

E-ISSN : 2614-8404
P-ISSN : 2776-3234



JISA

(JURNAL INFORMATIKA
dan SAINS)



VOL 9 No 1
June

Managed by :

Program Studi Teknik Informatika
UNIVERSITAS TRILOGI

2026

JISA

(Jurnal Informatika dan Sains)

Volume 9, Edition 1, June 2026

Virtual Geocultural Tour of the Baduy Tribe at Bayah Dome Geopark Toward an Inclusive Global Geopark

Dentik Karyaningsih, Ahmad Sugiyarta, Donny Fernando, Reza Pramudita, Irfan Hadi, Muhammad Aris Febriawan Mail

Artificial Intelligence Adoption in Learning Systems: Systematic Literature Review

Hendi Sama, Aisyah Nurkayla, Stefanus Eko Prasetyo

Classification of High-Risk Provinces for Fintech Lending Based on TWP90 from LPBBTI OJK Using Interpretable Model for Artificial Intelligence Ethical Transparency

Adi Rizky Pratama, Ayu Ratna Juwita, Antika Zahrotul Kamalia, Jan Everhard Riwurohi

Short-Term and Long-Term Forecasting of Global Gold Prices Using LSTM and GRU Models

Fitria Fitria, Muhammad Syahid Pebriadi, Nitami Lestari Putri

Development of a Dairy Cattle Facial Recognition System Based on CNN for Digital Livestock Data Management

Andika Muhammad Nur Kholiq, Arief Suryadi Satyawan, Mokh Mirza Etnisa Haqiqi, Arief Abdillah, R Helkhan Sultan Fajar, Yusril Kamil

A Data-Driven Machine Learning Framework for Cybersecurity Risk Prediction Using Behavioral and Temporal Features from Email Server Logs

Frowin Rabanus Kifaru

Development of a School Profile Website as a Digital Information Medium Using the Waterfall Method

Muhammad Mustakim, Sofiansyah Fadli, Ahmad Tantoni

Evaluation of Computer-Based Testing Website Quality Using the DeLone and McLean Information System Success Model

Muhamad Sakinul Wadi, Sofiansyah Fadli, Mohammad Taufan Asri Zaen

Evaluation of Fresnel Zone in Ubiquiti Wireless Simulation and Implementation Using 802.11ac

Ahmad Tantoni, Mohammad Taufan Asri Zaen

Accuracy Comparison of Support Vector Machine, Random Forest, and K-Nearest Neighbors for Sundanese Speech Classification

Laela Nur Rohmah, Abdul Halim Anshor, Wahyu Hadikristanto

Usability Evaluation of the Yellow Cart Feature in the TikTok Shop Application Using the System Usability Scale (SUS)

Hikmah Sani Nadia, Evy Nurmiati, Nuryasin Nuryasin

An AHP–TOPSIS-Based Decision Support System for Objective Pencak Silat Athlete Selection at KONI Lamongan

Ahmat Rendi Saputra, Brina Miftahurrohmah

Comparison of LDA and BERTopic in Identifying Public Issues in the MBG Program

Nur Hayati, Saikin Saikin, Hairul Fahmi

Deployment of a Machine Learning Based Indoor Air Quality Predictor on NodeMCU ESP32 Using EloquentTinyML

Yaddarabullah Yaddarabullah, Umar Al Faruq

Development of a Marker-Based Augmented Reality Application for Computer Hardware Learning Using the Multimedia Development Life Cycle

Muhamad Adila Syahputra, Ade Syahputra, Silvester Dian Handy Permana

Managed by:
Program Studi Teknik Informatika
Universitas Trilogi

JISA	Vol : 9	Ed :1	Page : 001-108	Jakarta, Jun 2026	e-ISSN: 2614-8404	p-ISSN: 2776-3234
-------------	----------------	--------------	---------------------------	------------------------------	------------------------------	------------------------------

JISA

(Jurnal Informatika dan Sains)

Volume 9, Edition 1, June 2026

Advisor

Assoc.Prof.,Silvester Dian Handy Permana.,ST.,MTI.,PhD

Editor in Chief

Assoc.Prof.Budi Arifitama, S.T., MMSI.,Ph.D : Universitas Trilogi, Indonesia

Editorial Board

Ade Syahputra, S.T., M.Inf.Comm.Tech.Mgmt. : Universitas Trilogi, Indonesia
Assoc.Prof. Ir.Yaddarabullah, M.Kom.IPM., ASEAN.Eng : Universitas Trilogi, Indonesia
Maya Cendana, S.T., M.Cs. : Universitas Bunda Mulia, Indonesia
Assoc.Prof. Silvester Dian Handy Permana, S.T., M.T.I.Ph.D : Universitas Trilogi, Indonesia
Ketut Bayu Yogha. B, S.Kom., M.Cs : Universitas Trilogi, Indonesia
Ninuk Wiliani.,S.Si.,M.Kom : Universitas Pancasila, Indonesia
Dwi Pebrianti,Ph.D : International Islamic University Malaysia, **Malaysia**
Dr.Wahyu Caesarendra : Opole University, **Poland**
Erneza Dewi Krishnasari.,S.Ds.,M.Ds : Universitas Trilogi, Indonesia
Muhammad Lahandi Baskoro.,Ph.D : Coventry University, **United Kingdom**
Gatot Tri Pranoto.,S.Kom.,M.Kom : Universitas Trilogi, Indonesia

Reviewers

Azida Zainol.,Ph.D :Sheffield Hallam University, **United Kingdom**
Prof. Ir. Suyoto, M.Sc. Ph.D : Universitas Atma Jaya Yogyakarta
Dr. Ir. Albertus Joko Santoso, M.T. : Universitas Atma Jaya Yogyakarta
Prof. Setiawan Assegaff, ST, MMSI, Ph.D : STIKOM Dinamika Bangsa, Jambi
Michael Marchenko, Ph.D : Universitas Trilogi, Jakarta
Dwi Pebrianti,Ph.D : Universiti Malaysia Pahang, **Malaysia**
Prof.Dr.Hoga Saragih.,ST.,MT : Universitas Bakrie
Isham Shah Hassan.,Ph.D : Port Dickson Polytechnic **Malaysia**
Prof.Dr Abdul Talib Bon : Universiti Tun Hussein Onn, **Malaysia**
Wiwin Armoldo Oktaviani, S.T, M.Sc : Universitas Muhammadiyah Palembang,
Yosi Apriani, S.T, M.T : Universitas Muhammadiyah Palembang,
Dr. Gandung Triyono.,M.Kom : Universitas Budi Luhur, Indonesia
Ir. Lukito Edi Nugroho, M.Sc., Ph.D : Universitas Gadjah Mada, Indonesia
Dr. Soetam Rizky Wicaksono : Universitas Ma chung, , Indonesia
Aulia Akhrian Syahidi.,M.Kom. : Politeknik Negeri Banjarmasin, Indonesia

Editorial Address

Fakultas Sains Teknik dan Desain Lantai 3
Jalan Taman Makam Pahlawan No. 1, Kalibata, Pancoran, RT.4/RW.4, Duren Tiga, Pancoran, Kota
Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12760Telp :(021) 798001

Managed by:
Program Studi Teknik Informatika
Universitas Trilogi

JISA	Vol : 9	Ed :1	Page : 001-108	Jakarta, Jun 2026	e-ISSN: 2614-8404	p-ISSN: 2776-3234
-------------	----------------	--------------	---------------------------------	------------------------------------	------------------------------------	------------------------------------

Table of Content

Virtual Geocultural Tour of the Baduy Tribe at Bayah Dome Geopark Toward an Inclusive Global Geopark	001-009
<i>Dentik Karyaningsih, Ahmad Sugiyarta, Donny Fernando, Reza Pramudita, Irfan Hadi, Muhammad Aris Febriawan Mail</i>	
Artificial Intelligence Adoption in Learning Systems: Systematic Literature Review	010-018
<i>Hendi Sama, Aisyah Nurkayla, Stefanus Eko Prasetyo</i>	
Classification of High-Risk Provinces for Fintech Lending Based on TWP90 from LPBBTI OJK Using Interpretable Model for Artificial Intelligence Ethical Transparency	019-024
<i>Adi Rizky Pratama, Ayu Ratna Juwita, Antika Zahrotul Kamalia, Jan Everhard Riwurohi</i>	
Short-Term and Long-Term Forecasting of Global Gold Prices Using LSTM and GRU Models	025-029
<i>Fitria Fitria, Muhammad Syahid Pebriadi, Nitami Lestari Putri</i>	
Development of a Dairy Cattle Facial Recognition System Based on CNN for Digital Livestock Data Management	030-035
<i>Andika Muhammad Nur Kholiq, Arief Suryadi Satyawan, Mokh Mirza Etnisa Haqiqi, Arief Abdillah, R Helkhan Sultan Fajar, Yusril Kamil</i>	
A Data-Driven Machine Learning Framework for Cybersecurity Risk Prediction Using Behavioral and Temporal Features from Email Server Logs	036-041
<i>Frowin Rabanus Kifaru</i>	
Development of a School Profile Website as a Digital Information Medium Using the Waterfall Method	042-051
<i>Muhammad Mustakim, Sofiansyah Fadli, Ahmad Tanton</i>	
Evaluation of Computer-Based Testing Website Quality Using the DeLone and McLean Information System Success Model	052-059
<i>Muhamad Sakinul Wadi, Sofiansyah Fadli, Mohammad Taufan Asri Zaen</i>	
Evaluation of Fresnel Zone in Ubiquiti Wireless Simulation and Implementation Using 802.11ac	060-067
<i>Ahmad Tanton, Mohammad Taufan Asri Zaen</i>	
Accuracy Comparison of Support Vector Machine, Random Forest, and K-Nearest Neighbors for Sundanese Speech Classification	068-074
<i>Laela Nur Rohmah, Abdul Halim Anshor, Wahyu Hadikristanto</i>	
Usability Evaluation of the Yellow Cart Feature in the TikTok Shop Application Using the System Usability Scale (SUS)	075-079
<i>Hikmah Sani Nadia, Evy Nurmiati, Nuryasin Nuryasin</i>	

JISA	Vol : 9	Ed.1	Page : 001-108	Jakarta, Jun 2026	e-ISSN: 2614-8404	p-ISSN: 2776-3234
-------------	----------------	-------------	---------------------------	------------------------------	------------------------------	------------------------------

An AHP–TOPSIS-Based Decision Support System for Objective Pencak Silat Athlete Selection at KONI Lamongan 080-087

Ahmat Rendi Saputra, Brina Miftahurrohmah

Comparison of LDA and BERTopic in Identifying Public Issues in the MBG Program 088-095

Nur Hayati, Saikin Saikin, Hairul Fahmi

Deployment of a Machine Learning Based Indoor Air Quality Predictor on NodeMCU ESP32 Using EloquentTinyML 96-102

Yaddarabullah Yaddarabullah, Umar Al Faruq Mail

Development of a Marker-Based Augmented Reality Application for Computer Hardware Learning Using the Multimedia Development Life Cycle 103-108

Muhamad Adila Syahputra, Ade Syahputra, Silvester Dian Handy Permana

JISA	Vol : 9	Ed.1	Page : 001-108	Jakarta, Jun 2026	e-ISSN: 2614-8404	p-ISSN: 2776-3234
-------------	----------------	-------------	---------------------------	------------------------------	------------------------------	------------------------------

Virtual Geocultural Tour of the Baduy Tribe at Bayah Dome Geopark Toward an Inclusive Global Geopark

Dentik Karyaningsih^{1*}, Ahmad Sugiyarta², Donny Fernando³, Reza Pramudita⁴,
Irfan Hadi⁵, Muhammad Aris Febriawan⁶

^{1*}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Serang Raya

²Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Serang Raya

³Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Serang Raya

⁴Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Serang Raya

⁵Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Serang Raya

⁶Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Serang Raya

Email: ^{1*}karya.tiek@gmail.com, ²ahmad.sugiyarta@unsera.ac.id, ³mr.donny2008@gmail.com,
⁴hanyqueen14@gmail.com, ⁵friexlif@gmail.com, ⁶arisfebriawan26@gmail.com

Abstract – The Bayah Dome Geopark in Lebak Regency possesses significant geological and cultural heritage; however, it currently holds the status of an "Aspiring Geopark." A primary challenge in optimizing its geotourism potential—particularly for the Baduy tribe—is the physical accessibility constraint caused by adverse weather conditions and rugged terrain. This study aims to develop and evaluate a mobile-based 360-degree virtual tour website as an inclusive digital geocultural and educational medium. The research follows a Research and Development (R&D) approach using the Multimedia Development Life Cycle (MDLC) framework, encompassing concept, design, material collection, assembly, testing, and distribution stages. The system integrates 360-degree panoramic visualizations with interactive navigation and informational hotspots to represent Baduy's local wisdom. Empirical results from black-box testing confirmed 100% functional validity across all system features. Furthermore, a User Acceptance Test (UAT) conducted using a Likert scale yielded an average acceptance index of 82.8%, categorizing the platform as "Very Good." This virtual tour serves as a strategic digital solution for cultural preservation and sustainable geotourism promotion, effectively surmounting geographical and temporal barriers for a global audience.

Keywords – Geopark Bayah Dome, Geocultural, Baduy Tribe. Virtual Tour 360, Website

I. INTRODUCTION

A Geopark is a single or composite geographical area that possesses valuable geological heritage sites (geosites) and landscapes, related to aspects of Geological Heritage, Geological Diversity, Biodiversity, and Cultural Diversity. Therefore, every region throughout Indonesia that has geological heritage, landscapes, and other forms of diversity should be designated as a Geopark. A Geopark also incorporates educational aspects as a means of disseminating knowledge in the field of geoscience, highlighting the uniqueness and diversity of Earth's heritage, as well as economic aspects through community involvement in area management as geotourism [1][2]. Lebak Regency possesses numerous geological heritage sites as well as rich biodiversity and cultural diversity with strong potential to be developed into a nationally and even internationally recognized Geopark. Consequently, the Ministry of Energy and Mineral Resources (ESDM) has designated the Bayah Dome Geopark, which comprises 32 geological heritage sites. The Bayah Dome Geopark is currently an Aspiring Geopark and has not yet been officially designated as either a National Geopark or a UNESCO Global Geopark. To achieve National Geopark status in 2023 and UNESCO Global Geopark status in 2024, comprehensive socialization, promotion, and dissemination of information are required, which must be widely accessible to the public through information and communication technology.

One form of cultural diversity is the Baduy Tribe, which represents cultural geotourism that upholds the preservation of local wisdom, including language,

customs, and other cultural elements [3]. The Baduy are one of Indonesia's indigenous communities that have consistently maintained their local wisdom to this day. The Baduy Tribe is located in Lebak Regency, Banten. The noble values embedded in Baduy customary law emphasize preserving and protecting nature without altering it, let alone damaging it [4]. The Baduy Tribe is a cultural tourism destination of world-class significance, as it is already recognized internationally [5]. However, in its development, local wisdom-based culture has gradually been eroded by modernization, affecting ethics, habits, norms, and traditions. As a result, many younger generations in Indonesia are still unfamiliar with the local cultural heritage of their respective regions [6]. Therefore, it is important to introduce local culture, particularly the Baduy local culture of Lebak Regency, to the wider community. Efforts to introduce Baduy local culture have included visiting the Baduy cultural villages and observing traditional ceremonies held at certain times. However, tourism activities or merely observing ceremonies have not been sufficient for optimal cultural introduction, considering the vast and hilly geographical conditions of the Baduy area. Additionally, unfavorable weather conditions may limit access, making it difficult to visit and fully experience the Baduy cultural villages. Currently, technology is widely used as a medium for promotion and education, particularly mobile-based technology, which offers various conveniences and interactive features that can attract public interest. Speed and convenience are among the positive values of the internet, as applications can be easily accessed by communities in various regions using only an internet



connection [7]. Geotourism destinations should be equipped with interactive applications that can serve as alternatives during unfavorable weather conditions and when there is a shortage of tour guide human resources during peak visitor periods. Through the use of interactive mobile-based applications, the community is also expected to develop awareness of the importance of preserving and not damaging the surrounding geotourism environment [8]. Based on this background, the problem formulation identified is the absence of promotional and educational media to introduce, disseminate, and educate the public about the Bayah Dome Geopark in support of conservation and sustainable economic development. To address this issue, this research focuses on the development of a mobile website-based 360-degree geocultural virtual tour application. This study employs a Research and Development (R&D) methodology, specifically utilizing the Multimedia Development Life Cycle (MDLC) framework—encompassing concept, design, material collecting, assembly, testing, and distribution—to ensure the systematic creation of an effective digital tourism platform.

Previous studies have been conducted under the title Design and Development of a Virtual Tour Application as a Tourism Information Media in Majalengka Regency Based on Android. The research method employed was Agile, consisting of the stages of Planning, Design, Development, Testing, Release, and Feedback [9]. Another study entitled Virtual Tour Application of Mount Pilar Natural Tourism Site in Serang Regency enabled users to view locations from unlimited and seamless perspectives [10]. A further study, entitled Interactive Multimedia-Based Virtual Tour Application for Qubu Resort Pontianak, utilized interactive multimedia as an information medium for the Qubu Resort Pontianak tourism destination and served as a promotional tool for the resort [11]. In addition, a study entitled Design and Development of a Virtual Reality Tour as a Tourism Promotion Media in Riau Province applied the ADDIE development model, which consists of five stages: analysis, design, development, implementation, and evaluation [12].

Based on this background, the problem formulation identified is the absence of promotional and educational media to introduce, disseminate, and educate the public about the Bayah Dome Geopark in support of conservation and sustainable economic development. To address this, the research methodology employed is Research and Development (R&D) utilizing the Multimedia Development Life Cycle (MDLC) framework—encompassing concept, design, material collecting, assembly, testing, and distribution. The result of this study is the development of a mobile website-based 360° geocultural virtual tour of the Baduy Tribe, which serves as an alternative medium for geotourism during unfavorable weather conditions as well as an effort to preserve and safeguard local wisdom.

II. RESEARCH METHODOLOGY

This study aims to introduce and preserve the Baduy Geocultural heritage of the Bayah Dome Geopark through a mobile-based 360° Virtual Tour application with

positioning features for the Bayah Dome Geopark. Through this application, Baduy geotourism is expected to be widely recognized by the public, even at a global level, while also providing educational value to the community in terms of preservation, conservation, and support for sustainable economic development through a website-based platform.

The development of the website-based 360-degree virtual tour follows the Multimedia Development Life Cycle (MDLC) methodology. This methodology is a structured framework specifically designed for creating multimedia-based products, ensuring a systematic and iterative process from initial planning to final deployment [13]. In this study, the MDLC approach—adopting the Luther-Sutopo model—is utilized to integrate various digital assets, such as 360-degree panoramic photography, interactive navigation, and informational hotspots. The process consists of six integrated stages: (1) **Concept**, to define the objectives and target audience; (2) **Design**, to create the storyboard and navigation structure; (3) **Material Collecting**, to gather all necessary digital assets; (4) **Assembly**, to develop the interactive platform; (5) **Testing**, to ensure functional validity; and (6) **Distribution**, to deploy the application on the Geopark's official website.

The development of the website-based Virtual Tour of Baduy Geocultural Diversity is carried out using an engineering approach, which consists of the stages of analysis, design, implementation, and evaluation. The research stages are illustrated in Figure 1 as follows:

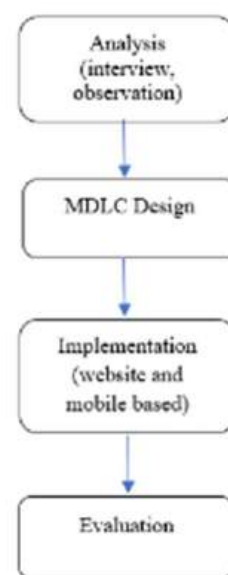


Figure 1. Research Stage

The analysis stage is the initial step in the development process of the website-based Virtual Tour of Baduy Geocultural Diversity as a medium for introducing cultural geotourism objects of the Bayah Dome Geopark. This stage consists of two main activities: field observation and data collection. Field observations were conducted at the Office of the Bayah Dome Geopark Management Agency in Lebak Regency, Banten, using interview methods, as well

as at the Lebak Regency Tourism Office and Kanekes Village, Ciboleger District, for data acquisition.

The design stage involves creating a program blueprint based on the results obtained during the analysis stage. The system development method used in this study is the Multimedia Development Life Cycle (MDLC), which adopts the Luther method with modifications to its stages. The process begins with the concept stage and ends with the distribution stage. However, the material collecting stage can be carried out in parallel with the assembly stage. The Luther–Sutopo version of the stages consists of concept, design, collecting, assembly, testing, and distribution [14].

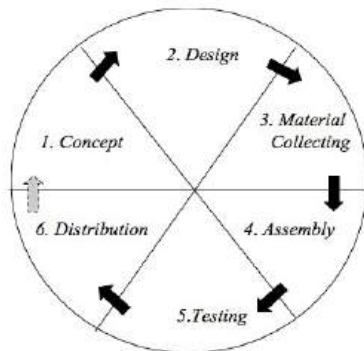


Figure 2. Luther–Sutopo Multimedia Development Cycle

- a. The concept stage is intended to determine the objectives of the multimedia and to identify the target audience. This stage also defines the type of application as interactive and establishes the application's purpose, namely to introduce geotourism objects representing the geocultural diversity of Baduy culture. Basic design rules are also determined at this stage, such as size and target specifications. The output of this stage is a narrative document on Saba Budaya Baduy, along with explanations of its location components.
- b. The design stage is the process of developing interface specifications and determining the required materials for the 360° panorama virtual tour application. At this stage, storyboards are used to illustrate the description of each scene by including all multimedia elements. In this design process, the researcher prepares the storyboard using applications such as Canva.
- c. The material collecting stage involves gathering materials in accordance with the project requirements. These materials may include photographs or videos. In this study, the materials used consist of 360-degree panoramic photographs.
- d. The assembly stage involves the creation of all multimedia objects and materials. Project development is carried out based on the design stage, such as storyboards, flowcharts, or navigation structures. This stage utilizes the finalized design and all collected materials, employing applications including Panorama Studio 360, 360 Studio, and HTML/CSS for the website-based platform.
- e. The testing stage is conducted after the assembly process is completed by running the project to ensure that no errors are present. This testing is referred to as the alpha testing stage (black-box testing), which is performed by experts in informatics engineering. Furthermore, system testing includes black-box testing, a software quality testing method that focuses on functional aspects [15].

f. At this stage, the project is stored and deployed on a storage medium in the form of a website. The virtual tour is hosted on the official website of the Bayah Dome Geopark Management Agency and can be accessed via personal computers, laptops, and Android mobile devices through the link <https://geoparkbayahdome.com/vt/>.

g. The implementation stage is carried out to apply the research results that have been tested and deemed feasible. The virtual tour will be distributed and utilized by the Bayah Dome Geopark Management Agency of Lebak Regency.

h. The evaluation stage aims to determine whether there are any issues arising from the implementation and to identify opportunities for further development of the Baduy Geocultural Diversity Virtual Tour system of the Bayah Dome Geopark.

III. RESULTS AND DISCUSSION

The results of this research demonstrate the systematic implementation of the system development process aimed at creating an immersive digital experience for the Bayah Dome Geopark. Following the methodology established in the previous section, the development of the website-based 360° virtual tour adheres strictly to the Multimedia Development Life Cycle (MDLC) stages. This structured approach ensures that each technical and creative phase—from initial conceptualization to final deployment—is aligned with the objective of preserving and promoting Baduy geocultural heritage.

1. Concept

The concept stage is the foundational phase intended to determine the specific objectives of the multimedia project and to identify the target audience. At this stage, the application is defined as an interactive medium with the primary purpose of introducing geocultural objects that represent the unique diversity of Baduy culture. Basic design rules, including display size and target specifications, are also established to ensure compatibility across various devices. The final output of this stage is a narrative document regarding Saba Budaya Baduy, which includes detailed explanations of the location components to be included in the virtual environment.

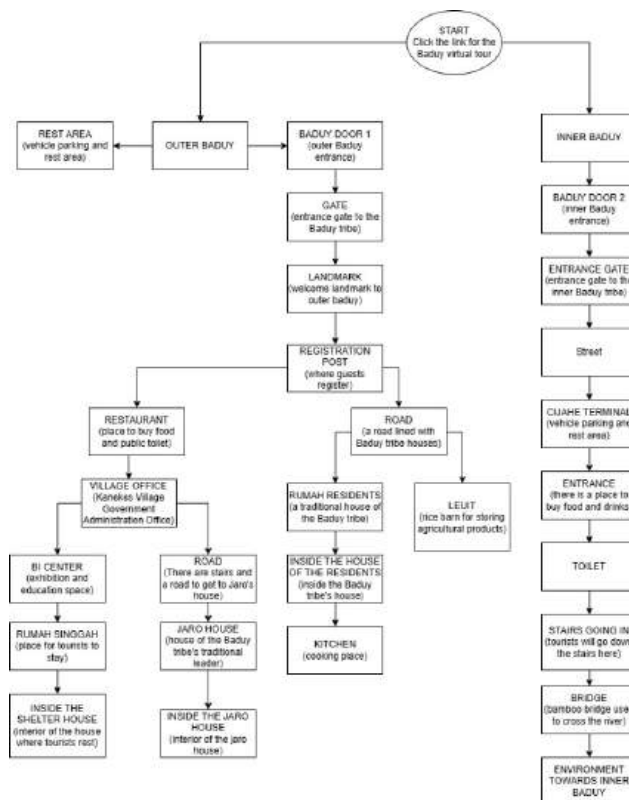


Figure 3. Concept

The research at this stage focuses heavily on the navigation flow of the Baduy Virtual Reality Tour to ensure a logical user experience. This navigation flow is structured before entering the design stage, serving as the primary reference framework for the creation of storyboards and the overall development of the system. To realize this structure, interactive hotspot buttons are required as the primary interaction medium, allowing users to transition between different panoramic scenes.

The application is operated by manipulating the 360-degree panoramic images, where users can touch and drag the view or select specific navigation buttons to determine their direction of exploration. By providing a comprehensive visual reconstruction of the Baduy area, this virtual tour serves as an ideal solution for individuals who wish to learn about the tribe but face limitations related to time, distance, or physical access to the Lebak Regency site.

2. Design

The design stage involves the development of detailed interface specifications and the determination of materials required for the 360-degree panoramic virtual tour. During this phase, storyboards are created to illustrate the description of each scene, incorporating all necessary multimedia elements to ensure a cohesive user experience. These storyboards serve as a visual blueprint and were developed using professional design applications such as Canva.

A critical aspect of the design stage is the specification of the user interface (UI) components that facilitate interaction within the virtual environment. The interface presented in this system includes the following components:

The navigation bar of the Baduy virtual tour website features a set of intuitive icons and buttons designed to provide seamless control over the 360-degree environment:



Figure 4. Navigation Button

- Rotation Control Buttons.**
These include the Left Rotation Button (←), Right Rotation Button (→), Upward Rotation Button (↑), and Downward Rotation Button (↓). Their primary function is to control the viewing direction, allowing users to look around the virtual space freely.
- Zoom Buttons.**
The interface provides a "Zoom In" feature represented by a Plus (+) icon and a "Zoom Out" feature represented by a Minus (-) icon. These buttons allow users to adjust the magnification level of the virtual tour display for a more detailed view.
- Auto-Rotate Button.**
Identified by a rotating symbol (↻) or play icon (▶), this button activates an automatic 360-degree rotation, enabling a hands-free panoramic view of the surroundings.
- Fullscreen Button.**
Represented by four outward-facing corners (⛶), this function allows users to toggle the fullscreen mode for a more immersive exploration experience.

To ensure optimal usability, all buttons are designed with colors and shapes that contrast sharply with the background, enhancing visibility. Furthermore, the buttons are equipped with hover effects to provide interactive feedback and are sized appropriately to ensure comfortable navigation on both desktop and mobile devices.

3. Material Collecting

The material collecting stage involves gathering all digital assets and information required to fulfill the project's specifications. In this study, the primary materials consist of 360-degree panoramic photographs captured through field observations at the Bayah Dome Geopark Management Agency, the Lebak Regency Tourism Office, and Kanekes Village. These materials are essential for constructing the immersive virtual environment. The following visual materials were collected to serve as the core content of the virtual tour:

- Rest Area**



Figure 5. Rest Area

Rest Area Visual data was collected at the starting point of the route, showcasing facilities such as seating areas, public toilets, and parking spaces intended for tourists.

- Entrance Gate**





Figure 6. Entrance Gate

Entrance Gate This material captures the traditional structure that serves as the official entry point to the Baduy region, reflecting the tribe's local wisdom through its architectural design.

c. Welcome Landmark



Figure 7. Welcome Landmark

Welcome Landmark Photographs were taken of the informational signage and welcome message that officially mark the entrance to Baduy territory .

d. Guest Registration Area



Figure 8. Guest Registration Area

Guest Registration Area Visuals of the registration post were gathered to represent the point where visitor data is collected and initial guidelines regarding rules and etiquette are provided.

e. Baduy Residential Settlement



Figure 9. Baduy Residential Settlement

Baduy Residential Settlement The collected materials display the distinctive layout of the residential area, highlighting the uniform architectural philosophy and the adaptation of houses to the natural environment .

f. Souvenir Center



Figure 10. Souvenir Center

The souvenir center is located in front of residents' houses, where local people sell distinctive Baduy products. Through the virtual tour, visitors can view various handicrafts such as woven fabrics, koja bags, machetes (golok), beaded jewelry, and other traditional products neatly arranged on the house veranda. This area illustrates the traditional economic activities of the Baduy community.

g. Leuit



Figure 11. Leuit

The leuit is a traditional stilted structure used for storing rice. The virtual tour presents the leuit structure made from natural materials with a high roof in the form of julang ngapak. Visitors can observe construction details specifically designed to keep rice dry and protected from pests, as well as the traditional ventilation system applied.

h. Kanekes Village Office



Figure 12. Kanekes Village Office

Kanekes Village Office Documentation was made of the administrative building, which serves as a public service center and uniquely combines traditional elements with modern functionality.

i. BI Center



Figure 13. BI Center

BI Center Visual content includes the information and education center, which houses photo displays, artifacts, and historical data regarding Baduy daily life.

j. Tourist Rest House



Figure 14. Tourist Rest House

Tourist Rest House Materials were gathered of the traditional lodging facilities, which utilize natural materials such as wood, bamboo, and palm fiber while maintaining the authentic Baduy style.

k. Jaro House



Figure 15. Jaro House

Jaro House Visuals capture the residence of the customary leader, which features a larger structure and specific architectural ornaments that signify the social status of its occupant .

l. Entrance Gate 2



Figure 16. Entrance Gate 2

Entrance Gate 2 Documentation was conducted of the simple bamboo and wood gate that serves as a transitional zone into the more sacred Inner Baduy territory.

m. Cijahe Terminal



Figure 17. Cijahe Terminal

Cijahe Terminal This material represents the final stopping point for motorized vehicles, where visitors must transition to traveling on foot.

n. Entrance Ticket Payment Area (Gate 2)



Figure 18. Entrance Ticket Payment Area (Gate 2)

Entrance Ticket Payment Area (Gate 2) Visual data includes the registration and payment point specifically for visitors traveling toward the Inner Baduy villages of Cibeo, Cikertawana, and Cikeusik.

4. Assembly

The assembly stage is the phase where all multimedia objects and materials are integrated into a functional system based on the finalized designs, such as storyboards and navigation structures. In this study, the project development was executed using a combination of specialized software, including Panorama Studio 360 and 360 Studio for panoramic processing, along with HTML and CSS for the website-based platform implementation. A key output of this stage is the creation of the interactive landing interface, which serves as the gateway for the entire virtual experience.



Figure 19. Main Page

The Main Page provides users with two primary location options to initiate their journey: Ciboleger, which grants access to the Outer Baduy Virtual Tour, and Cijahe, which leads to the Inner Baduy Virtual Tour (comprising Cibeo, Cikertawana, and Cikeusik). Users can intuitively select their preferred starting point by clicking or tapping on the respective location card. This interface is designed to be responsive, ensuring that the visual elements and selection buttons remain accessible across both desktop and mobile devices.

5. Testing

Following the assembly process, the testing stage is conducted to ensure the system is free of errors and satisfies the predefined requirements. This stage is divided into functional validation through black-box testing and user perception evaluation through a User Acceptance Test (UAT).

a. Blackbox Testing

Black-box testing focuses on evaluating the application's functionality without considering its internal structure or source code. This testing was performed by informatics experts to ensure the application operates smoothly and provides an optimal user experience.

Table 1. Blackbox Testing Result

Input Testing	Detail	Output Testing	Result
Click "Start Virtual Tour" button	Checking whether the button functions to start the virtual tour	Virtual tour starts and displays the initial 360° location view	Valid
Click navigation buttons (right, left, up, down arrows)	Ensuring navigation movement in the virtual tour functions properly	Camera moves according to the selected direction	Valid
Click information hotspot	Checking whether the hotspot displays information when clicked	Information pop-up appears with location/object details	Valid
Zoom in/out using mouse scroll	Ensuring zoom functionality operates properly	Virtual tour display is zoomed in/out	Valid
Drag virtual tour display	Checking whether users can drag the 360° view	Display moves according to mouse drag	Valid
Click "Fullscreen" button	Ensuring fullscreen mode can be accessed	Virtual tour is displayed in fullscreen mode	Valid
Mobile device rotation	Checking responsiveness on mobile devices	Display adjusts to screen orientation	Valid

The results confirm that 100% of all features function correctly according to the established specifications, with no significant technical issues identified.

b. User Acceptance Test

The User Acceptance Test was conducted with 20 participants using a Likert scale to measure perceptions, attitudes, and opinions regarding the application. Respondents indicated their level of agreement across ten

criteria using a five-point scale ranging from "Strongly Disagree" to "Strongly Agree".

Table 2. User Acceptance Test Result

No	Question	S (5)	S (4)	N (3)	T (2)	ST (1)	Total Score	Interpretation (Index %)
1	The virtual tour application interface display is attractive	35	28	15	2	0	80	80%
2	The information presented is easy to read	60	16	9	0	1	85	85%
3	The quality of the displayed 360° images is good	45	24	15	0	0	84	84%
4	The content is appropriate to Baduy culture and local wisdom	70	8	3	6	0	87	87%
5	The application runs smoothly without technical issues	25	48	6	2	0	81	81%

6	The application is responsive on various devices (PC/mobile)	50	20	15	0	0	85	85%
7	No errors occur while using the application	45	28	6	2	1	82	82%
8	The virtual tour provides an engaging experience	35	36	6	4	0	81	81%
9	Navigating within the virtual tour is easy to understand	35	40	6	0	1	82	82%
10	The zoom in/out function works properly	40	16	24	0	0	80	80%
Average	Total index (%) / number of questions							82.8%

The website-based Baduy Geocultural Virtual Tour achieved an average acceptance index of 82.8%, falling into the "Very Good" category. The highest individual score (87%) was for content suitability with Baduy culture, indicating that the virtual tour accurately reflects local

wisdom and traditional values. Overall, the UAT results confirm that the system meets user expectations and is highly suitable for sustainable geotourism promotion.

6. Distribution

The distribution stage is the phase where the finalized project is stored and deployed on a designated storage medium to ensure broad public access. In this study, the Baduy Geocultural Virtual Tour has been successfully integrated into a website-based platform, allowing it to serve as a sustainable digital asset for the region.

The application is officially hosted on the website of the Bayah Dome Geopark Management Agency and can be accessed through the following URL: <https://geoparkbayahdome.com/vt/>. To ensure an inclusive global reach, the system is designed with cross-device compatibility, enabling seamless exploration via personal computers, laptops, and Android mobile devices.

Furthermore, the research results have been formally distributed to and implemented by the Bayah Dome Geopark Management Agency of Lebak Regency. This implementation allows the agency to utilize the virtual tour as a strategic medium for geotourism promotion, particularly as an alternative solution during adverse weather conditions or for visitors who face geographical and time limitations. By hosting this platform on their official site, the agency supports the ongoing effort toward achieving UNESCO Global Geopark status by enhancing the dissemination of information through modern communication technology.

IV. CONCLUSION

This study successfully developed a website-based 360° Geocultural Virtual Tour of the Baduy Tribe within the Bayah Dome Geopark as an innovative medium for promotion, education, and cultural preservation. The virtual tour enables users to explore both the Outer Baduy and Inner Baduy areas virtually, providing an interactive and immersive experience that overcomes limitations related to distance, time, weather conditions, and accessibility. By integrating geocultural elements, local wisdom, and geotourism concepts, the developed system supports the dissemination of knowledge about Baduy culture while promoting sustainable geotourism within the Bayah Dome Geopark.

The development process followed an engineering approach using the Multimedia Development Life Cycle (MDLC), encompassing the stages of analysis, design, material collecting, assembly, testing, distribution, implementation, and evaluation. The application integrates 360° panoramic visualization, interactive navigation, hotspot information, and responsive design to ensure usability across desktop and mobile devices. Black-box testing results confirmed that all system functionalities operated correctly and met the predefined specifications, indicating system reliability and technical feasibility.

Furthermore, the User Acceptance Test (UAT) results demonstrated a high level of user satisfaction, with an average acceptance index of 82.8%, categorized as very good. Users positively evaluated the system in terms of interface design, ease of use, content clarity, cultural relevance, and overall experience. These findings indicate that the virtual tour application is effective as a digital

platform for cultural education, conservation awareness, and sustainable economic support. Overall, the website-based Baduy Geocultural Virtual Tour has strong potential to be utilized by the Bayah Dome Geopark Management Agency as a strategic tool for inclusive geotourism development and the preservation of local cultural heritage at both national and global levels.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Universitas Serang Raya for funding this research through the Internal Grant Scheme for the 2024 implementation period, which enabled the study to be conducted successfully. Appreciation is also extended to the Head of Kanekes Village, the Baduy community, as well as the Department of Culture and Tourism and the Bayah Dome Geopark Management Agency of Lebak Regency for granting permission and providing support for the implementation of this research.

REFERENCES

- [1] Darsiharjo, U. Supriatna, and I. M. Saputra, "Pengembangan Geopark Ciletuh Berbasis," *Jurnal Manajemen Resort dan Leisure*, vol. 13, no. 1, pp. 55–60, 2016.
- [2] D. Haryanto and W. Pirgana, "Pengembangan User Interface Pada Website Geopark Belitong," *Jurnal Ilmiah Informatika Global*, vol. 12, no. 1, pp. 53–58, 2021.
- [3] I. Wahab, "Identifikasi Cerita Anak Berbasis Budaya Lokal Untuk Membudayakan Literasi Di Sd," *Satya Widya*, vol. 35, no. 2, pp. 176–185, 2019, doi: 10.24246/j.sw.2019.v35.i2.p176-185.
- [4] D. Karyaningsih, D. Susandi, and E. Safaah, "Implementasi Mobile Learning Berbasis Android Kearifan Lokal Baduy Kabupaten Lebak Untuk Anak Usia Dini," *jurnal Teknik Informatika Unis*, vol. 9, no. 2, pp. 99–110, 2021.
- [5] L. M. Dewi et al., "Pelestarian Bahasa Baduy Sebagai Kearifan Lokal Melalui Aplikasi Tour Guide Berbasis Android Masyarakat Baduy merupakan Indonesia untuk wisatawan lokal dan mancanegara oleh para guide nya , Guide berbasis Mobile android merupakan alternatif utama dalam," vol. 19, pp. 378–389, 2023.
- [6] P. L. Ekawati and A. Z. Falani, "Pemanfaatan Teknologi Game Untuk Pembelajaran Mengenal Ragam Budaya Indonesia Berbasis Android," *Jurnal Link*, vol. 22, no. 1, pp. 30–36, 2015.
- [7] Ismai, "Perancangan Website Sebagai Media Promosi Dan Informasi," *Journal Of Informatic Pelita Nusantara*, vol. 3, no. 1, pp. 82–86, 2018.
- [8] H. Hermawan, "Geowisata Sebagai Model Pemanfaatan," 2018.
- [9] P. Putriani and T. F. Prasetyo, "Rancang Bangun Aplikasi Virtual Tour Sebagai Media Informasi Wisata di Kabupaten Majalengka Berbasis Android," *Konferensi Nasional Ilmu Komputer (KONIK)*, pp. 400–405, 2021.
- [10] Z. Muttaqin, T. Nurhadiyan, and Y. Arifin, "Aplikasi Virtual Tour Tempat Wisata Alam Gunung Pilar Di Kabupaten Serang," *ProTekInfo(Pengembangan Riset dan Observasi Teknik Informatika)*, vol. 8, no. 1, pp. 1–6, 2021, doi: 10.30656/protekinfo.v8i1.5056.
- [11] N. Safriadi, H. Sastypratiwi, and N. Tamara, "Aplikasi Virtual Tour Berbasis Multimedia Interaktif Pada Objek Wisata Qubu Resort Pontianak," *Seminar Nasional Teknologi Informasi Dan Komunikasi (SEMNASITIK) X*, pp. 411–416, 2018.
- [12] Y. Fatma, R. Hayami, A. Budiman, and Y. Rizki, "Rancang Bangun Virtual Tour Reality Sebagai Media Promosi Pariwisata Di Propinsi Riau," *Jurnal Fasilkom*, vol. 9, no. 3, pp. 1–7, 2019.
- [13] M. Sidik, "Perancangan dan Pengembangan E-commerce dengan Metode Research and Development," *Jtiust*, vol. 04, no. 1, pp. 99–107, 2019.
- [14] E. P. Harahap, U. Rahardja, and M. Salamuddin, "Aplikasi Panduan dan Pembayaran Tiket Masuk Mendaki Gunung Menggunakan Metodologi Sistem Multimedia Luther-Sutopo," *SATIN - Sains dan Teknologi Informasi*, vol. 4, no. 2, p. 9, 2019, doi: 10.33372/stn.v4i2.404.
- [15] L. Setiyani, "Pengujian Sistem Informasi Inventory Pada Perusahaan Distributor Farmasi Menggunakan Metode Black Box Testing," *Techno Xplore : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 4, no. 1, pp. 1–9, 2019, doi: 10.36805/technoxplore.v4i1.539.
- [16] V. H. Pranatawijaya, W. Widiatry, R. Priskila, and P. B. A. A. Putra, "Penerapan Skala Likert dan Skala Dikotomi Pada Kuesioner Online," *Jurnal Sains dan Informatika*, vol. 5, no. 2, pp. 128–137, 2019, doi: 10.34128/jsi.v5i2.185.
- [17] T. Nempung, T. Setiyaningsih, and N. Syamsiah, "Otomatisasi Metode Penelitian Skala Likert Berbasis Web," no. November, pp. 1–8, 2015.



Artificial Intelligence Adoption in Learning Systems: Systematic Literature Review

Hendi Sama¹, Aisyah Nurkayla^{2*)}, Stefanus Eko Prasetyo³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Internasional Batam
Email: ¹hendi@uib.ac.id, ²2231122.aisyah@uib.edu, ³stefanus@uib.ac.id

Abstract – The integration of Artificial Intelligence (AI) in education has accelerated rapidly, reshaping teaching and learning processes through intelligent, adaptive, and data-driven systems. Despite its widespread adoption, a consolidated understanding of implementation trends, benefits, and challenges across educational contexts remains limited. This study aims to examine the current state of AI adoption in education by identifying key applications, success factors, and implementation barriers. A systematic literature review (SLR) was conducted by analyzing 32 peer-reviewed journal articles published within the last five years and indexed in reputable academic databases. The reviewed studies focus on adaptive learning systems, intelligent tutoring systems, and learning analytics. The findings demonstrate that AI contributes significantly to personalized learning, real-time performance assessment, and improved learner engagement. However, effective implementation is strongly influenced by institutional readiness, educator digital literacy, data quality, and ethical governance. Major challenges identified include data privacy concerns, lack of standardization, and unequal access to technological infrastructure. This study concludes that AI has substantial potential to support more adaptive and inclusive educational systems, but sustainable integration requires coordinated efforts among educators, policymakers, and technology developers.

Keywords – *artificial intelligence, education technology, adaptive learning, intelligent tutoring system, learning analytics.*

I. INTRODUCTION

Artificial Intelligence (AI) has rapidly emerged as a transformative force across multiple sectors, with education being one of the most promising areas for innovation. In contemporary learning environments, the integration of AI technologies aims to enhance instructional effectiveness, personalize learning experiences, and improve overall student outcomes. Adaptive learning systems, intelligent tutoring systems, and learning analytics platforms represent the forefront of AI applications in education, offering tailored instructional content, real-time feedback, and dynamic assessment methods [1]. These AI-driven platforms are capable of adjusting content difficulty to match individual student proficiency, providing targeted feedback to reinforce understanding, and enabling adaptive assessments that optimize knowledge retention [2]. Furthermore, AI techniques such as classification, regression, and clustering empower educators to identify patterns in student learning, predict academic performance, and inform instructional design [3]. By leveraging these capabilities, instructors can monitor learning trends and intervene proactively, fostering responsive and data-informed learning environments.

Despite the promising potential, the adoption of AI in educational settings encounters significant barriers. High implementation costs, limited technological infrastructure, and institutional unpreparedness remain primary obstacles [4]. Social resistance and low digital literacy among educators further impede effective utilization, while inadequate training and perceived complexity of AI tools reduce adoption rates [5], [6]. The rapid pace of technological change may also overwhelm institutions, resulting in inconsistent application of AI solutions and limiting their long-term sustainability. These challenges

underscore the importance of supportive policies, targeted professional development, and strategic planning to ensure successful integration and lasting impact.

Inclusivity and ethical considerations represent critical dimensions of AI deployment in education. AI-powered systems can facilitate personalized learning for students with disabilities, providing adaptive resources and tailored support [7]. However, concerns regarding data privacy, algorithmic bias, and unequal access highlight the need for responsible and equitable implementation [8]. Overreliance on AI technologies may reduce human interaction, which is essential for social and emotional development, necessitating a balanced approach that integrates human instruction with AI support [9]. Learning analytics offers valuable insights into student engagement and performance [10], yet disparities in access and digital resources can exacerbate existing inequities [11]. These factors emphasize the importance of evaluating not only technological readiness but also social and ethical preparedness to ensure that AI adoption enhances educational equity.

Therefore, This study aims to provide a holistic analysis of AI adoption in educational settings by focusing on three objectives: identifying factors influencing AI adoption, including technical infrastructure, educator readiness, and institutional policies; examining the benefits and challenges associated with AI integration, including improvements in engagement, learning performance, and pedagogical efficiency; and assessing the impact of AI applications, such as intelligent tutoring systems and adaptive learning platforms, on student engagement and performance. By synthesizing recent empirical findings, this research offers insights into AI adoption trends, while highlighting ethical considerations and inclusive practices that can guide educators, policymakers, and technologists in implementing AI-driven learning solutions.



To achieve these objectives, the study employs a Systematic Literature Review (SLR) methodology. This approach enables comprehensive evaluation of mechanisms and outcomes of AI adoption, identifying both the drivers and barriers of acceptance and assessing the effects of AI on cognitive, behavioral, and emotional learner outcomes. The research is guided by three primary questions:

RQ1. What key factors influence the adoption of Artificial Intelligence (AI) in educational settings?

RQ2. What are the primary benefits and challenges associated with integrating AI into learning systems?

RQ3. How do AI applications such as intelligent tutoring systems and adaptive learning platforms affect student engagement and academic performance?

II. RESEARCH METHODOLOGY

This study employs a Systematic Literature Review (SLR) approach to comprehensively examine the application of Artificial Intelligence (AI) in education, focusing on adaptive learning systems and intelligent tutoring platforms. The aim is to identify, evaluate, and synthesize empirical evidence on factors influencing adoption, benefits, challenges, and the impact of AI on learning outcomes. The SLR method was selected for its structured and replicable framework in reviewing academic literature [12]. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, the review process consisted of four phases: identification, screening, eligibility, and inclusion.

A. Identification of Studies

The identification stage involved an extensive literature search conducted across two major academic databases, IEEE Xplore and ScienceDirect. These databases were selected due to their comprehensive coverage of peer-reviewed publications in Artificial Intelligence and educational technology. Boolean operators (AND, OR) were applied to combine keywords related to AI-based learning, adoption factors, and student outcomes. The search strings included:

- “Artificial Intelligence in Education” OR “AI-based Learning”
- “Adaptive Learning” OR “Intelligent Tutoring Systems”
- “Learning Analytics” OR “Educational Data Mining”
- “Technology Adoption” AND “Student Outcomes”

Search filters were applied to include peer-reviewed journal articles published between 2020 and 2025, written in English, and available in full-text format. This timeframe was chosen to capture recent developments and post-pandemic innovations in AI-based learning. The initial search yielded 500 articles, consisting of 165 articles from IEEE Xplore and 335 articles from ScienceDirect. All records were exported to Mendeley for reference management, and duplicate entries were removed. Subsequently, 285 articles were excluded due to restricted access, incomplete methodology, or limited relevance, resulting in 215 studies retained for the screening stage.

B. Screening and Study Selection

During the screening stage, the titles and abstracts of 215 studies were reviewed to assess their relevance to AI based learning in education. A total of 38 articles were excluded at this stage, leaving 177 articles for full-text assessment. The eligibility evaluation examined methodological rigor, empirical contribution, clarity of findings, and relevance to the research objectives. In this phase, 145 studies were excluded due to a non-AI focus ($n = 93$), conceptual or non-empirical nature ($n = 28$), or being review papers ($n = 24$). Eligibility assessments were conducted independently by reviewers, with discrepancies resolved through discussion. As a result, 32 studies were included in the final analysis. The overall selection process is illustrated in the PRISMA flow diagram (Fig. 1).

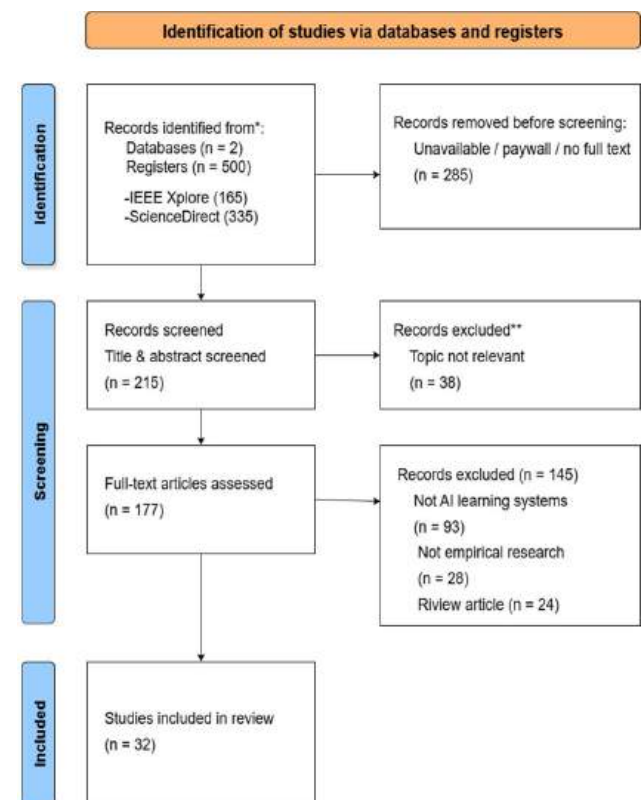


Figure. 1 Prisma Flow Diagram

C. Inclusion and Exclusion Criteria

Clear inclusion and exclusion criteria were established to ensure methodological rigor and relevance to the research objectives. Only peer-reviewed empirical studies focusing on AI applications in educational contexts were included. The criteria used in the selection process are summarized in Table 1.

Criteria	Inclusion	Exclusion
Year	Published articles between 2020 to 2025	Published before 2020.
Focus	AI in learning systems	Not related to AI in education
Type	Empirical research	Review or conceptual paper
Relevance	Related to adoption, impact, or use of AI	irrelevant topics to AI based learning

Table 1. Criteria for Selection Studies

D. Data Extraction

Relevant data were systematically extracted from each included study, including authors, publication year, research objectives, methodology, main findings, limitations, AI technology type, implementation context, and observed learning outcomes. Mendeley was used for reference management, while Microsoft Excel was utilized for data organization, synthesis, and pattern identification. This structured extraction process enabled comparative analysis and thematic synthesis related to AI adoption, benefits, and challenges.

Paper	Year	Focus / Main Topic
[13]	2025	Impact of AI-assisted learning on motivation and continuous learning intention.
[14]	2025	Personalized learning content generation using <i>Generative AI (AIGC)</i> .
[15]	2024	Student performance prediction using <i>deep learning</i> -based knowledge tracing.
[16]	2023	Adaptive formative assessment through AI-driven item difficulty adjustment
[17]	2022	Adaptive AI system to maintain the optimal learning zone and self-regulated learning.
[18]	2025	Impact of Generative AI on primary-level mathematics learning and attitudes.
[19]	2022	AI-based reading tool to enhance EFL students' oral reading fluency.
[20]	2023	Integration of <i>Open Learner Model (OLM)</i> and AI visualization for self-regulation.
[21]	2025	Emotionally enriched AI feedback for improving engagement and well-being.
[22]	2025	Use of emojis and AI to enhance interactivity in LMS environments.
[23]	2022	Data-driven learning design through AI and <i>Learning Analytics</i> .
[24]	2024	<i>ChatGPT-based Intelligent Tutoring System (ITS)</i> for programming education.
[25]	2025	Application of AI (NLP, ITS, VR/AR) to enhance engineering education.
[26]	2025	<i>Knowledge tracing</i> for analyzing conceptual relationships in learning.
[27]	2024	Relationship between AI literacy, attitudes, and motivation in EFL learning.
[28]	2023	Adaptive virtual assistant for promoting engagement and self-regulation.
[29]	2025	Effect of <i>adaptive learning platforms</i> on reading performance.
[30]	2025	Integration of AI-based adaptive learning in flipped classroom settings.
[31]	2025	AI framework (LAMB) for building intelligent learning assistants in LMS.
[32]	2024	Prediction of spelling errors using <i>machine learning</i> models.
[33]	2025	AI-driven adaptive system to enhance problem-solving skills in technical education.
[34]	2022	Student and instructor perceptions of personalized AI-supported learning.
[35]	2025	Peer network analysis for personalized content adaptation via AI.
[36]	2022	Real-time science learning outcome prediction using neurocognitive data.
[37]	2025	AI use in nursing education for improving clinical and cultural competence.
[38]	2025	Identifying student learning styles using adaptive AI modeling.
[39]	2025	Conversational AI agent for algebra learning and student interaction.
[40]	2025	Comparison of AI-based collaborative learning approaches.
[41]	2022	Personalization in MOOCs/LMS using <i>Intelligent Tutoring Systems (ITS)</i> .

[42]	2021	AI-based emotion recognition for detecting student engagement in online learning.
[43]	2025	Effects of <i>Generative AI</i> on essay revisions and student engagement.
[44]	2025	Automatic analysis of Cognitive Presence phases using Large Language Models.

Table 2. The general characteristics of the included studies

III. RESULTS AND DISCUSSION

This section presents and analyzes the synthesized findings from 32 peer-reviewed studies that met the established inclusion and exclusion criteria. The discussion integrates key insights on factors influencing AI adoption, the benefits and challenges of its integration, and its overall impact on student learning outcomes. Descriptive and interpretive analyses were combined to highlight relationships between variables, emerging trends, and educational implications. All data extraction and synthesis were conducted using Microsoft Word and Excel to ensure analytical consistency and transparency.

The results are structured thematically according to the formulated research questions, with an emphasis on interpretive connections between findings and prior theoretical models. In this comprehensive analysis, we delve deeper into the nuances of each theme, drawing on the collective evidence from the reviewed studies to provide a robust understanding of how AI is reshaping educational landscapes. By examining the interplay of technical, social, and pedagogical elements, we aim to offer educators, policymakers, and researchers a clearer roadmap for implementing AI in ways that are both effective and equitable. This approach not only summarizes the data but also interprets it within broader theoretical frameworks, such as the Technology Acceptance Model (TAM) and socio-technical systems theory, to underscore the multifaceted nature of AI integration in education. Furthermore, the discussion highlights potential gaps in literature and suggests avenues for future research, ensuring that the synthesis remains forward-looking and actionable.

A. RQ1: Main Factors Influencing the Adoption of AI in Learning Systems

1. Technical factors

The analysis reveals that AI adoption in learning environments is strongly influenced by technical factors, particularly ease of use and perceived usefulness, which are consistent with the Technology Acceptance Model (TAM) as demonstrated in [29]. Systems that provide real-time feedback, adaptive content, and seamless integration with institutional platforms tend to increase user acceptance and perceived system reliability. These findings align with recent educational technology studies emphasizing usability and interoperability as critical determinants of innovation diffusion. Furthermore, infrastructure readiness and educators' technical competence play a vital role, especially in higher education institutions where digital maturity significantly affects the sustainability of AI-based systems [34]. In



environments supported by robust digital infrastructure, such as universities utilizing advanced cloud computing capabilities, AI tools can be deployed more effectively, resulting in higher adoption rates. Conversely, limited bandwidth and outdated hardware often generate resistance due to system instability and delayed responsiveness, highlighting the importance of foundational IT investment as a prerequisite for successful AI implementation.

2. *Social and pedagogical factors*

Social and pedagogical dimensions also shape AI adoption through institutional support, AI literacy, and user trust. Study [27] identifies AI literacy as a mediating variable between user attitudes and sustained system adoption, indicating that confidence and competence in using AI tools foster long-term engagement. In addition, [37] emphasizes cultural adaptation and personalization as essential drivers of adoption, underscoring the need for AI systems to align with learners' contextual and cultural realities. Compatibility with Learning Management Systems (LMS), data security, and flexibility of AI frameworks such as LAMB are highlighted by [31] as enabling factors that reduce resistance among educators with limited programming skills. AI literacy in this context extends beyond technical knowledge to include awareness of ethical implications such as data privacy and algorithmic bias. Without adequate institutional training and policy support, educators may perceive AI as a threat to professional autonomy, thereby hindering adoption. Supportive institutional policies, professional development programs, and ethics-oriented workshops can cultivate a culture of innovation and encourage educators to view AI as a complementary instructional tool.

3. *Data readiness and systemic factors*

Beyond individual readiness, data availability and model adaptivity significantly influence implementation success. Studies such as [26] and [35] demonstrate that access to large, high-quality datasets and adaptive models capable of tracking learner behavior enhances personalization and system responsiveness. AI systems that dynamically adjust learning materials based on performance trends improve learner engagement and provide instructors with actionable insights. Integrative approaches, including eye-tracking-based AI [38], further enhance instructional feedback by enabling real-time observation of attention patterns and cognitive responses. Data readiness involves not only the volume of data but also its quality, diversity, and contextual relevance. Datasets representing diverse learner profiles reduce the risk of biased predictions, while adaptive models that continuously learn from new data inputs ensure long-term system relevance. In practice, this enables AI-driven learning environments to evolve alongside learners' needs, fostering more inclusive and learner-centered educational experiences.

Overall, AI adoption emerges as a socio-technical process shaped by the interaction between individual factors such as motivation, attitudes, and literacy and systemic factors, including infrastructure readiness and institutional policy. Challenges related to algorithmic bias [39] and digital inequality [37] persist, necessitating ethical governance frameworks and continuous professional development initiatives. By integrating technological capabilities with human-centered design principles, institutions can promote equitable and sustainable AI adoption. This perspective aligns with socio-technical systems theory, which emphasizes the alignment of social practices and technological infrastructures as a foundation for successful innovation in educational contexts.

B. *RQ2: The Main Benefits and Challenges of AI Integration in Education*

1. *Pedagogical and Learning Benefits*

The integration of AI into education provides substantial pedagogical benefits through personalized and adaptive instruction that optimizes learning paths based on individual progress and preferences. AI-driven learning systems enable adaptive feedback mechanisms that significantly enhance knowledge retention and learner motivation, with studies reporting improvements of up to 68.3% in learning outcomes [14], [16]. In flipped classroom models, AI increases learner autonomy and participation by allowing students to control their learning pace [30], while in Massive Open Online Courses (MOOCs), Intelligent Tutoring Systems (ITS) support self-regulated learning and adaptive assessment processes [41]. These benefits extend beyond efficiency, fostering learner ownership and accommodating diverse learning needs where traditional one-size-fits-all instructional approaches are often ineffective.

2. *Emotional and Motivational Impact*

AI also contributes to emotional engagement and motivational enhancement through empathy-driven and interactive learning environments. Emotion-enriched feedback and interactive features, such as emojis and sentiment-aware prompts, have been shown to cultivate positive affective responses without diminishing academic performance [21], [22]. In addition, Generative AI supports writing and creative tasks by reducing cognitive load and increasing learning motivation, particularly in disciplines that require expression and ideation. From an instructional standpoint, AI reduces teachers' workload by automating assessments, grading, and learning analytics generation, thereby supporting data-informed instructional decision-making [23], [31]. This automation enables educators to focus on personalized mentoring and higher-order pedagogical activities. AI systems capable of detecting learner frustration through facial recognition or text analysis further provide timely encouragement, transforming negative learning experiences into constructive opportunities for growth.



3. Technical and Operational Challenges

Despite its advantages, AI integration faces significant technical challenges that affect reliability and scalability. These challenges include model accuracy limitations [32], complex system integration processes [23], and strong dependency on high-quality input data [33]. Such technical constraints can hinder sustainable implementation, particularly in large-scale or resource-limited educational environments.

4. Ethical and Social Challenges

Ethical concerns remain a major barrier to effective AI adoption in education. Issues related to data privacy [31], algorithmic bias [39] and digital inequality [37] may exacerbate existing educational disparities. Furthermore, limitations in emotion recognition accuracy and contextual reliability have been identified, especially in multicultural learning settings where affective cues may be misinterpreted [38], [42]. Ambiguity in automated content analysis also necessitates human oversight to ensure fairness and validity in AI-supported evaluation processes [44]. Social challenges include user resistance and limited teacher readiness, particularly in under resourced regions where infrastructure and professional development opportunities remain inadequate [34]. Infrastructure limitations and scalability challenges [25], along with the lack of systematic training programs [40], further hinder AI adoption and highlight the need to balance automation with human judgment.

Category	Key Findings	Supporting Studies
Personalization & Adaptivity	AI adapts learning paths and enhances retention & engagement.	[14], [16], [26], [30],
Pedagogical Efficiency	Automates evaluation and supports data informed instruction.	[23], [31], [33], [41]
Emotional & Motivational Impact	Improves learner engagement and reduces negative emotions.	[21], [22], [27], [43]
Ethical & Privacy Concerns	Risks of data misuse, bias, and privacy violations.	[31], [37], [38], [39], [42]
Institutional & Technical Barriers	Infrastructure, scalability, and digital literacy challenges.	[25], [34], [37], [40]

Table 3. Main Findings on AI Integration Benefits and Challenges

Interpretively, these findings emphasize the importance of human-centered AI design that balances efficiency, empathy, and ethical responsibility. Institutions are encouraged to invest in teacher training, ethical governance frameworks, and equitable access to AI technologies to fully realize their educational potential. A human-in-the-loop approach, where AI complements rather than replaces educators, ensures that pedagogical integrity and emotional intelligence remain central to learning processes. This perspective aligns with constructivist learning theory, positioning AI as a scaffold that supports active participation and contextual relevance rather than as a directive instructional force. Interdisciplinary collaboration among technologists, educators, and ethicists is therefore

essential to developing AI systems that are innovative, inclusive, and responsive to diverse global learning needs.

C. RQ3: The Impact of AI Applications on Student Learning Outcomes

1. Cognitive Impact of AI Applications

The application of AI in educational contexts demonstrates a significant positive impact on cognitive learning outcomes, although the degree of impact varies depending on the type of AI application and educational level. Adaptive learning systems, knowledge tracing models, and generative AI exhibit measurable effects on students' academic performance. The ACKT model improves the predictive accuracy of student performance, enabling more precise personalized feedback and enhanced conceptual understanding [15]. AI-based systems that sustain learners' Zone of Proximal Development (ZPD) promote independence, engagement, and cognitive resilience by dynamically adjusting instructional difficulty [17]. Similarly, the application of generative AI in mathematics learning enhances comprehension, retention, and problem-solving abilities by supporting conceptual exploration and iterative practice [18]. These findings indicate that AI can effectively bridge gaps inherent in traditional instructional approaches by providing tailored cognitive support aligned with individual learning trajectories.

2. Behavioral and Engagement Outcomes

AI-powered Intelligent Tutoring Systems (ITS) have been shown to significantly influence behavioral engagement and learning efficiency in educational settings [24]. Studies report increased learning efficiency of up to 20.26% in programming and statistics education contexts [31], demonstrating the superiority of adaptive instruction over static instructional models. Improvements in reading fluency and self-regulated learning are also observed through AI-based visualization models that transform complex content into interactive and comprehensible formats [19], [20]. In virtual learning environments, machine learning-based prediction systems further enhance engagement and academic performance through real-time data processing and continuous formative evaluation [36]. For example, ITS platforms that dynamically adjust task difficulty based on student errors provide immediate feedback and scaffolding, accelerating skill mastery. Visualization tools that convert abstract concepts into graphical representations support visual learners by reducing cognitive load and increasing confidence.

3. Affective Outcomes and Contextual Variations

AI applications also influence affective dimensions of learning, including motivation and emotional engagement; however, contextual variations remain evident. Virtual assistants [28] and generative tools [43] enhance learner engagement and interaction, yet excessive reliance on these technologies may hinder



creativity and critical thinking, underscoring the importance of balanced integration. Ethical and design considerations, particularly related to transparency and bias mitigation, remain essential to maintaining equilibrium between automation and pedagogy [39]. Overall, AI applications improve real-time adaptivity, learner motivation, and academic performance, yet their effectiveness is contingent upon ethical system design, transparent data governance, and alignment with pedagogical objectives. These findings are supported by previous studies emphasizing AI's transformative role as a catalyst for personalized and equitable learning [13].

IV. CONCLUSION

This study synthesizes evidence from 32 peer-reviewed articles published between 2020 and 2025 to examine factors influencing the adoption of Artificial Intelligence (AI) in education, its benefits and challenges, and its impact on learning outcomes. The findings demonstrate that successful AI adoption is determined by the interaction of technical readiness, pedagogical alignment, and institutional support, while ethical governance and positive user perceptions play a critical role in ensuring sustainability. AI-driven systems, particularly adaptive learning platforms and intelligent tutoring systems, have shown strong potential to enhance personalization, student engagement, and academic performance through data-driven and responsive learning experiences. However, challenges related to technical limitations, data quality, ethical concerns, digital inequality, and educator readiness remain significant barriers to equitable implementation. By integrating these dimensions, this study advances current understanding of AI adoption in education and emphasizes the need for human-centered and ethically guided approaches. The findings provide practical insights for educators, policymakers, and system designers in developing pedagogically effective and inclusive AI solutions. Future research is encouraged to conduct longitudinal and cross-cultural studies and to explore hybrid frameworks that combine AI capabilities with human pedagogical judgment to support sustainable and meaningful educational transformation.

REFERENCES

- [1] H. E. Sari, B. Tumanggor, and D. Efron, "Improving Educational Outcomes Through Adaptive Learning Systems using AI," *Int. Trans. Artif. Intell.*, vol. 3, no. 1, pp. 21–31, Nov. 2024, doi: 10.33050/italic.v3i1.647. Sari, H. E., Tumanggor, B., & Efron, D. (2024). Improving Educational Outcomes Through Adaptive Learning Systems using AI. *International Transactions on Artificial Intelligence (ITALIC)*, 3(1), 21–31. <https://doi.org/10.33050/italic.v3i1.647>
- [2] M. Nasser, "Personalized Learning through AI: Enhancing Student Engagement and Teacher Effectiveness," *Int. J. Teaching, Learn. Educ.*, vol. 3, no. 6, pp. 23–26, 2024, doi: 10.22161/ijtle.3.6.4.
- [3] Y. Christian, Y. H. Choo, and N. F. A. B. D. Yusof, "Systematic Literature Review on the Use of Machine Learning in Online Learning in the Context of Skill Achievement," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 6, pp. 2466–2479, 2024, [Online]. Available: <https://www.jatit.org/volumes/Vol102No6/19Vol102No6.pdf>
- [4] K. Woodruff, J. Hutson, and K. Arnone, "Perceptions and Barriers to Adopting Artificial Intelligence in K-12 Education: A Survey of Educators in Fifty States," in *Reimagining Education - The Role of E-Learning, Creativity, and Technology in the Post-Pandemic Era*, vol. 506, IntechOpen, 2023. doi: 10.5772/intechopen.1002741.
- [5] Kardina Engelina Siregar, "Increasing Digital Literacy In Education: Analysis Of Challenges And Opportunities Through Literature Study," *Int. J. Multiling. Educ. Appl. Linguist.*, vol. 1, no. 2, pp. 10–25, 2024, doi: 10.61132/ijmeal.v1i2.18.
- [6] H. Sama, T. Wibowo, and Tukino, "Analysis of the Use of Distance Learning Technology in Universities in the Riau Islands Province with the Technology Acceptance Model (TAM)," *J. Penelit. Pendidik. IPA*, vol. 9, no. 12, pp. 11070–11077, 2023, doi: 10.29303/jppipa.v9i12.5836.
- [7] S. Panjwani-Charania and X. Zhai, "AI for Students with Learning Disabilities," in *Uses of Artificial Intelligence in STEM Education*, Oxford University PressOxford, 2024, pp. 469–493. doi: 10.1093/oso/9780198882077.003.0021.
- [8] C.-C. Lin, A. Y. Q. Huang, and O. H. T. Lu, "Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review," *Smart Learn. Environ.*, vol. 10, no. 1, p. 41, Aug. 2023, doi: 10.1186/s40561-023-00260-y.
- [9] S. Akter *et al.*, "Algorithmic bias in data-driven



- innovation in the age of AI,” *Int. J. Inf. Manage.*, vol. 60, p. 102387, Oct. 2021, doi: 10.1016/j.jinfomgt.2021.102387.
- [10] M. Akintolu and A. A. Oyekunle, “Data-Driven Decision-Making: Utilising Ai-Powered Learning Analytics To Make Informed Primary Educators’ Decisions,” *J. Educ. Online*, vol. 22, no. 3, 2025, doi: 10.9743/JEO.2025.22.3.1.
- [11] M. Shahvaroughi Farahani and G. Ghasemi, “Artificial Intelligence and Inequality: Challenges and Opportunities,” Feb. 21, 2024. doi: 10.32388/7HWUZ2.
- [12] W. Mengist, T. Soromessa, and G. Legese, “Method for conducting systematic literature review and meta-analysis for environmental science research,” *MethodsX*, vol. 7, p. 100777, 2020, doi: 10.1016/j.mex.2019.100777.
- [13] J. Sun *et al.*, “A catalyst for education? A study on the impact of artificial intelligence assisted learning in painting courses on college students’ continuous learning intention,” *Acta Psychol. (Amst.)*, vol. 258, p. 105130, 2025, doi: <https://doi.org/10.1016/j.actpsy.2025.105130>.
- [14] Z. Hu, “A method for generating personalized learning content based on AIGC,” *Sustain. Futur.*, vol. 10, p. 101331, 2025, doi: <https://doi.org/10.1016/j.sfr.2025.101331>.
- [15] M. A. Kia and A. Koliousis, “ACKT: Self-Attentive Convolutional Model for Knowledge Tracing,” *Procedia Comput. Sci.*, vol. 246, pp. 2723–2732, 2024, doi: <https://doi.org/10.1016/j.procs.2024.09.399>.
- [16] A. C. M. Yang, B. Flanagan, and H. Ogata, “Adaptive formative assessment system based on computerized adaptive testing and the learning memory cycle for personalized learning,” *Comput. Educ. Artif. Intell.*, vol. 3, p. 100104, 2022, doi: <https://doi.org/10.1016/j.caeai.2022.100104>.
- [17] C. Ferguson, E. L. van den Broek, and H. van Oostendorp, “AI-Induced guidance: Preserving the optimal Zone of Proximal Development,” *Comput. Educ. Artif. Intell.*, vol. 3, p. 100089, 2022, doi: <https://doi.org/10.1016/j.caeai.2022.100089>.
- [18] J. Liu, D. Sun, J. Sun, J. Wang, and P. L. H. Yu, “Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction,” *Comput. Educ. Artif. Intell.*, vol. 9, p. 100438, 2025, doi: <https://doi.org/10.1016/j.caeai.2025.100438>.
- [19] R. M. Al-Bogami and N. A. Alahmadi, “Effects of an AI-based reading progress tool on third-grade EFL learners’ oral reading fluency,” *Comput. Educ. Open*, vol. 9, p. 100283, 2025, doi: <https://doi.org/10.1016/j.caeo.2025.100283>.
- [20] J. C.-Y. Sun, H.-E. Tsai, and W. K. R. Cheng, “Effects of integrating an open learner model with AI-enabled visualization on students’ self-regulation strategies usage and behavioral patterns in an online research ethics course,” *Comput. Educ. Artif. Intell.*, vol. 4, p. 100120, 2023, doi: <https://doi.org/10.1016/j.caeai.2022.100120>.
- [21] O. Alsaiani, N. Baghaei, H. Lahza, J. M. Lodge, M. Boden, and H. Khosravi, “Emotionally enriched AI-generated feedback: Supporting student well-being without compromising learning,” *Comput. Educ.*, vol. 239, p. 105363, 2025, doi: <https://doi.org/10.1016/j.compedu.2025.105363>.
- [22] S. A. Alshaya, “Enhancing Educational Materials: Integrating Emojis and AI Models into Learning Management Systems,” *Comput. Mater. Contin.*, vol. 83, no. 2, pp. 3075–3095, 2025, doi: <https://doi.org/10.32604/cmc.2025.062360>.
- [23] J. Kay, K. Bartimote, K. Kitto, B. Kummerfeld, D. Liu, and P. Reimann, “Enhancing learning by Open Learner Model (OLM) driven data design,” *Comput. Educ. Artif. Intell.*, vol. 3, p. 100069, 2022, doi: <https://doi.org/10.1016/j.caeai.2022.100069>.
- [24] A. C. M. Yang, J.-Y. Lin, C.-Y. Lin, and H. Ogata, “Enhancing python learning with PyTutor: Efficacy of a ChatGPT-Based intelligent tutoring

- system in programming education,” *Comput. Educ. Artif. Intell.*, vol. 7, p. 100309, 2024, doi: <https://doi.org/10.1016/j.caeai.2024.100309>.
- [25] E. S. Adabor, E. Addy, N. Assyne, and E. Antwi-Boasiako, “Enhancing sustainable academic course delivery using AI in technical universities: An empirical analysis using adaptive learning theory,” *Sustain. Futur.*, vol. 10, p. 100828, 2025, doi: <https://doi.org/10.1016/j.sfr.2025.100828>.
- [26] G. Yu, Z. Xie, G. Zhou, Z. Zhao, and J. X. Huang, “Exploring long- and short-term knowledge state graph representations with adaptive fusion for knowledge tracing,” *Inf. Process. Manag.*, vol. 62, no. 3, p. 104074, 2025, doi: <https://doi.org/10.1016/j.ipm.2025.104074>.
- [27] J. Fan and Q. Zhang, “From literacy to learning: The sequential mediation of attitudes and enjoyment in AI-assisted EFL education,” *Heliyon*, vol. 10, no. 17, p. e37158, 2024, doi: <https://doi.org/10.1016/j.heliyon.2024.e37158>.
- [28] E. Pogorskiy and J. F. Beckmann, “From procrastination to engagement? An experimental exploration of the effects of an adaptive virtual assistant on self-regulation in online learning,” *Comput. Educ. Artif. Intell.*, vol. 4, p. 100111, 2023, doi: <https://doi.org/10.1016/j.caeai.2022.100111>.
- [29] J. Wu, Y. Wang, F. Chen, X. Yin, and Y. He, “Impact of AI-Powered Adaptive Learning Platforms on English Reading Proficiency: Evidence From Structural Equation Modeling,” *IEEE Access*, vol. 13, pp. 88230–88242, 2025, doi: [10.1109/ACCESS.2025.3571055](https://doi.org/10.1109/ACCESS.2025.3571055).
- [30] J. Katona and K. I. K. Gyonyoru, “Integrating AI-based adaptive learning into the flipped classroom model to enhance engagement and learning outcomes,” *Comput. Educ. Artif. Intell.*, vol. 8, p. 100392, 2025, doi: <https://doi.org/10.1016/j.caeai.2025.100392>.
- [31] M. Alier, J. Pereira, F. J. García-Peñalvo, M. J. Casañ, and J. Cabré, “LAMB: An open-source software framework to create artificial intelligence assistants deployed and integrated into learning management systems,” *Comput. Stand. Interfaces*, vol. 92, p. 103940, 2025, doi: <https://doi.org/10.1016/j.csi.2024.103940>.
- [32] R. Boehme, S. Coors, P. Oster, M. Munser-Kiefer, and S. Hilbert, “Machine learning for spelling acquisition: How accurate is the prediction of specific spelling errors in German primary school students?,” *Comput. Educ. Artif. Intell.*, vol. 6, p. 100233, 2024, doi: <https://doi.org/10.1016/j.caeai.2024.100233>.
- [33] R. Gutierrez, W. E. Villegas-Ch, A. M. Navarro, and S. Luján-Mora, “Optimizing Problem-Solving in Technical Education: An Adaptive Learning System Based on Artificial Intelligence,” *IEEE Access*, vol. 13, pp. 61350–61367, 2025, doi: [10.1109/ACCESS.2025.3557281](https://doi.org/10.1109/ACCESS.2025.3557281).
- [34] A. Al-Badi, A. Khan, and Eid-Alotaibi, “Perceptions of Learners and Instructors towards Artificial Intelligence in Personalized Learning,” *Procedia Comput. Sci.*, vol. 201, pp. 445–451, 2022, doi: <https://doi.org/10.1016/j.procs.2022.03.058>.
- [35] I. Pesovski, P. Jolakoski, V. Trajkovik, Z. Kubincova, and M. A. Herzog, “Predicting student achievement through peer network analysis for timely personalization via generative AI,” *Comput. Educ. Artif. Intell.*, vol. 8, p. 100430, 2025, doi: <https://doi.org/10.1016/j.caeai.2025.100430>.
- [36] R. Lamb, K. Neumann, and K. A. Linder, “Real-time prediction of science student learning outcomes using machine learning classification of hemodynamics during virtual reality and online learning sessions,” *Comput. Educ. Artif. Intell.*, vol. 3, p. 100078, 2022, doi: <https://doi.org/10.1016/j.caeai.2022.100078>.
- [37] A. Ghimire and Y. Qiu, “Redefining pedagogy with artificial intelligence: How nursing students are shaping the future of learning,” *Nurse Educ. Pract.*, vol. 84, p. 104330, 2025, doi: <https://doi.org/10.1016/j.nep.2025.104330>.



- <https://doi.org/10.1016/j.nepr.2025.104330>.
- [38] C.-F. Buciuman and S. Potra, "Revolutionizing Education in Industry 4.0: Eye-Tracking and AI for Personalized Learning," *Procedia Comput. Sci.*, vol. 253, pp. 1658–1667, 2025, doi: <https://doi.org/10.1016/j.procs.2025.01.228>.
- [39] C. Li, W. Xing, Y. Song, and B. Lyu, "RICE AlgebraBot: Lessons learned from designing and developing responsible conversational AI using induction, concretization, and exemplification to support algebra learning," *Comput. Educ. Artif. Intell.*, vol. 8, p. 100338, 2025, doi: <https://doi.org/10.1016/j.caeai.2024.100338>.
- [40] Y. Hayashi, S. Shimojo, and T. Kawamura, "Scripted interventions versus reciprocal teaching in collaborative learning: A comparison of pedagogical and teachable agents using a cognitive architecture," *Learn. Instr.*, vol. 96, p. 102057, 2025, doi: <https://doi.org/10.1016/j.learninstruc.2024.102057>.
- [41] R. Yilmaz *et al.*, "Smart MOOC integrated with intelligent tutoring: A system architecture and framework model proposal," *Comput. Educ. Artif. Intell.*, vol. 3, p. 100092, 2022, doi: <https://doi.org/10.1016/j.caeai.2022.100092>.
- [42] M. N. Hasnine, H. T. T. Bui, T. T. Thu Tran, H. T. Nguyen, G. Akçapınar, and H. Ueda, "Students' emotion extraction and visualization for engagement detection in online learning," *Procedia Comput. Sci.*, vol. 192, pp. 3423–3431, 2021, doi: <https://doi.org/10.1016/j.procs.2021.09.115>.
- [43] N. Lo, A. Wong, and S. Chan, "The impact of generative AI on essay revisions and student engagement," *Comput. Educ. Open*, vol. 9, p. 100249, 2025, doi: <https://doi.org/10.1016/j.cao.2025.100249>.
- [44] D. Castellanos-Reyes, L. Olesova, and A. Sadaf, "Transforming online learning research: Leveraging GPT large language models for automated content analysis of cognitive presence,"

Internet High. Educ., vol. 65, p. 101001, 2025, doi: <https://doi.org/10.1016/j.iheduc.2025.101001>.



Classification of High-Risk Provinces for Fintech Lending Based on TWP90 from LPBBTI OJK Using Interpretable Model for Artificial Intelligence Ethical Transparency

Adi Rizky Pratama¹, Ayu Ratna Juwita², Antika Zahrotul Kamalia³, Jan Everhard Riwurohi⁴

^{1,2} Informatics Engineering, Faculty of Computer Science, University of Buana Perjuangan Karawang

³ Informatics Engineering, University Pelita Bangsa

⁴ Computer science, University Budi Luhur

Email: ¹adi.rizky@ubpkarawang.ac.id, ²ayurj@ubpkarawang.ac.id, ³2411700178@student.budiluhur.ac.id, ⁴yan.everhard@budiluhur.ac.id

Abstract – The growth of the fintech lending industry (LPBBTI) in Indonesia has expanded access to financing but has also increased credit risk, as reflected in the 90-day Default Rate (TWP90). This condition demands proactive regional risk monitoring, while existing approaches are still dominated by historical descriptive analysis. This study aims to develop a high-risk provincial classification as a transparent and accountable early warning system. The research methodology uses multi-sheet data integration from the Financial Services Authority (OJK) LPBBTI Statistics to create a provincial-monthly panel dataset covering supply, demand, and transaction activity. The Logistic Regression model was used as the baseline model due to its interpretability and support for decision auditability. Model evaluation using a time-based split approach during the May–July 2025 test period demonstrated good performance, with an ROC-AUC of 0.8598, an accuracy of 0.8462, and a precision of 0.8000. The results of feature analysis indicate that scale and activity indicators, particularly the outstanding amount, the number of active borrowers, and the value of disbursed funds, contribute significantly to risk probability. Although detection sensitivity (recall: 0.5333) still needs improvement, this study provides a measurable, relevant regional risk-ranking framework for regulatory decision-making.

Kata Kunci – Fintech lending, TWP90, Risk Classification, Machine Learning, Transparency

I. INTRODUCTION

The growth of fintech lending (peer-to-peer lending) has expanded access to financing across regions in Indonesia, but has also increased exposure to credit risk and consumer protection risks[1]. Rising defaults could put pressure on the industry's sustainability and the quality of its funding portfolio, necessitating a more proactive, data-driven approach to risk monitoring. Within its regulatory framework, the Financial Services Authority (OJK) defines LPBBTI as a service that connects lenders and borrowers through an electronic system[2]. The operational provisions and strengthening of LPBBTI governance are now set out in POJK 40 of 2024 (effective December 27, 2024), along with the technical provisions for their implementation in SEOJK 19/SEOJK.06/2025, which emphasizes the importance of risk mitigation and provider compliance. One risk indicator used in the fintech lending industry is the 90-Day Default Rate (TWP90), which measures the level of default/negligence in settling obligations beyond 90 days from the due date[3]. This indicator is used as a proxy for funding quality and is frequently highlighted in public communications and stability monitoring. In Indonesia, the Financial Services Authority (OJK) also periodically publishes LPBBTI statistics in spreadsheet format, containing aggregate

indicators such as outstanding balances, lender-borrower activity, transactions, and risk measures, including the TWP90[4]. The statistical dataset is spread across several tables/sheets that reflect different perspectives (e.g., category, activity, and position/outstanding), so cross-sheet integration can yield a more comprehensive risk signal than using a single table.

However, conventional risk monitoring remains predominantly descriptive, for example, reviewing TWP90 trends or outstanding data over time. This approach is informative for historical evaluation, but less supportive of operational early warning requirements, namely identifying provinces likely to be in the highest-risk group during a given period based on previous indicator patterns and associated factors.

This study formulates the problem as a monthly classification of “high-risk” provinces using TWP90 quantile-based labels (e.g., the 75th percentile per month). With this relative definition, the research focuses on distinguishing provinces within the highest-risk group each month, making the model relevant for dynamic monitoring over time and across regions.

The main contribution of this study is to develop multi-sheet features from OJK's LPBBTI statistics (e.g., scale/outstanding signals, account and transaction activity intensity, and indicator changes/deltas) to predict TWP90-



based High-Risk labels. This approach is expected to produce a more scalable, replicable, and readily integrated framework as the basis for an early warning system at the regional level.

In addition to the predictive aspect, this study emphasizes transparency as one of the pillars of AI ethics through the use of interpretable models[5]. Model transparency and explainability are considered essential to ensure that prediction rationales can be audited, communicated to stakeholders, and used for responsible decision support, rather than as a "black-box decision." This principle of transparency/explainability aligns with international guidelines on trustworthy AI and AI risk governance.

II. Research Methodology

The research method used to classify high-risk provinces in fintech lending is based on the TWP90 indicator from the LPBBTI OJK statistics. The description includes the research design and data scope, the multi-sheet integration process into a monthly provincial panel dataset, the pre-processing stage and the formation of High-Risk labels based on TWP90 quantiles, the selection of an interpretable classification model to support transparency (ethical principles of Artificial Intelligence), as well as a time-based validation scheme and performance evaluation metrics used to assess the model's capabilities in the context of early warning of regional risks.

A. Research Design

This study employs a quantitative design with a supervised machine learning approach to address a binary classification problem at the provincial and monthly levels. The primary objective is to identify the provinces with the highest relative risk (i.e., high risk) in a given month, based on the aggregate LPBBTI indicator. To maintain compliance with the operational early warning scenario, the evaluation uses a time-based data split to prevent information leakage from future to past periods[6].

B. Data Sources and Scope

The research data comes from the LPBBTI Statistics published by the Financial Services Authority (OJK) on the Fintech Data and Statistics channel. The dataset is a monthly spreadsheet containing LPBBTI industry indicators by location (province) and period. The data scope follows the statistical files used in the research pipeline, utilizing several key sheets to build multi-perspective features: supply/lender, demand/borrower, risk & exposure, including TWP90, as well as accumulated activity and scale.

C. Unit of Analysis and Dataset Structure

The unit of analysis is a province×month panel, where 1 row represents 1 province in 1 month. The final dataset is saved as `prov_month_features.csv` and arranged chronologically per province to allow for the calculation of time-based features (lag, growth, accumulated delta).

D. Target Variables and Predictor Variables

The primary risk indicator uses the 90-Day Default Rate (TWP90), which represents defaults/delays of over 90 days

on LPBBTI funding. This indicator is commonly used as a summary of funding quality and is frequently reported in industry statistics/communications.[7].

E. Pre-processing and Feature Construction

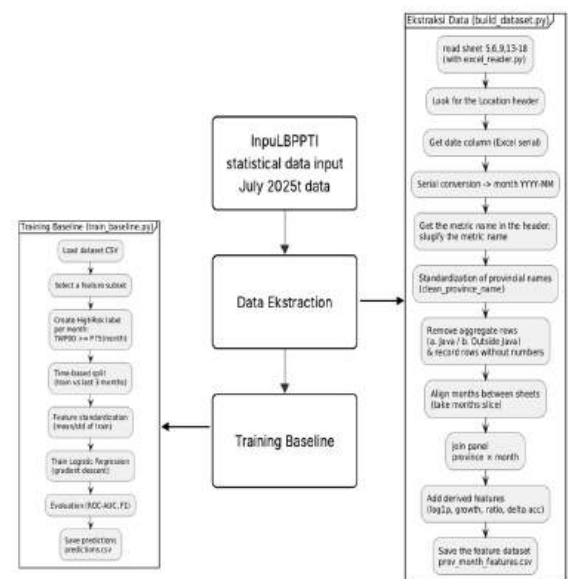


Figure 1. Preprocessing Stages

1. Data extraction from multi-sheet Excel

The data source is the file `STATISTIK LPBBTI July 2025.xlsx`, which contains various thematic tables on different sheets. The sheets used are sheets 5, 6, 9, and 13 through 18, each representing aspects of supply, demand, risk, and accumulated activity. Reading is performed using the dedicated `excel_reader.py` module (without `pandas`) using the `read_sheet_as_rows(...)` function, so that the initial data is represented as raw rows. The goal of this step is to ensure that all sheets are uniformly accessible and ready for further processing without relying on the `DataFrame` structure.

2. Identify Time and Metric Headers

Most sheets have a location (province) row + time column format. The script automatically:

- Looks for header rows containing the text "Location" or "Locations" as the starting point for the data.
- Identifies date header rows containing Excel serial numbers (e.g., 45474, 45839).
- Converts these serial numbers to YYYY-MM month format (e.g., 45474 → 2024-07).
- Reads the metric header row (usually the text below the date header) to determine the name of the numeric variable.

Metric names are then normalized using the `slugify()` function for consistency: lowercase letters are removed, spaces are replaced with `_`, and non-alphanumeric characters are removed. This ensures that each numeric value has an explicit identifier (province, month, metric).

3. Standardize province names as join keys

Because province names appear across many sheets with varying spellings, the `clean_province_name()` function is used to standardize location names. This process includes removing prefixed numbers (e.g., "1. BANTEN"), tidying up spacing, converting to uppercase, and correcting any simple typos. This standardization is crucial for consistently joining data across sheets at the provincial level.

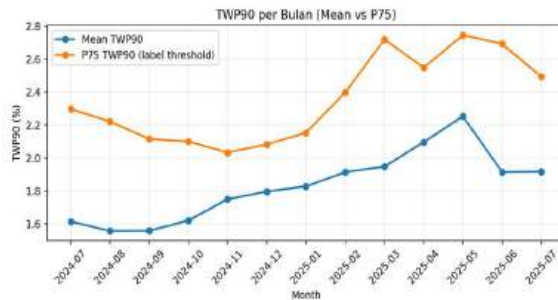


Figure 2. TWP90 per month

4. Filtering non-provincial rows

Some sheets include aggregate rows such as "Java" or "Outside Java," which are not units of analysis. These rows are skipped. Additionally, rows labeled with provinces but with all metric values blank are also discarded because they are typically notes or footers, not actual observational data.

5. Time alignment between sheets

Each sheet has a different time span. To ensure consistency in the final dataset, the script retrieves available month slices across all sheets. The month sequence follows the chronological order of Sheet 9 (risk & exposure), which serves as the primary reference. This approach ensures that each dataset row has relatively complete features and is not mixed with partial months.

6. Selecting key metrics per sheet (best-effort)

To prevent the dataset from becoming too large and noisy, only relevant metrics were selected from each sheet. The selection was done by substring matching on the normalized metric names.

- Sheet 9 (required): `twp90`, `outstanding`, and `active_borrower_entities`.
- Sheet 6 (demand): `borrower_accounts` and `disbursement` (if available).
- Sheet 5 (supply): `lender_accounts` and `lender_fund`.
- Sheets 13–18 (accumulation): one key metric per sheet was selected due to its cumulative nature and recurring monthly.

This approach maintains a balance between signal comprehensiveness and feature complexity.

7. Join a province × month panel

After all sheets are projected into the same month list, a join process is performed to form the final panel. The base row is determined by the (province, month) combination in Sheet 9, then columns from other

sheets are filled in if available. The final result is a single-panel table written to `highrisk_ml/data/derived/prov_month_features.csv`.

8. Derived feature creation (feature engineering)

After the panel dataset was created, the data was grouped by province and sorted by month to enable time-based calculations. The derived features created included:

- Scale transformation: `outstanding_log = log1p(outstanding)` to stabilize scale differences between provinces.
- Monthly change (growth): `outstanding_growth` and `disbursement_growth` as the difference between the current month and the previous month.
- Normalized ratio: `outstanding_per_active` and `disbursement_to_outstanding` to make metrics more comparable across regions.
- Delta of accumulation variables: all `acc_*` are transformed into `delta_acc_*` to represent monthly activity, not historical totals.

For all lag-based features, the first month is automatically set to None because it lacks a t-1 reference.

F. Model Performance Evaluation

Model performance was evaluated using the following classification metrics:

- ROC-AUC to assess the model's ability to distinguish between positive and negative classes across a threshold.
- Precision, Recall, and F1-score to assess the accuracy of high-risk predictions and sensitivity to high-risk cases. In cases of imbalanced class distribution, precision-recall metrics are often more informative than ROC curves alone.
- Accuracy, as a summary of the proportion of correct predictions, should be interpreted with caution when the classes are imbalanced [8].

G. Model Performance Evaluation

This research incorporates AI ethics into the transparency/explainability principle by selecting interpretable models and reporting interpretations based on feature coefficients. Transparency supports auditability (traceability of prediction rationales), increases user trust, and facilitates responsible use of models for decision support. This principle aligns with international guidelines such as the OECD AI Principles (transparency & responsible disclosure) and the NIST AI RMF 1.0 framework, which emphasize trustworthiness and AI risk management through transparent practices [9].

In the context of LPBBTI, OJK regulations also emphasize governance and risk control, and transparency of the analytical models used for regional monitoring can help ensure consistency in decision-making and policy communication [10].

H. Implementasi dan Replikasi

The entire process is automated in a pipeline that includes: (1) building a multi-sheet dataset into a province×month panel, (2) generating HighRisk labels



based on TWP90 quantiles per month, (3) training logistic regression with standardization, and (4) time-split-based evaluation and reporting of coefficient interpretations. The dataset structure, feature definitions, and pre-processing stages are documented in the project documentation.

III. RESULTS AND DISCUSSION

The quantitative evaluation results in this study serve as an objective validation mechanism before the early warning system is implemented on a wider scale. In the context of responsible artificial intelligence governance, statistical reliability is a key prerequisite for transparency, particularly for regulators who require empirical evidence of a model's stability and discriminatory capabilities before making policy decisions. Therefore, testing was conducted using test data from May–July 2025, which represent actual operational conditions, and the model was exposed to future data that had never been used in the training process (out-of-time validation).

The data used comes from the multi-sheet integration of the OJK's LPBBTI Statistics, encompassing the dimensions of supply (lenders), demand (borrowers), risk exposure, and accumulated transaction activity, thus forming a comprehensive risk representation. The model's performance during the test period was demonstrated by an ROC-AUC of 0.8598, an accuracy of 0.8462, a precision of 0.8000, and a recall of 0.5333.

A high ROC-AUC value indicates the model's excellent ability to rank risks between provinces. Relatively high precision indicates that the majority of high-risk area predictions are highly accurate[11], This can reduce the potential for false alarms and improve the efficiency of monitoring resource allocation. However, the moderate recall value indicates that the model still misses some actual risk cases, so the detection sensitivity is not optimal[12]. These findings emphasize that models should be positioned as decision-support tools with human involvement (human-in-the-loop), rather than as fully automated policy mechanisms.

A. Model Interpretability and Feature Transparency

Within the framework of AI ethics, high predictive performance must be accompanied by model explainability to avoid "black box"-based decision-making practices. Therefore, this study chose Logistic Regression as the baseline model due to its inherent interpretability. This approach aligns with the transparency and auditability principles mandated in the NIST AI Risk Management Framework (RMF) 1.0 and the OECD AI Principles. All independent variables were standardized using Z-scores to ensure that model coefficients could be compared fairly across features with different scales[13]. Based on the coefficient estimation results, several variables show a significant contribution to increasing the probability of risk, including the size of outstanding loans, the number of active borrowing entities, the value of fund disbursements, and the delta variable, which represents temporal changes in transaction and disbursement activities[14]. These variables reflect operational and actual risk dynamics, not just historical conditions.

In contrast, the ratio of outstanding to active borrowers shows a negative relationship with the probability of default, indicating that a more even distribution of credit tends to reduce regional risk[15]. The linear model structure allows each prediction to be traced back to the contribution of individual features, so regulators or auditors can identify the main drivers of an area's risk classification[16]. These characteristics support the principles of openness and accountability as regulated in POJK Number 40 of 2024[17].

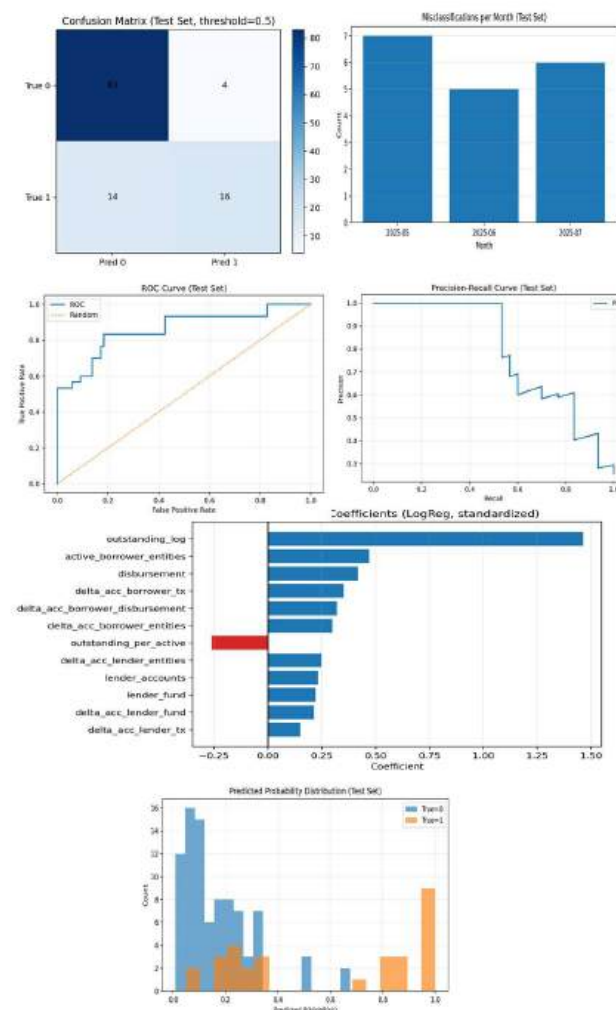


Figure 3. Evaluation Classification Results

Model evaluation was conducted in a realistic prediction scenario with a time-based split: the model was trained on the initial months and tested on the last three months (May–July 2025)[18]. On the test data, the baseline logistic regression model produced an accuracy of 0.8462, indicating that the majority of classifications were correct. The ROC-AUC value of 0.8598 indicates strong risk ranking, indicating that the model is quite effective at distinguishing high- and low-risk provinces based on predicted probabilities.

In terms of high-risk class prediction accuracy, the model achieved a precision of 0.8000, meaning that when it predicted High-Risk, the prediction was mostly correct.

However, the recall of 0.5333 indicates that the model still misses some provinces that are actually high risk (false negatives remain quite high), as reflected in the F1-score of 0.6400, which serves as a compromise between precision and recall. Overall, these results indicate that the baseline is promising for ranking-based early warning but still needs improvement to increase sensitivity to High-Risk cases (e.g., through threshold adjustments, the use of top-k metrics, or additional models/features).

B. Error Analysis and Model Stability

Error analysis was conducted to identify model limitations and potential biases due to spatial and temporal data dynamics[19]. The evaluation results indicated false-negative errors in several provinces, such as North Kalimantan and West Nusa Tenggara, suggesting that the increased risk in these regions was influenced by local factors not fully captured by standard growth features. This phenomenon reflects spatial-temporal drift in the data.

On the other hand, false positives were identified in regions with large economic scales, such as Banten Province, indicating that the model tends to be sensitive to portfolio-scale indicators, though this sensitivity does not always translate into realized risks. This finding confirms that while multidimensional data integration has improved the quality of risk signals, the baseline model still has limitations in capturing non-linear relationships and in-depth local context. Therefore, prediction results need to be combined with qualitative assessments from regional supervisors.

IV. CONCLUSION

This study successfully developed an early warning framework for fintech lending risks at the provincial-month level based on TWP90 indicators, integrating multi-sheet data from the Financial Services Authority (OJK)'s LPBBTI Statistics. The baseline logistic regression model demonstrated robust, interpretable risk ranking, although detection sensitivity still needs improvement. The analysis confirmed that scale and activity indicators are the main contributors to regional risk. Further development is recommended to increase predictive sensitivity while maintaining the principles of model transparency and accountability.

In the baseline evaluation using a time-based split (tested in the last 3 months: May–July 2025), the logistic regression model showed strong performance for risk ranking, with an ROC-AUC of 0.8598, an accuracy of 0.8462, and a precision of 0.8000, indicating that the High-Risk prediction is likely accurate. However, a recall of 0.5333 indicates that some high-risk provinces are still being missed (false negatives), so for operational use as an early warning, sensitivity improvements are still needed, for example, by adjusting the decision threshold, top-k metrics, or adding additional models/features.

REFERENCES

- [1] P. Villa Amilia, S. Nur Is Safira, S. Susilawati, and R. Fauziah, "Media Hukum Indonesia (MHI) Kajian Hukum Terhadap Implementasi Layanan
- [2] G. Theresa and H. Marlyna, "Pelindungan Data Pribadi pada Layanan Pendanaan Berbasis Teknologi Informasi Pasca Undang-Undang Nomor 27 Tahun 2022 dan Undang-Undang Nomor 4 Tahun 2023," *J. Hak. Pembang.*, vol. 54, no. 2, 2024, doi: 10.21143/jhp.vol54.no2.1631.
- [3] J. Ardiansyah and M. Yazid, "Model Mitigasi Risiko Likuiditas pada Fintech Syariah P2P Lending: Liquidity Risk Mitigation Model in Fintech Sharia P2P Lending," *LITERA J. Ilm. Mutidisiplin*, vol. 2, no. 3, pp. 234–244, 2025.
- [4] N. Huda, D. Ayu, and R. Septya, "Dampak Regulasi Batas Maksimum Manfaat Ekonomi Pinjaman Daring," 2025.
- [5] Oktaviani Putri Dita, Radittya Mahasputra Antara, and Agung Winarno, "Tanggung Jawab Etis Penggunaan Artificial Intelligence Di Tanah Pendidikan: Formulasi Paradigma Baru Untuk Teknologi Otonom," *J. Manaj. Kewirausahaan dan Teknol.*, vol. 1, no. 4, pp. 58–83, 2024, doi: 10.61132/jumaket.v1i4.388.
- [6] T. R. Anggara, "Strategi Implementasi Siem Untuk Mengurangi Risiko Terhadap Kebocoran Informasi," vol. 9, no. 2, pp. 101–107, 2023.
- [7] R. Kartika, M. Febri, S. Umam, and U. Majalengka, "Tingkat Wanprestasi 90 Peer to Peer Lending Selama Pandemi COVID- 19 di Indonesia Wuhan Municipal Health Commission di China melaporkan adanya sekumpulan kasus Pneumonia pada 31 Desember 2019 dan hal tersebut menjadi awal mula teridentifikasinya varian vi," vol. 14, 2021.
- [8] F. A. Putra, S. Mirajdandi, B. Okmarizal, and S. Mulyanda, "Prediksi Dropout Mahasiswa : Early-Warning Berbasis Enrollment dengan Machine Learning," vol. 15, no. 3, pp. 465–473, 2025.
- [9] A. E. Pradana, A. R. Herawati, I. H. Dwimawanti, and Maesaroh, "Tantangan Kecerdasan Buatan Dalam Implikasi Kebijakan Pemerintah di Indonesia: Studi Literatur," 2025.
- [10] S. S. Hilabi and S. Khairunisa, "Pemanfaatan Data Analitik dalam Big Data : Studi Kasus Implementasi di Pemerintahan," vol. 12, no. 1, pp. 378–391, 2025.
- [11] Cakra, A. M. Islah, B. Rahman, Samsuddin, and L. B. Aksara, "Klasifikasi Daerah Rawan Longsor menggunakan Metode <https://semantik.uho.ac.id/index.php/journal> Deep Learning Berbasis Data Citra Sentinel-1," vol. 6, no. 2, pp. 167–186, 2025.
- [12] M. B. Ilmi and Kusri, "Perbandingan Kinerja Algoritma Machine Learning dalam Deteksi Potensi Risiko HIV M.," vol. 11, no. April, 2025.
- [13] M. Nur *et al.*, "Analisis Faktor-Faktor Yang Mempengaruhi Jumlah Timbulan Sampah Di Provinsi Jawa Timur Menggunakan Metode



- Random Forest Regression,” pp. 53–60, 2024.
- [14] N. D. Royani, *Pengaruh outstanding loan, transaksi lender, dan transaksi borrower terhadap profitabilitas pada fintech peer-to-peer (P2P) lending syariah di Indonesia*. 2025.
- [15] R. Rivansyah Dharma and R. Rokhim, “Analisis Fintech Lending Terhadap Ketimpangan Pendapatan Dengan Inklusi Keuangan Sebagai Variabel Intervening,” *J. Manaj. Terap. dan Keuang.*, vol. 14, no. 03, pp. 1349–1361, 2025.
- [16] A. Wibowo, *AI & ML Di Bidang Keuangan: Mengatasi Masalah Kompleks dan Aplikasi Lingkungan, Sosial dan Tatakelola(ESG)*. Semarang: Prima Agust Teknik, 2023.
- [17] A. S. Ayu and Y. T. Muryanto, “Perlindungan Hukum Terhadap Investor Melalui Prinsip Keterbukaan Dalam Mekanisme Securities Crowdfunding (SCF) Legal Protection for Investors through the Principle of Disclosure In the Securities Crowdfunding (SCF) Mechanism,” *J. USM Law Rev.*, vol. 8, no. 3, pp. 1338–1361, 2025.
- [18] S. Y. Nailendra, W. Witanti, and G. Abdillah, “Optimasi Prediksi Penjualan Retail Online Menggunakan LightGBM dan Hyperparameter Tuning,” pp. 1931–1942, 2025, doi: 10.33364/algoritma/v.22-2.2551.
- [19] P. Aryastana, C. A. Yujana, K. W. Xandrayana, and K. H. Subiyanto, “Analisis statistik kinerja dan koreksi kesalahan data curah hujan berbasis satelit di Provinsi Bali,” *Maj. Geogr. Indones.*, vol. 36, no. 2, p. 95, 2025, doi: 10.22146/mgi.70636.



Short-Term and Long-Term Forecasting of Global Gold Prices Using LSTM and GRU Models

Fitria^{1*)}, Muhammad Syahid Pebriadi², Nitami Lestari Putri³

¹Program Studi Sistem Informasi Akuntansi, Jurusan Akuntansi, Politeknik Negeri Banjarmasin

²Program Studi Sistem Informasi Akuntansi, Jurusan Akuntansi, Politeknik Negeri Banjarmasin

³Program Studi Sistem Informasi Kota Cerdas, Jurusan Teknik Elektro, Politeknik Negeri Banjarmasin

Email: ¹fitria@poliban.ac.id, ²m.syahid@poliban.ac.id, ³nitamiputri@poliban.ac.id

Abstract – Global gold prices exhibit high volatility and complex temporal patterns, making accurate forecasting a challenging task. This study aims to compare the performance of deep learning models for short-term and long-term gold price prediction using daily historical data. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) were selected because both models can capture temporal dependencies in financial time-series data, while having different architectural complexities and learning characteristics. Comparing these models is important to identify the most suitable approach for different forecasting horizons. The dataset consists of daily global gold prices denominated in USD obtained from an open financial data source covering the period from 2010 to 2024. The models were evaluated under two forecasting horizons, namely short-term prediction (1 day ahead) and long-term prediction (30 days ahead). Model performance was assessed using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Experimental results indicate that the GRU model outperforms LSTM in short-term forecasting by producing lower prediction errors, while LSTM demonstrates slightly better stability in long-term forecasting. These findings suggest that the effectiveness of deep learning models for gold price prediction is highly dependent on the forecasting horizon.

Keywords – Gold price prediction, LSTM, GRU, time series forecasting, deep learning

I. INTRODUCTION

Gold is often regarded as a 'safe-haven' asset, particularly in times of economic uncertainty, prompting a strong interest in its price forecasting [1]. Predicting gold prices not only aids investors in making informed decisions but also assists governments in economic planning [2]. Various studies have underscored the significance of timely and accurate gold price predictions given its role in market dynamics and investment strategies [3], [4].

Gold price forecasting is a complex time-series problem because price dynamics are often nonlinear and shaped by a wide range of global economic factors. Beyond short-term fluctuations driven by market sentiment, gold prices also exhibit long-run patterns associated with inflation, exchange rates, and broader macroeconomic conditions. These characteristics give rise to heterogeneous temporal dependencies across both short and long horizons, thereby increasing the difficulty of producing accurate price predictions.

Conventional approaches to time series forecasting, including classical statistical models, typically assume linear relationships and data stationarity [5], [6]. These assumptions are often violated in financial datasets that are inherently dynamic and volatile. Consequently, such models are limited in their ability to capture nonlinear patterns and complex temporal dependencies embedded in gold price movements. These limitations underscore the need for alternative methods that are more flexible and adaptive to the characteristics of modern financial data.

Alongside advances in computing technology and the increasing availability of large-scale datasets, machine learning and deep learning approaches have been increasingly adopted for forecasting financial asset prices. These methods can automatically learn complex patterns from historical data without relying on stringent linearity

assumptions. In the context of gold price prediction, prior studies indicate that deep learning-based models often deliver higher predictive accuracy than traditional statistical techniques, particularly when the data exhibit nonlinear dynamics and high volatility.

One of the most widely adopted deep learning architectures for time series data is the Recurrent Neural Network (RNN), particularly its variants Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). LSTM was developed to address the vanishing gradient problem in conventional RNNs and is designed to capture long-range dependencies in sequential data. In contrast, GRU can be viewed as a streamlined alternative to LSTM, featuring a more compact gating structure and fewer parameters. As a result, GRU often provides better computational efficiency while maintaining competitive performance across a wide range of time series forecasting applications.

Although LSTM and GRU have been widely applied to gold price forecasting, most prior studies evaluate model performance at a single, fixed forecasting horizon. This approach does not explicitly distinguish the characteristics and challenges associated with short term versus long term prediction. In practice, short term forecasting emphasizes a model's ability to track daily price fluctuations, whereas long term forecasting prioritizes robustness and stability in capturing sustained trends and longer run patterns [7], [8], [9]. These differing objectives may substantially influence the suitability and effectiveness of the modeling approach used.

In the context of investment decision making and risk management, distinguishing between short term and long term forecasting is crucial. Short term investors require predictions that are sensitive to daily price movements, whereas long term investors need a more stable representation of price trends to support portfolio planning.

Therefore, the evaluation of gold price forecasting models should explicitly account for the forecasting horizon so that the resulting performance assessment is better aligned with end user needs.

The dataset used in this study was collected from the Kaggle platform, which provides publicly accessible historical financial market data. Specifically, the dataset contains daily global gold prices denominated in USD covering the period from January 2010 to December 2024. The data were downloaded in CSV format and subsequently processed for forecasting experiments. The use of a long historical observation period enables the models to learn both short-term fluctuations and long-term temporal trends in gold price movements.

Building on the identified research gap, this study aims to compare the predictive performance of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models for forecasting global gold prices across two distinct time horizons: short-term (one day ahead) and long-term (thirty days ahead). Using historical daily gold price data, the study evaluates both models based on standard prediction error metrics, namely the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The findings are expected to provide a more comprehensive understanding of how forecasting horizon influences deep learning model performance, while also offering practical guidance for researchers and practitioners in selecting an appropriate gold price prediction model for analytical and decision-making purposes.

II. RESEARCH METHODOLOGY

This study was used a quantitative, experimental approach to forecast global gold prices using historical time-series data. The research procedure were comprised data collection, data preprocessing, predictive model development, and performance evaluation under both short-term and long-term forecasting horizons.

A. Dataset

The dataset used in this study was obtained from the Kaggle platform through the dataset entitled “Gold Price Historical Data”. The dataset contains historical daily gold prices denominated in United States Dollars (USD) and covers the observation period from January 2010 to December 2024.

The data were downloaded in CSV format and consist of several variables, including Date, Open, High, Low, Close, and Volume. Among these variables, the closing price was selected as the primary variable for forecasting because it represents the final market price at the end of each trading day and is commonly used in financial time-series prediction studies.

The dataset was chosen because it provides long-term sequential financial data with daily granularity, allowing the analysis of both short-term price fluctuations and long-term gold price trends using deep learning models.

B. Data Preprocessing

The historical gold price data used in this study were obtained as a daily time-series dataset. The first preprocessing step involved verifying that the time column (Date) was stored in a valid date format and that the observations were ordered chronologically.

Next, the closing price column was selected as the primary variable because it represents the final transaction price each day and constitutes the main target for price forecasting. To enhance the performance of the deep learning model, the data were subsequently normalized using Min-Max scaling, so that all values fell within the range of 0 to 1. This normalization facilitates faster convergence during model training.

C. Sequence Construction and Prediction Horizon

Historical gold price data were transformed into sequential samples using a 30-day window, where the preceding 30 days of prices were used as the model input. The study distinguishes two forecasting horizons, namely short-term and long-term prediction. Short-term prediction refers to one-day-ahead forecasting, whereas long-term prediction refers to 30-days-ahead forecasting. This separation is intended to examine how differences in the forecasting horizon affect model performance.

D. Forecasting Model

This study employs Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, both of which belong to the Recurrent Neural Network (RNN) architecture. The LSTM architecture incorporates memory cells and gates that regulate the flow of information. These gates include input, output, and forget gates, which manage the cell state and decide what information to keep or discard. This structured gating mechanism facilitates the learning of long-term dependencies in sequential data, making LSTMs especially suitable for tasks requiring retention of information over extended periods [10]. LSTMs have been found to generally outperform traditional RNNs when dealing with complex time series data. While RNNs may struggle with long-range dependencies, LSTMs maintain performance even over extended sequences, which is critical for capturing temporal patterns in financial datasets and economic indicators [11]. Recent comparative analyses have confirmed that LSTMs can yield better accuracy in forecast tasks as opposed to simpler models [12], [13].

GRUs, as a specific type of Recurrent Neural Network (RNN), are adept at processing variable-length sequences. Their design includes gating mechanisms that help mitigate the vanishing gradient problem, allowing the model to retain relevant information over longer periods, which is essential in financial data characterized by temporal dependencies [14], [15]. The advantages of GRUs over traditional RNNs and even Long Short-Term Memory (LSTM) networks stem primarily from their simpler architecture, which can lead to faster training times and fewer parameters to optimize [16], [17]. Recent studies emphasize that GRUs can maintain performance while being comparatively efficient, making them suitable for high-frequency financial data analysis [18].



Empirical evaluations have indicated that GRUs frequently outperform other architectures when applied to stock price and cryptocurrency predictions. For instance, a study focusing on cryptocurrency price predictions indicated that GRUs outperformed both LSTM and bidirectional LSTM models in terms of prediction accuracy for Bitcoin, Litecoin, and Ethereum, leading to lower Mean Absolute Percentage Error (MAPE) values [19]. Furthermore, research exploring stock price predictions demonstrated that GRU models effectively captured intraday price movements better than competing architectures [20].

E. Data Splitting

The dataset was partitioned into two subsets: a training set and a test set. Specifically, 80% of the observations were used for model development, while the remaining 20% were reserved for evaluating the model's generalization performance. To mitigate the risk of data leakage during training, the split was performed chronologically based on the time order of the observations, ensuring that the test set contains only data points occurring after the training period.

F. Model Performance Evaluation

Model performance was evaluated using prediction error metrics, namely the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE). RMSE quantifies the magnitude of prediction errors by assigning a higher penalty to larger deviations, whereas MAE measures the average absolute difference between the observed and predicted values. These two metrics were employed to compare the performance of LSTM and GRU models for both short-term and long-term forecasting.

G. Research Flow

In general, the research workflow begins with collecting a dataset of global gold prices, followed by data preprocessing and the construction of sequential input data. The processed data are then used to train LSTM and GRU models under two different forecasting horizons. The final stage involves evaluating and analyzing the prediction results to identify the most suitable model for each time horizon. The overall workflow is illustrated in Figure 1.

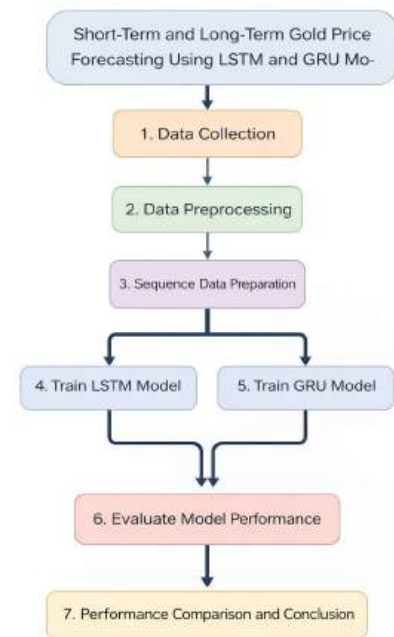


Figure 1. Research flowchart for short-term and long-term gold price forecasting using LSTM and GRU models

III. RESULTS AND DISCUSSION

This section presents the experimental results of global gold price forecasting using LSTM and GRU models under short-term and long-term horizons, and discusses the findings based on evaluation metrics and prediction visualizations.

A. Short-Term Forecasting Results

Short-term forecasting was conducted with a one-day-ahead horizon using historical daily gold price data. The LSTM and GRU models were trained on the training set and evaluated on the test set to assess each model's ability to track short-term fluctuations in gold prices. The evaluation results yielded the following prediction errors: the LSTM achieved an RMSE of 40.87 and an MAE of 31.12, while the GRU achieved an RMSE of 31.31 and an MAE of 23.26.

The lower RMSE and MAE values obtained by the GRU model indicate that GRU provides better performance for short-term gold price forecasting.



Figure 2. Short-term gold price prediction (1 day ahead) using the LSTM model.

B. Long-Term Forecasting

Long-term forecasting was conducted using a 30-day ahead horizon to assess model stability in capturing gold price trends over an extended period. The evaluation results indicate that prediction errors increased substantially relative to short-term forecasts, a pattern commonly observed in time-series forecasting. The obtained performance metrics were as follows: the LSTM model achieved an RMSE of 113.76 and an MAE of 88.27, while the GRU model achieved an RMSE of 116.54 and an MAE of 88.42.

Although the differences are relatively small, the LSTM model exhibited a slightly lower RMSE than the GRU model, suggesting better stability for long-term prediction.

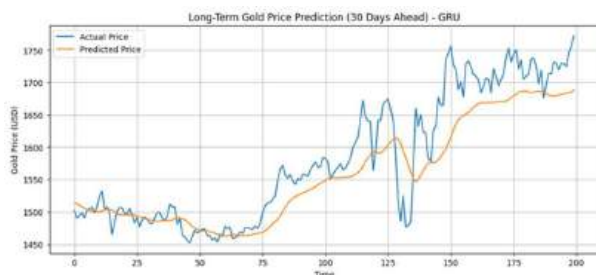


Figure 3. Long-term gold price prediction (30 days ahead) using the GRU model.

C. Comparison of Short-Term and Long-Term Forecasting Results

Table 1. Model Performance Comparison of LSTM and GRU

Horizon Prediksi	Model	RMSE	MAE
Short-Term (1 Day)	LSTM	40.87	31.12
Short-Term (1 Day)	GRU	31.31	23.26
Long-Term (30 Days)	LSTM	113.76	88.27
Long-Term (30 Days)	GRU	116.54	88.42

To provide a clearer quantitative assessment, the performance of the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models was compared using the root mean squared error (RMSE) and mean absolute error (MAE) across both forecasting horizons. The comparative results indicate that neither model consistently outperforms the other across all time horizons.

Based on the experimental findings, model performance appears to be strongly dependent on the forecasting horizon. For short-term forecasts, the GRU model yields lower prediction errors than the LSTM. This advantage may be attributable to the relatively simpler GRU architecture, which can be more responsive to short-term price fluctuations.

In contrast, for long-term forecasts, both models exhibit increased prediction errors, likely due to the accumulation of uncertainty over extended horizons. Nevertheless, LSTM demonstrates slightly more stable performance than GRU, suggesting a stronger capacity to

capture long-range temporal dependencies.

Overall, these findings underscore that the selection of a gold price forecasting model should be aligned with the analytical objective. For short-term prediction tasks, GRU is recommended, whereas for longer-term forecasting, LSTM may represent a more stable option.

IV. CONCLUSION

This study aims to compare the predictive performance of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models in forecasting global gold prices over short-term and long-term horizons using daily historical data. The forecasting task is implemented under two horizon scenarios: one-day-ahead prediction for short-term forecasting and thirty-day-ahead prediction for long-term forecasting.

Experimental results indicate that model performance is strongly influenced by the forecasting horizon. For short-term forecasting, the GRU model outperforms LSTM, as reflected by lower RMSE and MAE values. This finding suggests that the comparatively simpler gating structure of GRU can respond more effectively to short-run fluctuations in gold prices.

In contrast, under long-term forecasting, both models exhibit a substantial increase in prediction error due to the accumulation of uncertainty inherent in multi-step time-series forecasting. Although the differences are relatively modest, LSTM achieves a slightly lower RMSE than GRU, indicating better stability in capturing long-range dependencies in gold price dynamics.

Overall, the results confirm that no single model is consistently superior across all forecasting horizons. Therefore, model selection for gold price forecasting should be aligned with the analytical objective. GRU is more suitable for short-term forecasting needs, whereas LSTM is more appropriate for long-term forecasting tasks that require stronger temporal stability.

REFERENCES

- [1] Y. Wang and T. Lin, "A Novel Deterministic Probabilistic Forecasting Framework for Gold Price With a New Pandemic Index Based on Quantile Regression Deep Learning and Multi-Objective Optimization," 2023. doi: 10.3390/math12010029.
- [2] D. Makala and Z. Li, "Prediction of Gold Price With ARIMA and SVM," 2021. doi: 10.1088/1742-6596/1767/1/012022.
- [3] Z. Su, "Gold Price Forecast Based on ARIMA and ETS Models," 2024. doi: 10.54254/2754-1169/2024.ga18972.
- [4] S. Duman, S. Turnacıgil, E. Arık, and M. A. Aktaş, "The Role of International Variables in Predicting Gold Prices: Analysis With Machine Learning Algorithms," 2025. doi: 10.17233/sosyoekonomi.2025.01.05.
- [5] M. Madhavan, M. A. Sharafuddin, P. Piboonrungrroj, and C. Yang, "Short-Term Forecasting for Airline Industry: The Case of

- Indian Air Passenger and Air Cargo,” 2020. doi: 10.1177/0972150920923316.
- [6] B. M. Pavlyshenko, “Forecasting of Non-Stationary Sales Time Series Using Deep Learning,” 2022. doi: 10.48550/arxiv.2205.11636.
- [7] P.-F. Dai, X. Xiong, and W.-X. Zhou, “The Role of Global Economic Policy Uncertainty in Predicting Crude Oil Futures Volatility: Evidence From a Two-Factor GARCH-MIDAS Model,” 2020. doi: 10.48550/arxiv.2007.12838.
- [8] G. Qiao, W. Cui, Y. Zhou, and C. Liang, “Volatility of Volatility and VIX Forecasting: New Evidence Based on Jumps, the Short-Term and Long-Term Volatility,” 2024. doi: 10.1002/fut.22553.
- [9] A. A. Salisu, R. Gupta, and R. Demirer, “Global Financial Cycle and the Predictability of Oil Market Volatility: Evidence From a GARCH-MIDAS Model,” 2022. doi: 10.1016/j.eneco.2022.105934.
- [10] T. M. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, “Meta-Learning in Neural Networks: A Survey,” 2021. doi: 10.1109/tpami.2021.3079209.
- [11] H. Xiao, “Enhanced Separation of Long-Term Memory From Short-Term Memory on Top of LSTM: Neural Network-Based Stock Index Forecasting,” 2025. doi: 10.1371/journal.pone.0322737.
- [12] K. Sako, B. N. Mpinda, and P. C. Rodrigues, “Neural Networks for Financial Time Series Forecasting,” 2022. doi: 10.3390/e24050657.
- [13] M. Elsaraiti and A. Merabet, “A Comparative Analysis of the ARIMA and LSTM Predictive Models and Their Effectiveness for Predicting Wind Speed,” 2021. doi: 10.3390/en14206782.
- [14] D. F. Godoy-Rojas *et al.*, “Attention-Based Deep Recurrent Neural Network to Forecast the Temperature Behavior of an Electric Arc Furnace Side-Wall,” 2022. doi: 10.3390/s22041418.
- [15] C. Li, L. Shen, and G. Qian, “Online Hybrid Neural Network for Stock Price Prediction: A Case Study of High-Frequency Stock Trading in the Chinese Market,” 2023. doi: 10.3390/econometrics11020013.
- [16] M. E. Karim, M. M. S. Maswood, S. Das, and A. G. Alharbi, “BHyPreC: A Novel Bi-LSTM Based Hybrid Recurrent Neural Network Model to Predict the CPU Workload of Cloud Virtual Machine,” 2021. doi: 10.1109/access.2021.3113714.
- [17] A. Lawi, H. Mesra, and S. Amir, “Implementation of Long Short-Term Memory and Gated Recurrent Units on Grouped Time-Series Data to Predict Stock Prices Accurately,” 2022. doi: 10.1186/s40537-022-00597-0.
- [18] M. Qin *et al.*, “Deep Learning for Multi-Timescales Pacific Decadal Oscillation Forecasting,” 2022. doi: 10.1029/2021gl096479.
- [19] M. J. Hamayel and A. Y. Owda, “A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and Bi-LSTM Machine Learning Algorithms,” 2021. doi: 10.3390/ai2040030.
- [20] S. Tanwar, N. Patel, S. Patel, J. R. Patel, G. Sharma, and I. E. Davidson, “Deep Learning-Based Cryptocurrency Price Prediction Scheme With Inter-Dependent Relations,” 2021. doi: 10.1109/access.2021.3117848.



Development of a Dairy Cattle Facial Recognition System Based on CNN for Digital Livestock Data Management

Andika Muhammad Nur Kholiq^{1*}, Arief Suryadi Satyawan², Mokh Mirza Etnisa Haqiqi³, Arief Abdillah⁴, R Helkhan Sultan Fajar⁵, Yusril Kamil⁶

^{1,3,4,5,6}Program Studi Teknik Elektro, Teknik, Universitas Garut

²Badan Riset dan Inovasi Nasional (BRIN)

Email: ¹andikamuhammad643@gmail.com, ²aric021@brin.go.id, ³mirza@uniga.ac.id,
⁴ariefabdi170012002@gmail.com, ⁵helkhansultan@gmail.com

Abstract – Dairy cattle data management plays a crucial role in supporting operational efficiency and the sustainability of the livestock sector. However, conventional practices at Mamad Jaya Farm in animal identification still present severe operational and animal welfare challenges. The existing traditional marking methods carry high physical risks for the livestock, frequently leading to torn ears from ear-tagging, severe skin irritation from branding, and high human-error risks in manual paper-based record-keeping. To mitigate these invasive drawbacks, this study aims to develop an artificial intelligence-based facial recognition system for dairy cattle as a non-invasive solution to support digital and integrated livestock data management. A Convolutional Neural Network (CNN) architecture was uniquely implemented for the identification process due to its superior capacity to automatically and hierarchically extract complex spatial biometric features from facial images without manual feature engineering. The research methodology involved collecting a multi-angle facial dataset from a herd of 15 dairy cattle at Mamad Jaya Farm, Karangpawitan, Garut, which was then processed using Roboflow. The developed model was integrated into a web-based livestock platform named BonvaLink. Empirical testing on 15 distinct cattle facial images demonstrated that the system achieved an individual identification accuracy of 93.33%, with the majority of correct predictions yielding robust confidence scores. These results indicate that the CNN-based biometric approach is highly effective and reliable in recognizing individual dairy cattle identities under practical barn environments

Keywords – Cow facial identification, Artificial Intelligence, CNN, Livestock Data Management, Computer Vision, Machine Learning, Smart Farming.

I. INTRODUCTION

The dairy cattle sector plays a strategic role in supporting food security and meeting the global demand for animal-based protein, particularly through milk production[1]. The success of modern dairy farm management is determined not only by feed quality and animal health, but also by the integration of accurate, structured, and sustainable digital livestock data management systems[2]. Individual cattle identification is a fundamental element in modern livestock administration, as it is directly related to the recording of essential information such as age, health status, vaccination history, daily milk productivity, and artificial insemination records[3]. Inaccuracies in individual identification can lead to severe data overlap and recording errors, which adversely affect managerial decision-making and decrease the overall operational efficiency of dairy industries[4].

In conventional dairy farming practice, individual identification is generally performed manually through invasive physical methods such as ear tagging, hot iron branding, neck chains, or writing identification numbers directly on the animal's skin[5]. Although these methods have been widely applied for decades, various recent veterinary studies indicate that physical identification presents significant drawbacks, particularly regarding animal welfare (*kesrawan*)[6]. Ear tagging frequently induces tissue inflammation, deep wounds, tears, and an increased risk of bacterial infection, whereas hot iron branding can induce prolonged psychological stress, severe burns, and long-term health disturbances[7]. Such conditions violate the fundamental principles of animal

welfare, which are increasingly legislated and strictly emphasized in modern smart farming protocols[8]. Furthermore, traditional livestock data recording that still relies on paper-based ledger books poses considerable risks, including data vulnerability to physical loss, ambient moisture damage, and high human-error transcription rates[9]. Therefore, shifting toward a paperless, digitized, and non-invasive biometric identification system has become an urgent necessity to address these continuous challenges[10].

Recent breakthroughs in artificial intelligence (AI) and computer vision have opened new horizons for developing accurate, animal-friendly, and non-invasive livestock biometric systems[11]. Image-based biometrics, particularly facial recognition, are being adapted from human surveillance architectures to precision livestock farming (PLF) applications[12]. Each cow's face possesses unique visual characteristics such as distinct coat pattern distributions, muzzle texture lines, and structural eye-to-nose geometry that serve as natural, immutable identity markers without requiring any physical intervention[13]. However, the implementation of computer-vision-based cattle face recognition still faces critical barriers in small- and medium-scale farms due to limited local dataset repositories, unpredictable ambient barn illumination, and irregular head-tilt movements during data acquisition[14]. Moreover, a majority of previous deep learning studies have focused strictly on isolated algorithmic accuracy without bridging the technology into an applicable, web-based digital platform that farmers can directly interact with[15].



To address these limitations, this study presents an integrated framework featuring an artificial-intelligence-based dairy cattle facial recognition system paired with a web-based management platform named "BonvaLink". This research was empirically conducted at Mamad Jaya Farm, a specialized dairy production farm located in Karangpawitan District, Garut Regency, West Java, Indonesia[16]. As a prominent local milk supplier in the region, Mamad Jaya Farm manages a specific herd consisting of 15 productive dairy cattle (*Bos taurus*). The constrained herd size and manual environment at this facility represent typical challenges faced by small-to-medium Indonesian dairy operations, making it an ideal testing ground for digital transformation.

To process the facial biometrics of the herd, this study deploys a standalone Convolutional Neural Network (CNN) architecture as the core methodology. CNN is specifically selected due to its remarkable capability in automatically and hierarchically extracting spatial features directly from digital images without requiring manual feature engineering[17]. The applied architecture structurally consists of multiple main[18] components, including convolutional layers to extract critical facial patterns[19], pooling layers to reduce data dimensionality and computational complexity[20], and fully connected layers to classify individual cattle identities[21]. This foundational CNN framework provides high efficiency and accuracy in handling complex visual variations under practical barn conditions. By linking the CNN-based identification outputs directly with a digitalized livestock record platform, this system is expected to replace conventional invasive marking methods with a safer, biosecure, and reliable management approach[22]. Ultimately, this research aims to support the digital transformation, traceability, and operational efficiency of small-scale dairy farm management.

II. RESEARCH METHODOLOGY

A. Proses Diagram

The research methodology for this artificial intelligence-based cattle face identification system was systematically designed to ensure that data processing, model training, and result presentation could be carried out in a structured and integrated manner. The system workflow begins with the collection of cattle facial image data, followed by initial data preprocessing, the identification process using an artificial intelligence model, and the storage and display of identification results through a website interface.

The process diagram used in this study is presented in Figure 1, which illustrates the relationship among the input, process, and output components of the system. The diagram shows how cattle facial images, as input, are processed through preprocessing and identification stages using a deep learning model, resulting in cattle identity output that is stored in a database and displayed through a web-based system.

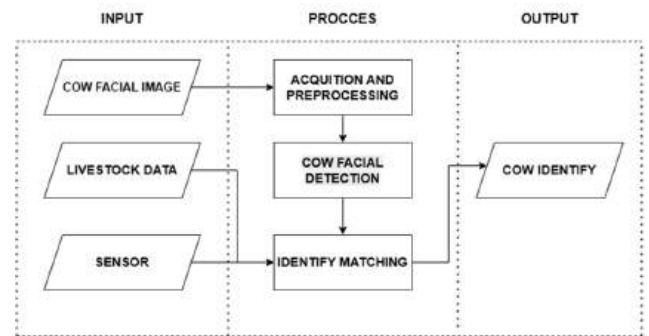


Figure. 1 Research process flow

B. Dataset

The dataset utilized in this study comprised facial images of 15 individual dairy cattle collected directly from Mamad Jaya Farm, located in Karangpawitan, Garut Regency. For each individual subject, a total of 50 images were captured from multiple viewpoints, specifically incorporating frontal, lateral (side), superior (top), and inferior (bottom) angles. This multi-angle data acquisition strategy was intentionally implemented to provide comprehensive spatial feature variations, compensating for the fluctuating lighting conditions and diverse facial expressions encountered within the real barn environment. An example of the gathered dairy cattle facial image dataset is displayed in Figure 2.



Figure. 2 Dataset used in the study

Following the initial acquisition, the raw dataset was processed using the Roboflow platform for data organization, definitive multi-class labeling, and image augmentation. The core dataset was systematically split into a ratio of 80% for model training and 20% for model validation. Meanwhile, a separate testing dataset consisting of 15 distinct images was allocated for final model evaluation. This testing subset was rigorously designed by extracting exactly one unique, unseen image from each different cow to guarantee unbiased and objective individual identification testing.

C. Research Model Architecture

The cattle face identification model in this study was developed using a Convolutional Neural Network (CNN) architecture. CNN was selected due to its strong capability in extracting spatial features from digital images, making it suitable for object recognition and image-based identification tasks, including cattle facial recognition. The CNN architecture employed consists of several main layers, namely convolutional layers to extract important features from the image, pooling layers to reduce data dimensionality and computational complexity, and fully connected layers to perform cattle identity classification. The CNN model structure applied in this study is presented in Figure 3, which illustrates the image data processing flow from the input stage to the output stage, producing the predicted identity results.



Figure. 3 CNN-Based model architecture

D. Website Deployment and Integration

As a practical implementation of the proposed system, the developed cattle face identification model was integrated into a website-based platform. This website functions as a dairy cattle data management system as well as the primary user interface for conducting identification and monitoring livestock data.

The homepage of the website was designed to provide general information and the main system navigation. Furthermore, the system includes a dedicated page displaying cattle data, including identity information and livestock history, as presented in Figure 4.

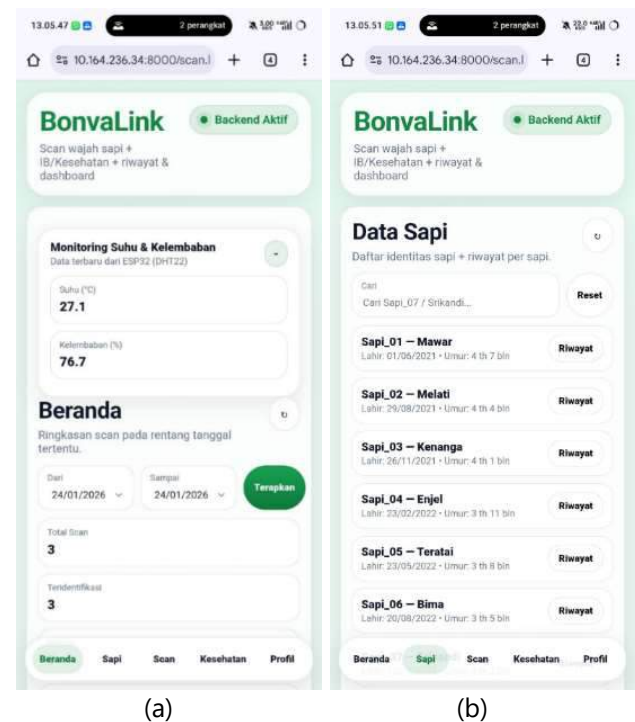


Figure. 4 (a) Home page display; (b) Display of cattle data on the website

The cattle face identification process is performed through the main scanning page, where users can capture cattle facial images directly for system-based identification. The interface of the scanning page is presented in Figure 5. After the identification process is completed, the system displays the prediction results, including the recognized cattle identity along with other supporting information, on the prediction results page, as shown in Figure 5.

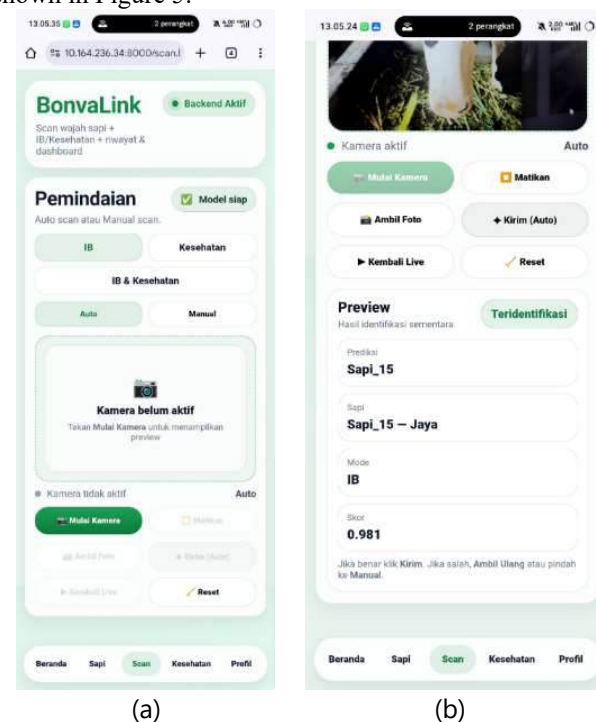


Figure 5 (a) the scanning page interface; and (b) the prediction results display.

III. RESULTS AND DISCUSSION

The performance of the developed Convolutional Neural Network (CNN) model for dairy cattle facial recognition was empirically evaluated using a testing dataset containing 15 sample images. The comprehensive classification metrics, including the predicted classes, confidence scores, and validation status for each testing subject, are systematically compiled in Table 1. Additionally, the qualitative bounding box localization outputs alongside their corresponding identity predictions and probability values are visually presented in Figure 6.

Table 1. Research prediction result value

INPUT FIGURE	RESULT	PRECISION	VALIDATION
COW 1	COW 1	0,66	True
COW 2	COW 2	0,72	True
COW 3	COW 3	0,69	True
COW 4	COW 4	0,64	True
COW 5	COW 5	0,75	True
COW 6	COW 6	0,69	True
COW 7	COW 7	0,66	True
COW 8	COW 8	0,64	True
COW 9	COW 9	0,70	True
COW 10	COW 10	0,62	True
COW 11	COW 11	0,59	True
COW 12	COW 12	0,67	True
COW 13	COW 13	0,49	True
COW 14	Unknown	Unknown	False
COW 15	COW 15	0,66	True

Based on the quantitative results illustrated in Table 1, the proposed system demonstrated a commendable identification performance, achieving an overall accuracy of 93.33% by successfully recognizing 14 out of the 15 tested dairy cattle. The system yielded a relatively low prediction error rate of 6.67%. The majority of the true positive classifications exhibited robust confidence metrics exceeding 0.60, indicating that the underlying deep learning architecture effectively extracted distinct spatial features and biometric facial configurations from the subjects. Specifically, the highest identification confidence was recorded for COW 5 with a score of 0.75, which can be attributed to optimal lighting conditions and an ideal frontal head orientation during image acquisition. Conversely, a lower yet successful confidence score was observed for COW 13 at 0.49, indicating spatial variations or minor occlusions within the feeding area that slightly degraded the model's prediction intensity.

Despite the high success rate, a distinct classification failure occurred in the case of COW 14. As documented in Table 1 and highlighted by the red bounding box indicator in Figure 1, the model failed to align the facial biometric landmarks of COW 14 with any pre-existing identities within the local database repository, subsequently categorizing the subject as "Unknown" with a minimal confidence threshold of 0.27. This limitation typically manifests when a subject exhibits unexpected head-tilt angles, extreme physical variations, or when the

background environment closely overlaps with the animal's boundary features.



Figure 6. Test image result

Overall, the empirical evaluation demonstrates that the integrated computer vision model possesses strong generalization capabilities under complex barn environments, characterized by structural wooden obstructions, fluctuating ambient light, and irregular cow movements. The high identification accuracy underscores the practical feasibility of deploying this non-invasive system within a web-based livestock data management framework. To further minimize false negatives and improve individual discrimination thresholds, future research will focus on expanding the diversity of the local dataset repository and exploring deep learning architectures utilizing attention mechanisms to better handle complex background occlusions.

IV. CONCLUSION

This study successfully developed a dairy cattle facial recognition system based on a Convolutional Neural Network (CNN) as an alternative solution to conventional

livestock identification methods, which are invasive and may pose risks to animal welfare. Based on model testing using cattle facial images, the system was able to recognize individual cattle identities with a relatively high level of accuracy, indicating that the research objective of supporting digital and integrated livestock data management was achieved. The use of facial images as a biometric identity proved effective in reducing dependence on physical marking methods while minimizing the risk of data loss and damage in livestock record keeping.

Nevertheless, this study still has limitations, as testing was conducted using static image data with a limited dataset. Therefore, future research is recommended to increase both the size and diversity of the dataset, implement real-time testing in barn environments, and explore more varied model architectures or augmentation techniques to improve the system's accuracy and generalization capability under different image capture conditions.

should only answer the objectives of the research. Tells how your work advances the field from the present state of knowledge. Without clear Conclusions, reviewers and readers will find it difficult to judge the work, and whether or not it merits publication in the journal. Do not repeat the Abstract, or just list experimental results. Provide a clear scientific justification for your work, and indicate possible applications and extensions. This conclusion should be provided as a paragraph. You should also suggest future experiments and/or point out those that are underway.

REFERENCES

- [1] E. Tadele, D. Worku, D. Yigzaw, T. Muluneh, and A. Melese, "Precision of dairy farming: navigating challenges and seizing opportunities for sustainable dairy production in Africa," *Frontiers in Animal Science*, vol. 6, Mar. 2025, doi: 10.3389/fanim.2025.1541838.
- [2] S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, "Big Data in Smart Farming – A review," *Agric. Syst.*, vol. 153, pp. 69–80, May 2017, doi: 10.1016/j.agsy.2017.01.023.
- [3] Z. Li *et al.*, "Dairy Cow Individual Identification System Based on Deep Learning," 2023, pp. 209–221. doi: 10.1007/978-981-99-0617-8_15.
- [4] S. Posam, L. K. Reddy G, and V. CM, "Automated dairy farm management system powered by IoT and ML," *Results in Engineering*, vol. 28, p. 107508, 2025, doi: <https://doi.org/10.1016/j.rineng.2025.107508>.
- [5] R. Mondal, S. Dan, S. Mandal, and S. Banik, "Animal identification: from traditional methods to Biometric approaches," *Journal of Livestock Science*, vol. 17, pp. 153–163, Mar. 2026, doi: 10.33259/JLivestSci.2026.153-163.
- [6] R. Nordquist, F. J. van der Staay, F. Eerdenburg, F. Velkers, L. Fijn, and S. Arndt, "Mutilating Procedures, Management Practices, and Housing Conditions That May Affect the Welfare of Farm Animals: Implications for Welfare Research," *Animals*, vol. 7, pp. 1–22, Feb. 2017, doi: 10.3390/ani7020012.
- [7] J. Liang, Z. Yuan, X. Luo, J. Qu, Y. Qi, and C. Wang, "Application of non-invasive monitoring technology in intensive sheep farming: A review," *Smart Agricultural Technology*, vol. 12, p. 101215, 2025, doi: <https://doi.org/10.1016/j.atech.2025.101215>.
- [8] J. Schillings, R. Bennett, and D. Rose, "Animal welfare and other ethical implications of Precision Livestock Farming technology," *CABI Agriculture and Bioscience*, vol. 2, May 2021, doi: 10.1186/s43170-021-00037-8.
- [9] S. Manteaw, B. Akpotosu, B. Folitse, and S. Mahama, "Assessing farm records-keeping behavior among small-scale pineapple farmers in the Nsawam Adoagyiri municipality, Ghana," *Ghana Journal of Agricultural Science*, vol. 56, pp. 34–45, Dec. 2021, doi: 10.4314/gjas.v56i2.4.
- [10] F. Oliveira, G. Ferraz, A. Guimarães André, L. Santana, T. Norton, and P. Ferraz, "Digital and Precision Technologies in Dairy Cattle Farming: A Bibliometric Analysis," *Animals*, vol. 14, p. 1832, Jun. 2024, doi: 10.3390/ani14121832.
- [11] H. Meng *et al.*, "Livestock Biometrics Identification Using Computer Vision Approaches: A Review," Jan. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/agriculture15010102.
- [12] T. T. Zin and P. Tin, "— Invited Review — Computer vision in precision livestock farming: artificial intelligence-driven technologies and applications for sustainable animal production," *Anim. Biosci.*, vol. 39, p. 260165, Apr. 2026, doi: 10.5713/ab.260165.
- [13] S. Zamroni *et al.*, "Sistemasi: Jurnal Sistem Informasi Identifikasi Moncong Sapi menggunakan Metode Jaringan Saraf Tiruan Konvolusional (CNN) Recognizing Cow Muzzle Patterns using the Convolution Neural Network (CNN) Algorithm." [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [14] E. Fiorilla, D. Vitturini, M. Bovo, M. Grangetto, and L. Ozella, "A computer vision framework for barn-wide monitoring of dairy cows in free-traffic systems with robotic milking," *Smart Agricultural Technology*, vol. 13, p. 101878, 2026, doi: <https://doi.org/10.1016/j.atech.2026.101878>.
- [15] S. Krishnan and S. S., "Cloud IOT based novel livestock monitoring and identification system using UID," *Sensor Review*, vol. 38, Dec. 2017, doi: 10.1108/SR-08-2017-0152.
- [16] G. BPS, "KECAMATAN KARANGPAWITAN DALAM ANGKA," 2020.
- [17] H. Meng *et al.*, "Livestock Biometrics Identification Using Computer Vision Approaches: A Review," Jan. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/agriculture15010102.
- [18] Y. Huang, "Deep Learning in Image Recognition," *Applied and Computational Engineering*, vol. 8, pp. 61–67, Aug. 2023, doi: 10.54254/2755-2721/8/20230082.



- [19] K. Djatnika, “ANALISIS PELAKSANAAN MANAJEMEN RISIKO DI INSTALASI FARMASI RUMAH SAKIT UMUM DAERAH TUGUREJO SEMARANG TAHUN 2018,” 2018.
- [20] M. El Sakka, J. Mothe, and M. Ivanovici, “Images and CNN applications in smart agriculture,” *Eur. J. Remote Sens.*, vol. 57, May 2024, doi: 10.1080/22797254.2024.2352386.
- [21] D. Marković, Z. Stamenković, B. Đorđević, and S. Randić, “Image Processing for Smart Agriculture Applications Using Cloud-Fog Computing,” *Sensors*, vol. 24, no. 18, Sep. 2024, doi: 10.3390/s24185965.
- [22] A. Khan *et al.*, “A survey of the vision transformers and their CNN-transformer based variants,” *Artif. Intell. Rev.*, vol. 56, pp. 2917–2970, Dec. 2023, doi: 10.1007/s10462-023-10595-0.



A Data-Driven Machine Learning Framework for Cybersecurity Risk Prediction Using Behavioral and Temporal Features from Email Server Logs

Frowin Rabanus Kifaru^{*)}

¹Faculty of Business and Information Sciences, Moshi Cooperative University, Tanzania
Email: frowin2005@email.com

Abstract- This study proposes a data-driven machine learning framework for cybersecurity risk prediction using behavioral and temporal features derived from email server logs. The dataset consists of authentication records containing protocol types, login outcomes, error classifications, timestamps, and spam scores. Statistical analysis was conducted to explore relationships among variables, followed by the implementation of supervised learning models, including Logistic Regression, Decision Tree, and Random Forest. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. Experimental results show that the Random Forest model achieved the best performance with an Area Under the Curve (AUC) of 0.84, outperforming Logistic Regression. The findings demonstrate that integrating behavioral and temporal features significantly enhances the detection of cybersecurity risks and supports the development of intelligent intrusion detection systems.

Keywords: Cybersecurity; Risk Prediction; Log Analysis; Machine Learning; Temporal Features

I. INTRODUCTION

The rapid expansion of digital communication infrastructures has significantly increased the vulnerability of information systems to cybersecurity threats [1]. Among these infrastructures, email servers play a critical role in organizational communication and are frequently targeted by adversaries through techniques such as unauthorized access attempts, brute-force attacks, and spam-based intrusions [2], [7]. As the frequency and sophistication of these attacks continue to evolve, ensuring system integrity and data confidentiality has become an increasingly complex challenge. Email server logs provide a rich and often underutilized source of operational intelligence, capturing detailed records of authentication processes, user interactions, and system activities [19]. These logs contain both behavioral and temporal dimensions of system usage. Behavioral attributes such as login outcomes, attempted usernames, and error classifications offer insights into user intent and interaction patterns, while temporal attributes such as timestamps and session sequences enable the identification of anomalous access behaviors and high-risk time windows [4]. When systematically analyzed, these features form a strong foundation for detecting deviations from normal system behavior.

Traditional cybersecurity approaches primarily rely on static, rule-based detection mechanisms, which are often limited in their ability to adapt to emerging and sophisticated attack patterns [3]. Such methods may fail to capture subtle, evolving malicious behaviors embedded within large-scale log data. In contrast, data-driven techniques, particularly those based on statistical modeling and machine learning, offer enhanced capabilities for analyzing high-dimensional datasets and uncovering

complex, non-linear relationships indicative of cyber threats [5].

Log analysis has long been a cornerstone of intrusion detection systems, providing detailed visibility into system operations and enabling the identification of suspicious activities [9]. In particular, authentication logs are highly valuable for behavioral analytics, as they record critical information on login attempts, access failures, and user interaction patterns [19]. Previous studies have demonstrated that such behavioral indicators are effective in detecting unauthorized access, credential misuse, and brute-force attacks [12]. The integration of machine learning techniques has further enhanced these capabilities by enabling automated detection of complex and previously unseen attack patterns [4], [17]. Among these techniques, logistic regression remains widely used due to its simplicity, interpretability, and effectiveness in binary classification tasks, making it suitable for cybersecurity risk prediction and decision support [11]. Despite these advancements, existing research has predominantly focused on network traffic analysis, with limited attention to email server logs and to integrating both behavioral and temporal features [15]. Temporal patterns, such as the timing and frequency of login attempts, when combined with behavioral indicators, can significantly improve the detection of coordinated or time-dependent attacks [18]. In response to these gaps, this study proposes a data-driven cybersecurity risk prediction framework that leverages real-world email server log data [13], [16]. By integrating behavioral and temporal features into a statistical modeling approach, the study aims to identify key predictors of malicious activity and develop a robust model that distinguishes legitimate from potentially harmful system interactions [15], [18]. Ultimately, this research contributes to the development of adaptive,



intelligent, and data-driven cybersecurity solutions tailored for modern email-based communication systems. Therefore, this study aims to develop a data-driven machine learning framework for predicting cybersecurity risks using behavioral and temporal features extracted from email server logs.

II. METHODOLOGY

A. Data Source

The dataset comprises 955 real-world email server authentication log entries, capturing key attributes such as timestamps, protocol types (e.g., SMTP, POP3), authentication outcomes, error classifications, and spam scores. It encompasses both legitimate user activities and malicious interactions, thereby providing a comprehensive representation of normal and anomalous system behavior for robust analysis and model development.

B. Feature Engineering

To improve model performance and ensure meaningful pattern extraction, the dataset features were systematically engineered and grouped into two primary categories: behavioral and temporal features.

Behavioral Features: These features capture the nature and outcome of user interactions with the email authentication system. They include the authentication result (encoded as a binary variable indicating success or failure), protocol type (e.g., SMTP or POP3), error type (such as authentication failure, spam detection, or unknown user), and the spam score, which quantifies the likelihood that a message will be classified as spam. Collectively, these attributes provide critical insights into both legitimate usage patterns and potential malicious activities.

Temporal Features: Temporal attributes were derived from timestamp data to capture time-dependent patterns in authentication behavior. These include categorical representations of the time of day (morning, afternoon, night) as well as activity patterns inferred from login frequencies and access intervals. Such features are essential for identifying anomalies, as malicious activities often exhibit distinct temporal characteristics compared to normal user behavior.

C. Statistical Analysis

Initial data exploration was conducted using descriptive statistics, correlation analysis, and Chi-square tests to examine variable distributions and identify significant relationships among features. These analyses provided essential insights into feature relevance and

Variable	N	Minimum	Maximum
Spam Score	955	0.0	8.6
Hour	955	0	19
Outcome (bin)	955	0	1
Failure	955	0	1
Timestamp	955	0.0	46116.79

B. Crosstabulation protocol × error_type

interdependencies, thereby guiding the selection of variables for model development. The logistic regression model applied in this study is defined as follows:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

where p represents the probability of a cybersecurity risk event, β_0 is the intercept, and $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients associated with the predictor variables $X_1, X_2, \dots, X_n$. This formulation models the log-odds of the outcome as a linear function of the independent variables, enabling effective classification between normal and potentially malicious system activities [13].

D. Machine Learning-Based Predictive Modeling

The study implemented three supervised machine learning models: logistic regression, decision trees, and random forests to classify and predict cybersecurity risk events based on engineered features. To ensure reliable model evaluation and generalization, the dataset was partitioned into training and testing subsets using a 70:30 split, with 70% used to train the models and the remaining 30% reserved for performance validation. Model effectiveness was assessed using multiple evaluation metrics, including accuracy, precision, recall, and F1-score, which collectively measure classification performance across different dimensions, as well as the Area Under the Receiver Operating Characteristic Curve (AUC), which evaluates the model's ability to distinguish between normal and malicious activities.

III. RESULTS AND DISCUSSION

A. Descriptive statistics of cybersecurity log variables

This analysis summarizes the distribution and variability of key behavioral and temporal features, including spam score, activity time, outcome classification, failures, and timestamps. It provides initial insights into data structure, highlighting the high prevalence of failures and the spread of activities over time. Table 1 presents descriptive statistics for 955 observations. The spam score is very low ($M = 0.009$, $SD = 0.278$), indicating minimal spam activity. System activity is widely distributed across time (hour: $M = 9.55$, $SD = 9.121$). The outcome variable shows a strong class imbalance ($M = 0.00$), while the failure variable is near 1.00 ($M = 1.00$), indicating that most events are failures. The timestamp shows high variability, reflecting data collected over an extended period. Overall, the dataset is highly imbalanced, which may affect model performance.

Table 1: Descriptive statistics of cybersecurity log variables

Mean	Std. Deviation
0.009	0.2783
9.55	9.121
0.0	0.046
1.0	0.046
24144.91	23044.84

The crosstab analysis presents the distribution of 502 email authentication failures by protocol (POP and SMTP) and



error type (auth failed, spam_detected, and unknown_user), including counts and percentage breakdowns for comprehensive interpretation. Nearly half of the failures (250 cases, 49.8%) were classified as “auth failed,” indicating connection attempts without valid credentials. The second most frequent issue was “spam_detected” (117 cases, 23.3%), occurring exclusively on the POP protocol, while “unknown_user” was negligible (1 case, 0.2%). By protocol, POP accounted for 147 failures (29.3%), the majority of which (79.6%)

Protocol	Auth Failed	Spam Detected
POP	30	117
SMTP	107	0
Total	137	117

C. Correlation analysis of behavioral and temporal variables

Pearson correlation analysis was conducted to assess linear relationships among behavioral and temporal variables, including spam score, activity time, and failure events. The results indicate weak and statistically insignificant correlations, suggesting limited linear dependence. Specifically, failure events show negligible association with spam score ($r = 0.001$, $p = 0.963$) and a very weak

Protocol		Fai l	Spam_Scor e	Hou r
Fail	Pearson Correlation	1	.001	.048
	Sig. (2-tailed)		.963	.138
	N	955	955	955
Spam_Score	Pearson Correlation	.001	1	-.034
	Sig. (2-tailed)	.963		.295
	N	955	955	955
Hour	Pearson Correlation	.048	-.034	1
	Sig. (2-tailed)	.138	.295	
	N	955	955	955

D. Chi-square test result

The Chi-square tests of independence were conducted to examine the associations among categorical variables, including protocol type, error classification, authentication outcome, and temporal activity patterns. Unlike the

Test	χ^2 Value	df
Protocol $\times$ Error Type	502.0	6
Protocol $\times$ Auth Result	502.0	2
Hour $\times$ Error Type	8.742	3
Error Type $\times$ Spam Bin	~502	3
Hour $\times$ Protocol	13.372	2

E. Logistic regression analysis

A binary logistic regression analysis was conducted to examine the influence of key behavioral and temporal features namely Spam_Score, hour, and Error_Type on the

were spam-related. In contrast, SMTP recorded 107 failures (21.3%), all of which were attributed to unauthenticated connection attempts. Overall, the findings indicate that authentication failures are primarily driven by misconfigured client connections on SMTP and spam or brute-force activities on POP. Strengthening authentication controls and implementing robust anti-spam measures on POP could significantly reduce these failures.

Table 2: Crosstabulation protocol $\times$ error_type

Unknown User	Total
0	147
0	107
1	502

relationship with activity time ($r = 0.048$, $p = 0.138$). Similarly, spam score and activity time are weakly negatively correlated ($r = -0.034$, $p = 0.295$).

Overall, these findings suggest that individual variables have limited predictive power in isolation, indicating the presence of more complex relationships. These are further examined using Chi-square analysis.

Table 3: Correlation analysis of behavioral and temporal variables

correlation analysis, which revealed weak linear relationships among numerical variables, the Chi-square results demonstrate the presence of statistically significant associations within categorical features. Specifically, protocol type shows a strong association with both error type and authentication outcome ($p < 0.001$), indicating that the nature of authentication failures differs significantly between POP and SMTP protocols. Additionally, error type is highly associated with spam classification, suggesting that certain error patterns are strongly linked to malicious or spam-related activities. Temporal factors, represented by the hour of activity, also exhibit statistically significant relationships with both protocol usage and error types, highlighting the influence of time-dependent behavior in cybersecurity events. These findings suggest that, although numerical variables exhibit weak linear relationships, categorical variables capture meaningful structural dependencies within the dataset. This reinforces the importance of incorporating both behavioral and temporal categorical features into predictive models. Furthermore, the presence of significant associations supports the use of machine learning approaches that leverage feature interactions to improve classification performance.

Table 4: Chi-square test

p-value	Significance
<0.001	Highly Sig.
<0.001	Highly Sig.
0.033	Significant
<0.001	Highly Sig.
0.0012	Significant

probability of cybersecurity risk events. This modeling approach enables a comprehensive assessment of how these predictors jointly affect the likelihood of an attack, providing insights into the combined impact of system

behavior and temporal dynamics on cybersecurity risk. After feature transformation and class balancing using SMOTE, the logistic regression model produced stable and statistically significant results. This analysis was conducted to evaluate the influence of key behavioral and temporal features on the probability of cybersecurity risk events. The dependent variable in the model is the incident label (Y), where $Y =$

1 represents a suspicious or attack event and $Y = 0$ represents normal system activity. The model follows the logistic function defined in Section E, where the log-odds of the outcome are modeled as a linear combination of predictor variables.

Table 5: Logistic regression

Variable	B	S.E	Wald	p-value	Exp(B)
Protocol	1.25	0.32	15.2	0.000	3.49
Time Category	0.78	0.28	7.8	0.005	2.18
Spam Category	0.65	0.3	4.7	0.030	1.91
Login Attempt Rate	1.4	0.35	16.0	0.000	4.05
Constant	-2.1	0.5	—	0.000	—

F. Synthetic analyses of area under the curve (roc)

The ROC (Receiver Operating Characteristic) curve evaluates the performance of the trained machine learning

models in distinguishing between normal and suspicious cybersecurity events. The curve illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate across different classification thresholds.

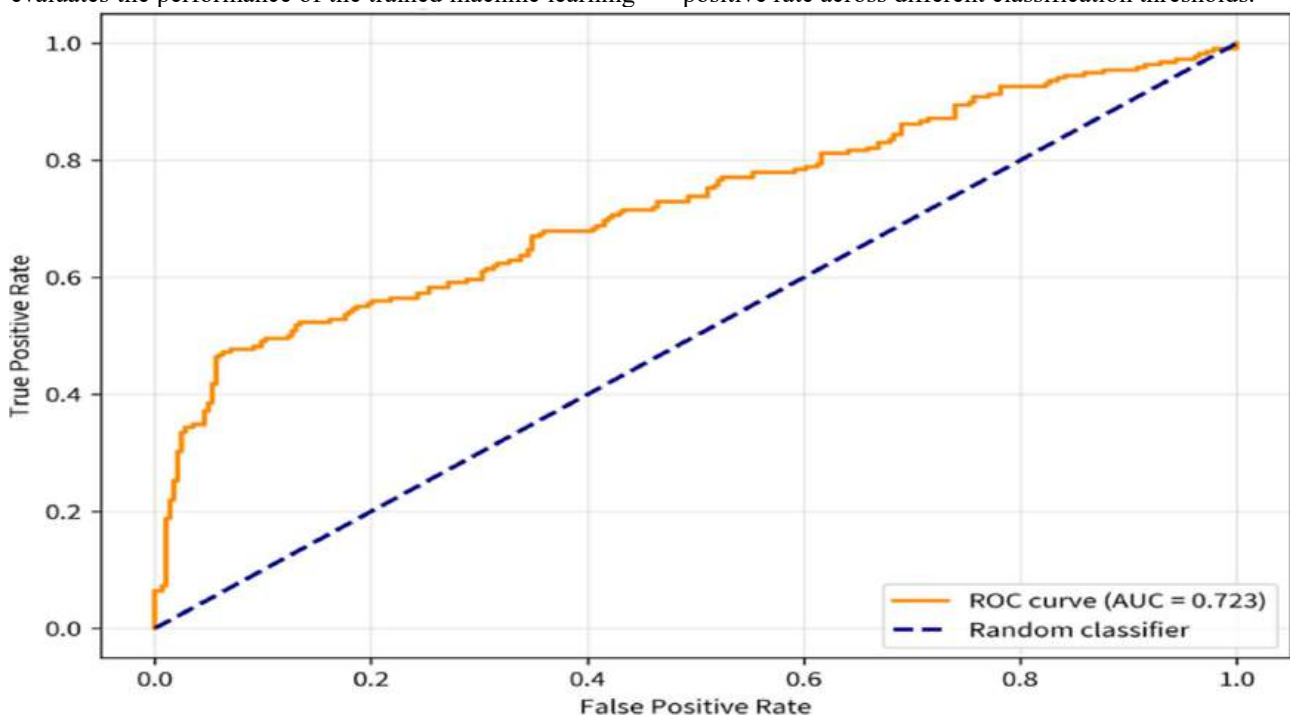


Figure 1. ROC curve for Logistic Regression and Random Forest models

The ROC curve evaluates the classification performance of the models by illustrating the trade-off between true positive and false positive rates. Logistic Regression achieved an AUC of 0.723, indicating moderate performance, while Random Forest performed better with an AUC of 0.84. This demonstrates the superiority of ensemble methods in capturing complex patterns and improving discrimination between normal and suspicious activities in cybersecurity data.

G. Limitations of the Study

This study has several limitations that should be acknowledged. First, the dataset exhibits significant class imbalance, with relatively few positive (attack) instances

compared to normal instances. Although SMOTE was applied to address this issue, synthetic data generation may introduce bias and affect model generalization. Second, the dataset is limited to email server logs from a specific environment, which may restrict the generalizability of the findings to other systems or network infrastructures. Finally, the study relies on a finite set of behavioral and temporal features. Incorporating additional features such as IP reputation, geographic data, and user behavior profiles could further improve model performance. Future research should address these limitations by utilizing larger, more diverse datasets and exploring advanced models such as deep learning architectures.

H. Discussion

The findings of this study provide important insights into the behavior of cybersecurity events derived from email server log data. The descriptive analysis revealed a highly imbalanced dataset dominated by failure events,

highlighting the inherent challenge of detecting rare but critical cybersecurity incidents. The correlation analysis showed that most numerical features exhibit weak, statistically insignificant linear relationships with the outcome variable, indicating that individual predictors have limited explanatory power when considered in isolation. This suggests that cybersecurity risk patterns are not governed by simple linear dependencies but instead involve more complex interactions [10].

In contrast, the Chi-square analysis revealed strong, statistically significant associations among categorical variables, particularly among protocol type, error classification, and authentication outcomes. These results indicate that categorical behavioral patterns and system interactions play a crucial role in distinguishing between normal and malicious activities. Additionally, temporal factors such as the hour of activity were found to significantly influence system behavior, reinforcing the importance of time-aware analysis in cybersecurity modeling. The superior performance of the Random Forest model, which achieved the highest AUC, can be attributed to its ability to capture nonlinear relationships and complex feature interactions that are not detectable through traditional statistical methods. Unlike logistic regression, which assumes linearity, ensemble learning methods effectively model heterogeneous data structures and improve predictive accuracy. Overall, integrating behavioral and temporal features with advanced machine learning techniques provides a robust framework for cybersecurity risk prediction. These findings highlight the limitations of conventional statistical approaches and emphasize the need for intelligent, data-driven models capable of capturing hidden patterns in high-dimensional log data [14], [17].

IV. CONCLUSION

This study developed a data-driven machine learning framework for predicting cybersecurity risks using behavioral and temporal features derived from email server logs. The findings demonstrate that while traditional statistical methods provide useful preliminary insights, their ability to capture complex relationships within the data is limited. Specifically, correlation analysis revealed weak linear associations among numerical variables, whereas Chi-square analysis identified strong and statistically significant relationships among categorical features, particularly between protocol type, error classification, and authentication outcomes.

These results highlight the multidimensional nature of cybersecurity data, in which risk patterns are governed by nonlinear interactions and feature dependencies rather than by isolated predictors. By integrating behavioral indicators with temporal dynamics, the proposed framework enables a more comprehensive representation of system activity and enhances the detection of anomalous behavior.

Among the evaluated models, the Random Forest algorithm demonstrated superior performance, achieving the highest predictive accuracy and AUC. This confirms the effectiveness of ensemble learning methods in capturing complex patterns and improving classification

robustness in cybersecurity applications. The key contribution of this study is to demonstrate that combining statistical analysis with machine learning provides a systematic approach to feature understanding and model development for cybersecurity risk prediction. The proposed framework is practical and adaptable, making it suitable for real-world deployment in intrusion detection systems and intelligent security monitoring platforms.

Despite these contributions, the study is limited by dataset imbalance and the scope of available features. Future research should focus on incorporating additional contextual attributes, such as IP reputation, geographic information, and user behavior profiles, and on exploring hybrid deep learning architectures (e.g., LSTM-based models) to better capture temporal dependencies and further enhance predictive performance.

REFERENCES

- [1] Admass, W. S., Munaye, Y. Y., & Diro, A. A. (2024). Cyber security: State of the art, challenges and future directions. *Cyber Security and Applications*, 2, 100031. <https://doi.org/10.1016/j.csa.2023.100031>
- [2] Alladi, T., Chamola, V., Sahu, N., & Guizani, M. (2020). Artificial intelligence and blockchain for cybersecurity: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(4), 3432–3465. <https://doi.org/10.1109/COMST.2020.3012460>
- [3] Alshamrani, A., Myneni, S., Chowdhary, A., & Huang, D. (2024). Machine learning in cybersecurity: Applications, challenges, and future directions. <https://doi.org/10.32628/CSEIT24102125>
- [4] Flick, C., & Worrall, K. (2022). The ethics of creative AI. In *The Language of Creative AI* (pp. 73–91). Springer. https://doi.org/10.1007/978-3-031-10960-7_5
- [5] Gandotra, E., & Gupta, D. (2021). An efficient approach for phishing detection using machine learning. In *Multimedia Security* (pp. 239–253). Springer. https://doi.org/10.1007/978-981-15-8711-5_12
- [6] He, Z. (2025). Machine learning for cybersecurity: A survey of applications. *Electronics*, 14(23). <https://doi.org/10.3390/electronics14234563>
- [7] Janati, M., & Messaoudi, F. (2025). Intrusion detection system-based network behavior analysis. *International Journal of Advanced Computer Science and Applications*, 16(3), 793–802. <https://doi.org/10.14569/IJACSA.2025.0160378>
- [8] Kaushik, S. S., et al. (2025). Robust machine learning based intrusion detection system. *Scientific Reports*, 15, 3970. <https://doi.org/10.1038/s41598-025-88286-9>
- [9] Landauer, M. M., et al. (2020). System log clustering approaches for cyber security applications. *Computers & Security*, 92, 101739. <https://doi.org/10.1016/j.cose.2020.101739>
- [10] Lu, C., Cao, Y., & Wang, Z. (2024). Research on intrusion detection based on an enhanced random forest algorithm. *Applied Sciences*, 14(2), 714. <https://doi.org/10.3390/app14020714>
- [11] Ojo, A. O. (2025). A review on the effectiveness of AI in cybersecurity. *JKLST*, 4(1), 1–12.



- [12] Parlanti, T. S., & Catania, C. A. (2025). Temporal analysis framework for intrusion detection systems. *arXiv*. <https://doi.org/10.48550/arXiv.2511.03799>
- [13] Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42. <https://doi.org/10.12691/ajams-8-2-1>
- [14] Singh, A., et al. (2024). Machine learning-based intrusion detection systems for cybersecurity applications. *Alexandria Engineering Journal*. <https://doi.org/10.1016/j.aej.2024.01.013>
- [15] Strauss, C., Harr, M. D., & Pieper, T. M. (2025). Analyzing digital communication. *Management Review Quarterly*, 75(4), 3119–3157. <https://doi.org/10.1016/j.mrq.2025.101002>
- [16] Talukder, M. A., et al. (2024). Machine learning-based network intrusion detection. *Journal of Big Data*, 11, 33. <https://doi.org/10.1186/s40537-024-00886-w>
- [17] Wardana, A. A., et al. (2024). Federated random forest with feature selection for collaborative intrusion detection in IoT. *Procedia Computer Science*, 246, 20–29. <https://doi.org/10.1016/j.procs.2024.09.193>
- [18] Wu, Y., Zou, B., & Cao, Y. (2024). Deep learning-based intrusion detection models. *Journal of Imaging*, 10(10), 254. <https://doi.org/10.3390/jimaging10100254>
- [19] Zhang, Y., Chen, X., Li, S., & Wang, L. (2021). A machine learning approach for cybersecurity intrusion detection. *IEEE Access*, 9, 34567–34578. <https://doi.org/10.1109/ACCESS.2021.3051234>
- <https://doi.org/10.1007/s11301-024-00455-8>



Development of a School Profile Website as a Digital Information Medium Using the Waterfall Method

Muhammad Mustakim¹, Sofiansyah Fadli^{2*}, Ahmad Tanton³

^{1,3}Department of Information Systems, STMIK Lombok, Praya, Indonesia

²Department of Informatics Engineering, STMIK Lombok, Praya, Indonesia

Email: ¹akimtemun9@gmail.com, ²sofiansyah182@gmail.com, ³ahmad.tantoni@students.amikom.ac.id

Abstract - The development of information technology encourages educational institutions to utilize digital media in delivering information more effectively. However, SMPN 4 Praya Tengah still uses conventional methods in disseminating information, resulting in limited reach and inefficiency. This study aims to design and implement a school profile website as a digital information medium that can be widely accessed by students, parents, and the community. The method in this study uses the Waterfall model, which was selected due to its systematic and structured development stages that are well-suited for projects with clearly defined and stable requirements. The Waterfall model includes the stages of analysis, design, implementation, and testing. This research integrates PHP and JavaScript programming languages with a MySQL database. The resulting system provides comprehensive features ranging from school profile, vision and mission, news, announcements, gallery, to online admission services (PPDB). The website is designed with a simple and user-friendly interface so it can be well managed by the school. Through Black Box Testing, this website has been proven to function optimally in presenting information accurately and in a structured manner. System validation was conducted through two methods: Black Box Testing, which confirmed all features operate according to specifications, and User Acceptance Testing (UAT) involving 50 respondents with 16 evaluation items, yielding an average score of 78.50% (categorized as 'Agree'). These results empirically validate that the implementation of this website provides measurable benefits in improving the effectiveness of information delivery, expanding access, and supporting digital transformation within the school environment. This system is expected not only to facilitate digital transformation in schools but also to serve as a reference for the development of other school websites.

Keywords: *design, implementation, web-based, information system, waterfall method.*

I. INTRODUCTION

Advances in information technology have brought profound transformation to various aspects of human life, particularly in the Education sector [1]. The implementation of web-based information systems is one form of digital transformation that educational institutions widely apply to optimize the distribution of information effectively and efficiently [2]. An educational institution's website functions as a publication medium that can present important data quickly, widely, and easily accessible to the public [3]. Through the website, information on the school profile, vision and mission, academic and non-academic activities, announcements, and other relevant details can be delivered in an integrated manner to students, parents, and the general public. Managing information through a web-based system also allows for the presentation of data in a systematic and precise manner and supports the dynamic updating of information according to the relevance of user needs [4].

The utilization of websites as a medium for educational information is also considered capable of improving the transparency of educational institutions, as well as the quality of information services to the public [5]. A school website serves as a public communication medium that facilitates interaction between the school and the community [6]. In addition, the website can also be used as a means of promoting educational institutions, documenting school activities, and providing information that can be accessed anytime and from anywhere [7]. Thus,

the existence of a school website not only supports the delivery of information more effectively, but also contributes to enhancing the image and public trust in the educational institution [8].

Even so, there are still educational institutions that do not have a website as an official means of information dissemination [9]. One example is SMPN 4 Praya Tengah, which is located in the Praya Tengah District, Central Lombok Regency. Based on the results of observations and initial interviews conducted in January 2026, it was found that the process of delivering information at the school is still carried out conventionally, such as through the school notice board or by putting up banners [10]. The use of this conventional method shows several systemic weaknesses, including limited information dissemination reach, a relatively small capacity of information that can be conveyed, and requiring more time and cost in the delivery process [11]. This condition causes school information to not be conveyed optimally to students, parents, or the general public.

Various previous studies have examined the development of school websites as digital information media [12]. Most previous research tends to focus on developing school websites as a medium for promoting educational institutions or integrating them into a larger framework, such as academic information systems and e-learning [12]. In addition, some studies only focus on the system design stage without continuing to the stage of implementing and thoroughly testing the system [13]. Other research is also widely applied at various levels of



education, ranging from preschool, primary education, to vocational secondary education, with various methodological approaches, such as Research and Development (R&D) as well as qualitative approaches [13].

Based on a review of previous research, there are still limitations in studies that specifically discuss the design and implementation of school profile websites as an official information medium in public junior high schools with a systematic software development approach [14]. In addition, literature that comprehensively combines the stages of needs analysis up to system testing in a single systematic development process is still relatively limited [15]. Therefore, research is needed that not only focuses on system design, but also directly implements the system so that it can be used by schools as a digital information medium.

In information system development, the selection of software development methods is one of the important factors in producing a system that meets user needs [16]. One of the methods widely used in software development is the Waterfall method [17]. This method has a development stage that is systematic and structured, which includes needs analysis, system design, implementation, and system testing [17]. The Waterfall method is considered suitable for this research compared to other methods, such as Agile or Prototype, because the system requirements have been clearly identified through observation and interviews prior to development, and are not expected to change significantly during the process. Several previous studies have also confirmed that the Waterfall method is effective for developing information systems with well-defined and stable requirements, particularly in educational website development contexts [17][18][19]. In contrast, Agile is more appropriate for projects with frequently changing requirements, while Prototype is better suited when user feedback is needed at early design stages. Since the scope of this system is focused and well-documented, Waterfall provides a more structured and measurable development process [17].

Based on these issues, this study aims to design and implement a school profile website as a digital information medium at SMPN 4 Praya Tengah using the Waterfall method [19]. The developed website is designed to provide various important information about the school, such as the school profile, announcements, event agendas, school news, documentation gallery, and contact information that can be widely accessed by students, parents, and the general public. In addition, the developed system is also designed to be easily managed by the school so that the process of delivering information can be carried out more effectively and sustainably [20].

The main contribution of this research is to produce a school profile website that is structured, easy to manage, and capable of improving the effectiveness of disseminating school information to the community. The system was validated through Black Box Testing to ensure functional correctness, and through User Acceptance Testing (UAT) with 50 respondents to empirically confirm that the website provides real benefits to its users [8]. In addition, this study is also expected to serve as a reference for the development of school profile websites at other

educational institutions as a simple yet effective digital information medium in supporting digital transformation in the educational environment [21].

II. RESEARCH METHODOLOGY

2.1 Types of Research

This research is a study on information system development that aims to design and implement a school profile website as a digital information medium [19]. The approach used in this research is the software engineering approach, which focuses on the processes of analysis, design, implementation, and structured system testing [16].

The software development method used in this research is the Waterfall method [16]. The method in this study uses the Waterfall model, which was selected due to its systematic and structured development stages that are well-suited for projects with clearly defined and stable requirements. The Waterfall model includes the stages of analysis, design, implementation, and testing. The Waterfall method was chosen because it has systematic and structured development stages, making it easier to develop the system gradually and measurably, from the initial analysis phase to the final functionality testing stage [16]. This method is also considered suitable for the development of school profile websites because the system requirements are relatively clear and do not undergo significant changes during the development process [22].

2.2 Research Location and Time

This research was carried out at SMPN 4 Praya Tengah, which is located in Praya Tengah District, Central Lombok Regency. The selection of the research location was based on the condition of the school, which does not yet have a website as a digital information medium, so the delivery of information is still carried out conventionally.

As for the research implementation time, it was carried out in several stages, which included the initial observation process, data collection, system design, system implementation, and system testing. This research was conducted within the time span from January 2026 to April 2026.

2.3 Data Collection Techniques

In this study, data collection techniques were conducted to obtain the information needed in the system development process. The data collection techniques used include:

a. Observation

Observation was conducted by directly observing the information delivery process that takes place at SMPN 4 Praya Tengah. Through this observation, the researcher can find out how the information delivery system is used by the school and identify various problems that occur in the process.

b. Interview

The data collection process was carried out through interview techniques with relevant school parties, such as the principal, administrative staff, and school operators. The interviews aimed to obtain direct information regarding system requirements, the types of information that need to be provided on the school



website, as well as the challenges faced in delivering information to students and the community.

c. Literature Study

A literature study is conducted by reviewing various reference sources relevant to the research, such as books, scientific journals, and previous studies related to the development of web-based information systems and school profile websites. This literature study aims to obtain a theoretical foundation and strengthen the scientific basis of the research.

2.4 System Development Method

The system development method used in this research is the Waterfall method. The Waterfall method is one of the software development methods that has systematic and sequential development stages [16]. The stages of the Waterfall method include requirements analysis, system design, implementation, and system testing [22].

The stages of system development in this research are explained as follows:

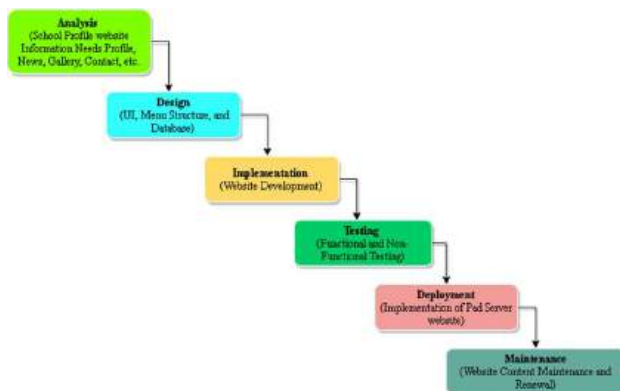


Figure 1. Waterfall Method

a. Requirement Analysis

At the analysis stage, the process of identifying system requirements is carried out based on the results of observations and interviews that have been conducted. The needs analysis aims to determine the type of information required by users and to define the features that will be provided on the school profile website [16] [23].

b. System Design

The system design stage is carried out after the system requirements have been successfully identified. At this stage, the design of the system structure is conducted, which includes database design, user interface design, and system flow design. The purpose is to provide an overview of how the system will be built before the implementation process is carried out. [16] [23][24].

c. Implementation

The implementation stage is the stage of applying the results of system design into program form. System implementation integrates the PHP and JavaScript programming languages with a MySQL database. The designed system is a school profile website that can be accessed by users through the internet network [22].

d. Testing

After the system is implemented, the next stage is system testing to ensure that the system can operate according to the functions that have been designed. System testing is conducted using the Black Box Testing method, which focuses on testing the system's functions without looking at the program code structure. This testing is carried out to ensure that all system features can work properly according to user needs [25] [24].

2.5 System Design

System design is carried out to illustrate the structure and workflow of the system to be developed. In this study, system design is carried out using several system modeling diagrams, including:

- A Use Case Diagram is used to show user interactions with the system.
- Flowchart used to depict the workflow of the system
- An Activity Diagram is used to describe the steps of the activities of the system.

Through the design of this system, it is expected that the process of developing a school profile website can be carried out more structured and systematically [18].

a. Use Case Diagram

A Use Case Diagram is one of the components in the Unified Modeling Language (UML) that is used to depict functional interactions between actors and the system. This diagram helps in identifying system requirements as well as the functional boundaries that will be developed [16]. Based on recent research, use case diagrams are very effective to use at the requirements analysis stage because they can provide a clear and easily understood overview of the system for both developers and users [26].

In this study, the Use Case Diagram is used to describe the interaction relationship between users and the admin with the school profile website system. Users can access various information such as the school profile, vision and mission, categories, gallery, directory, downloads, and contact. Meanwhile, the admin has access rights to manage all system data through the login process and content management.

The use of use case diagrams in this research aligns with the requirements analysis stage in the Waterfall method, where all system requirements must be clearly defined before entering the design and implementation stages [22].

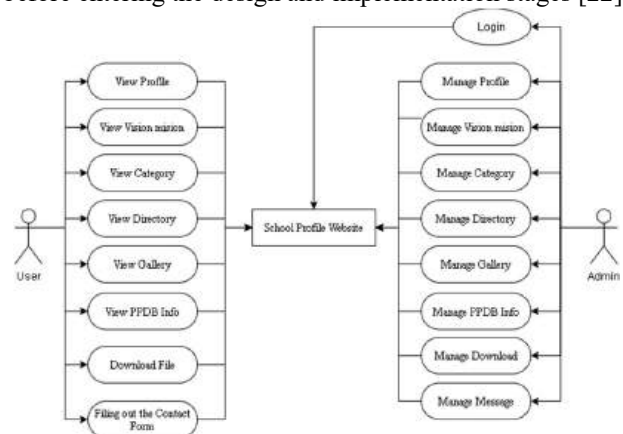


Figure 2. Use Case Diagram

b. Flowchart

A flowchart is a graphic diagram used to visualize the flow of a system process systematically with standard symbols. Flowcharts are used to visualize process steps, making it easier to understand the workflow of the system [22][27]. Recent research shows that flowcharts are very helpful in improving the clarity of system flow and minimizing errors during the implementation stage [26].

In this study, a flowchart is used to illustrate the system flow of the school profile website from the perspective of users and admins. The process starts with the user accessing the website, selecting a menu, and then the system processes the request and displays the information. On the admin side, the process starts from login, validation, and data management carried out through the dashboard.

The use of flowcharts in this research supports the system design stage in the Waterfall method, as it helps in illustrating the process flow in a structured manner before the system is implemented [22].

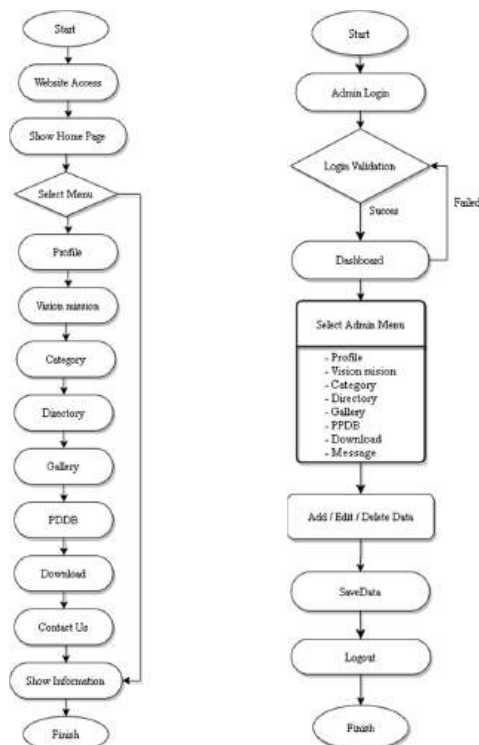


Figure 3. User and Admin System Flowchart

c. Activity Diagram

An Activity Diagram is a UML diagram used to illustrate the flow of activities in the system in more detail, including interactions between users and the system. This diagram shows the sequence of activities as well as the control flow within the system. Based on recent research, activity diagrams are very effective in modeling system processes dynamically and improving understanding of the system workflow [26].

In this study, an activity diagram is used to depict the interaction between users, the system, and the admin in the school profile website. The process begins when the user accesses the website, and then the system displays the main page. Next, the user selects a menu, and the system processes the request by retrieving data from the database before displaying the information.

On the admin side, the process starts with login, system validation, and then data management such as categories, gallery, directory, and downloads. Every data change will be saved and displayed directly on the system.

The use of activity diagrams in this research supports the system design process in the Waterfall method because it can depict the flow of activities in detail and in an integrated manner [22].

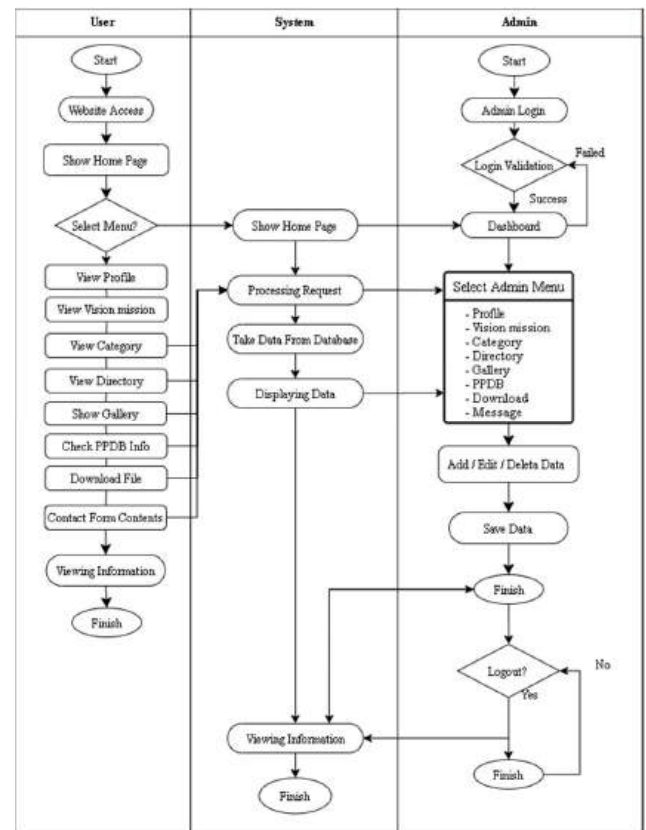


Figure 4. Activity Diagram

2.6 System Testing Methods

The testing method used in this study is Black Box Testing. This method is used to test the system functions based on user requirements without considering the internal structure of the program. Testing is carried out by examining each feature available on the school profile website, such as the school profile management feature, announcements, news, gallery, activity agenda, and the school contact feature [25].

The results of this testing are used to ensure that all system features can function properly and meet user needs, so that the school profile website can be used optimally as a digital information medium [22].

III. RESULTS AND DISCUSSION

3.1 Admin Login

The school profile website system consists of two user levels, namely the website admin and the user. After accessing the system, the admin will be directed to the login page [12].

Each admin role has specific functions and access rights, whereas users act as visitors who can access the information available on the website as follows:



Figure 5. Admin Login Page

On this page, the admin is required to input a username/email and password, which will then be validated by the system. If the data is valid, the admin will be directed to the dashboard, whereas if it is not valid, the system will deny access and request re-entry. This mechanism plays an important role in maintaining system security and preventing illegal access to data.



Figure 6. Admin Homepage

After going through the authentication process, the admin can access various main menus available in the system. The admin homepage functions as a control center in system management, designed to provide easy navigation and quick access to all system functions, thereby improving efficiency in data management. The admin dashboard or homepage plays an important role in enhancing the system's usability because it presents information and features in a structured and easily accessible manner.

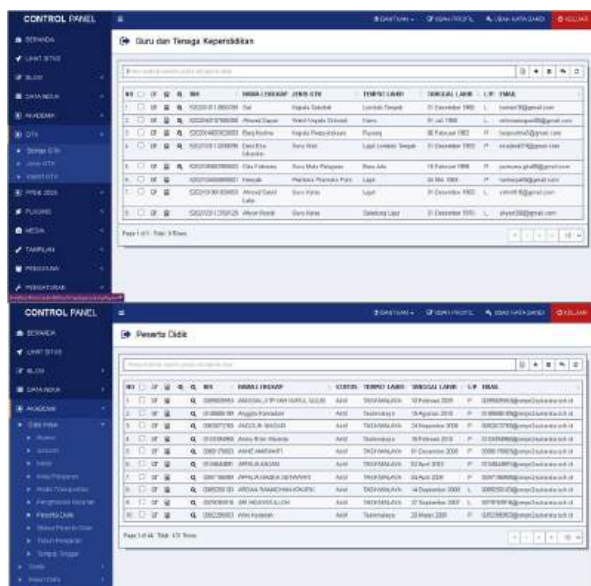


Figure 7. GTK and Student Data Page

The GTK (Teachers and Education Personnel) and Student Data Page is a feature used to manage teacher, education staff, and student data in a centralized system. This page displays important information such as full name, gender, place, and date of birth, as well as other supporting data in a structured table format. This feature makes it easier for admins to manage data, such as searching, adding data, editing data, and deleting data efficiently. The presentation of data in an interactive table format also helps to increase information access speed and accuracy in data processing. A web-based data management system can improve administrative efficiency and simplify the data management process within the education environment.

3.2 Login User

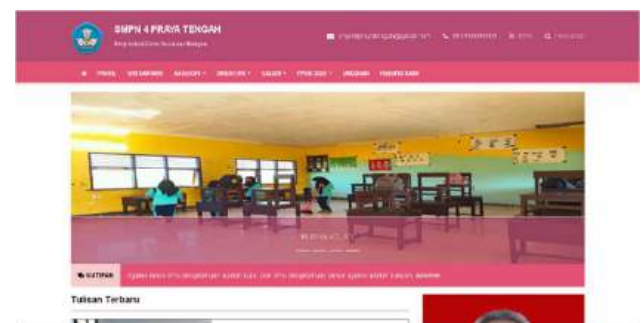


Figure 8. User Home page

The user's homepage is the initial view accessed by users when opening the school profile website via a link or website domain. This page serves as the main medium for conveying information to users, such as the school profile, news, announcements, student admission information (PPDB), and navigation menus to other features. The homepage design is made simple and informative with the aim of making it easy for users to obtain the information they need. A structured and user-friendly homepage layout can enhance the user experience as well as the effectiveness of information delivery in a web-based system.



Figure 9. Profile Page

The school profile page is one of the main components of a school website that serves to provide general information about the identity and overview of the institution to the public. This page becomes a digital information source that can be accessed by students, parents, and the community to get to know the school comprehensively without having to visit in person.



Figure 10. Vision and Mission Page

The vision and mission page is one of the important components on a school website that serves to present the goals, direction, and core values that serve as guidelines in the development of the institution. This page becomes a digital information source that can be accessed by students, parents, and the public to understand the school's policy direction and aspirations comprehensively without having to come in person.



Figure 11. Photo Gallery Page



Figure 12. Video Gallery Page

The gallery page is a feature used to display documentation of school activities in visual and audiovisual forms to users. On this page, users can view a collection of photos and videos arranged systematically, making it easier to access information related to school activities. Presenting content in the form of images and videos aims to increase the website's appeal and provide information that is more communicative and interactive compared to text alone. The use of visual and audiovisual media on the website can enhance user engagement as well as the effectiveness of information delivery in a web-based system.



Figure 13. News Category Page

The news category page is a feature used to group and display school news information based on certain categories, making it easier for users to find relevant information. On this page, news is organized in a structured and systematic manner, allowing users to quickly access the latest information or information based on the desired topic. Grouping information in the form of categories also helps to improve content orderliness and ease of navigation within the system. Organizing content based on categories on a website can enhance the efficiency of information searching as well as the user experience in accessing a web-based system.



Figure 14. PPDB Form Page



Figure 15. PPDB Form Print Page



Figure 16. PPDB Selection Page

The PPDB page has several main features, namely the registration form, form printing, and selection results. The form page is used by prospective students to input registration data online with a structured format that has been adjusted. Once the data is successfully saved, users can access the form printing page to download or print the registration proof as an archive. In addition, the selection results page allows users to check the acceptance status by entering certain data, such as registration number and date of birth. The integration of these three features supports a more efficient, transparent, and easily accessible registration process. The web-based PPDB system is able to improve administrative efficiency, data accuracy, and ease of information access for users compared to conventional methods.

3.3 System Testing

The developed system was then tested before being fully implemented. Testing was carried out using the Black Box Testing method, which focuses on the system's functions without considering the internal code structure. The test results are shown in Table 1, which includes various test scenarios to ensure that all system features function according to specifications [22].

Table 1. Blackbox Testing

Feature Test	Expected Result	Result
Admin Login	The system is able to validate the username/email and password input, and direct the user to the admin dashboard page if the entered data is valid.	Valid
Admin Home Page	The system is able to display the admin dashboard page containing brief information as well as the system data management menu.	Valid
GTK and Student Data Page	The system can display teacher and student data in a structured form in tables and provides data management functions (add, edit, and delete)	Valid
User Home Page	The system is able to display the main page of the website, which contains general information and navigation to other features.	Valid
Profile Page	The system is capable of displaying school profile information completely, in a	Valid

	structured manner, and is easy for users to understand.	
Vision and Mission Page	The system is able to present the school's vision and mission information clearly and informatively.	Valid
Photo Gallery Page	The system is able to display a collection of photo documentation of school activities that can be accessed by users.	Valid
Video Gallery Page	The system is able to display video content of school activities and ensure that the videos can be played properly.	Valid
News Category Page	The system is able to display news information based on the categories selected by the user systematically.	Valid
PPDB Form Page	The system is able to display the PPDB registration form and process the data entered by the user.	Valid
PPDB Form Print Page	The system is able to display and provide a printing feature for the registration form data according to the data that has been entered.	Valid
PPDB Selection Page	The system is able to display the PPDB selection results accurately based on the registration data that has been processed by the system.	Valid

The results of Black Box Testing indicate that all test scenarios have been successfully executed according to the specifications. Furthermore, user acceptance testing was conducted using the User Acceptance Test (UAT) method, where respondents were asked to answer several questions with a predetermined evaluation weight [25].

Table 2. Assessment Weights

Score Weight	Description
1	Strongly Disagree
2	Disagree
3	Neutral
4	Agree
5	Strongly Agree

In addition to Black Box Testing, User Acceptance Testing (UAT) was also conducted to validate the system's benefits from the user's perspective, ensuring that the developed website truly meets user needs and expectations [25]. These criteria are used as a reference in determining the UAT assessment weights presented in Table 3.

Table 3. UAT Assessment Criteria

Percentage	Description
81% - 100%	Strongly Agree
61% - 80%	Agree
41% - 60%	Neutral
21% - 40%	Disagree
0% - 20%	Strongly Disagree

The questions posed by the researcher to the users consist of 16 items that must be answered by the respondents, as shown in Table 4.

Table 4. List of Questions

NO	Questions
1	Is the Home Page layout easy for users to understand?
2	Is the information presented on the Home Page complete?
3	Is the information on the school Profile Page presented clearly and easily understood?
4	Does the Profile Page provide a sufficient overview of the school?
5	Are the displayed vision and mission easy to understand?
6	Is the information about the vision and mission presented engagingly?
7	Is the Photo Gallery layout attractive and easy to access?
8	Does the displayed photo provide relevant information about school activities?
9	Can the displayed video be played properly?
10	Does the video content help you understand school activities or the school profile?
11	Is the presented news easily accessible based on categories?
12	Is the displayed news information current and useful?
13	Is the PPDB form easy for users to fill out?
14	Are the instructions for filling out the form clear?
15	Can the form print feature be used easily?
16	Does the printed form match the data entered?

The number of respondents involved was 50 people. After that, the answers given by the respondents were calculated using weights that had been determined based on the answers received, according to the calculation formula.

$$\frac{\text{Total Respondent Weigh} / \text{Number of Respondents}}{\text{Maximum Weigh}} \quad (1)$$

1. $\frac{147/50}{5} \times 100\% = 58,80\%$ "Neutral"
2. $\frac{191/50}{5} \times 100\% = 76,40\%$ "Agree"
3. $\frac{234/50}{5} \times 100\% = 93,60\%$ "Strongly Agree"
4. $\frac{175/50}{5} \times 100\% = 70,00\%$ "Agree"
5. $\frac{191/50}{5} \times 100\% = 76,40\%$ "Agree"
6. $\frac{218/50}{5} \times 100\% = 87,20\%$ "Strongly Agree"
7. $\frac{182/50}{5} \times 100\% = 72,80\%$ "Agree"
8. $\frac{194/50}{5} \times 100\% = 77,60\%$ "Agree"
9. $\frac{209/50}{5} \times 100\% = 83,60\%$ "Strongly Agree"
10. $\frac{179/50}{5} \times 100\% = 71,60\%$ "Agree"
11. $\frac{205/50}{5} \times 100\% = 82,00\%$ "Strongly Agree"

12. $\frac{217/50}{5} \times 100\% = 86,80\%$ "Strongly Agree"
13. $\frac{157/50}{5} \times 100\% = 62,80\%$ "Neutral"
14. $\frac{201/50}{5} \times 100\% = 80,40\%$ "Strongly Agree"
15. $\frac{241/50}{5} \times 100\% = 96,40\%$ "Strongly Agree"
16. $\frac{199/50}{5} \times 100\% = 79,60\%$ "Agree"

The next step is to calculate the average percentage of all questions using the following formula: $(58.80\% + 76.40\% + 93.60\% + 70.00\% + 76.40\% + 87.20\% + 72.80\% + 77.60\% + 83.60\% + 71.60\% + 82.00\% + 86.80\% + 62.80\% + 80.40\% + 96.40\% + 79.60\%) / 16 = 78.50\%$. The result of calculating the average percentage of all questions shows that the UAT testing for the School Profile Website falls under the "Agree" criteria".

IV. CONCLUSION

Based on the research that has been conducted, the design and implementation of the School Profile Website as a digital information medium at SMPN 4 Praya Tengah was successfully developed using the Waterfall method, which includes the stages of requirements analysis, system design, implementation, and testing. The website provides various main features such as school profile, vision and mission, news, announcements, photo and video gallery, news categories, as well as an online-based student admission service that can be accessed by students, parents, and the wider community. This system is designed with a simple, structured, and easy-to-use interface, making it easier for the school to manage and deliver information more effectively. The testing results using the Black Box Testing method showed that all system features functioned according to the design without any significant functional issues. In addition, the User Acceptance Test (UAT) involving 50 respondents with 16 questions obtained an average percentage score of 78.50%, which falls into the "Agree" category. These results indicate that the developed system has been well received by users and is able to enhance the effectiveness of information delivery more quickly, accurately, and efficiently compared to previous conventional methods. The implementation of this website is expected to support the digital transformation process in the educational environment and improve the quality of school information services. For further development, the system can be added with more comprehensive features such as academic system integration, e-learning, student attendance, and other administrative services to optimize the website's functions. In addition, the development of a more interactive and responsive user interface across various devices also needs to be considered to enhance the user experience. Further research can also consider using other software development methods such as Agile or Prototype so that the development process is more flexible to changes in user needs, as well as expanding testing methods by involving more respondents and usability evaluations so that the system's quality can continue to be improved.



REFERENCES

- [1] J. Manajemen Pendidikan and Y. Putri Nirwana Boang Manalu, "Strategi Kepala Sekolah Dalam Membudayakan Literasi Dan Numerasi Melalui Program Kuliners Di Smpn 3 Sidikalang," vol. 11, no. 1, 2026, doi: 10.34125/jmp.v11i1.1057.
- [2] R. Khofifah, A. A. Harahap, A. Ratnasari, R. Nur, and R. Dzakiyullah, "Analisis Kualitas Website Smk Ma'arif 1 Yogyakarta Menggunakan Metode Webqual 4.0," *Indexia: Informatic and Computational Intelligent Journal*, vol. 07, no. 02, pp. 120–129, 2025.
- [3] R. Hidayat¹, R. Prayudhi², A. Fatoni³, and D. D. Disetujui, "Analisis Kualitas Layanan Pada Website Perpustakaan Seskoal Terhadap Kepuasan Pengguna Dengan Metode Webqual," *Journal Computer Science*, vol. 3, no. 2, 2024.
- [4] N. Wati, A. Nehemia Toscani, B. Irawan, and F. I. Komputer, "Analisis Kualitas Website Smkn 2 Tanjung Jabung Timur Menggunakan Metode Webqual 4.0," *Jurnal Manajemen Teknologi dan Sistem Informasi (JMS)*, vol. 3, no. 2, 2023, doi: 10.33998/jms.v3i1.
- [5] M. Shobri, "Peran Sistem Informasi Manajemen Pendidikan dalam Meningkatkan Transparansi dan Akuntabilitas di Lembaga Pendidikan Islam," *AKSI: Jurnal Manajemen Pendidikan Islam*, vol. 2, no. 2, pp. 78–88, Mar. 2024, doi: 10.37348/aksi.v2i2.302.
- [6] Afkar Hanif Syaifuddin, "Pengembangan Sistem Manajemen Pendidikan Islam Berbasis Web untuk Transparansi dan Akuntabilitas," *IHSANIK: Jurnal Pendidikan Agama Islam*, vol. 3, no. 1, pp. 308–320, Jan. 2025, doi: 10.59841/ihsanika.v3i1.2211.
- [7] F. Zahroh, A. Zulaikha Agustina, L. Cahyani, N. Aini, and E. Yuni Wijaya, "Pengelolaan Website Informasi Sekolah Sebagai Media Promosi Digital Di Smpn 1 Labang School Information Website Management As Digital Promotional Media At Smpn 1 Labang." [Online]. Available: <http://jurnal.faperta-unras.ac.id/index.php/pakdemas>
- [8] J. P. Masyarakat *et al.*, "Perkembangan Portal Informasi Berbasis Website Di SMK YKWI Pekanbaru".
- [9] H. Nabila and S. Listyo Prabowo, "Model Implementasi Sistem Informasi Manajemen Berbasis Web pada Lembaga Pendidikan," *Jurnal Teknologi Pendidikan*, vol. 3, no. 2, 2026, doi: 10.47134/jtp.v3i2.2394.
- [10] R. Maulana Muhammad, S. Marwiyah, R. Aranggraeni, V. Nandya Paramitha, and U. Soetomo, "Perlindungan Hukum Bagi Konsumen Dalam Transaksi Digital Dengan Modus Penipuan," *Bureaucracy Journal: Indonesia Journal of Law and Social-Political Governance*, vol. 5, no. 3, doi: 10.53363/bureau.v5i3.734.
- [11] Roudhotum Mawardania and Henny Dwi Bhakti, "Perancangan Sistem Presensi Sekolah Berbasis Web Di SD Muhammadiyah Gresik," *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 4, no. 1, pp. 78–90, Jan. 2024, doi: 10.55606/juisik.v4i1.745.
- [12] A. T. Zy, M. Najamuddin, D. Miharja, A. S. Rakhmat, M. Hatta Fahamsyah, and U. Pelita Bangsa, "Implementasi Media Promosi dan Informasi pada SMP Insan Kamil Cikarang Berbasis Website." [Online]. Available: <https://lenteranusa.id/>
- [13] Y. Saputra and D. Kurniadi, "Pengembangan Sistem Layanan Informasi Berbasis Web di SMK Negeri 1 Padang," *MASALIQ*, vol. 6, no. 2, pp. 679–707, Feb. 2026, doi: 10.58578/masaliq.v6i2.9064.
- [14] F. S. Farhan, H. Aspriyono, and A. Al Akbar, "Perancangan sistem informasi sekolah berbasis web pada TK Aisyiyah Bustanul Athfal 1 Bengkulu," *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 4, no. 1, pp. 65–81, Feb. 2025, doi: 10.24246/itexplore.v4i1.2025.pp65-81.



- [15] H. Nabila and S. Listyo Prabowo, "Model Implementasi Sistem Informasi Manajemen Berbasis Web pada Lembaga Pendidikan," *Jurnal Teknologi Pendidikan*, vol. 3, no. 2, 2026, doi: 10.47134/jtp.v3i2.2394.
- [16] "Dokumen Software Requirement Specification (SRS) Sistem Informasi Pemasaran Usaha Jasa Percetakan Dan Iklan".
- [17] S. Sallu, Y. Harsono, and O. Fajarianto, "Implementation of Waterfall Method in Model Development to Improve Learning Quality of Computer Network Courses," vol. 25, no. 3, pp. 496–513, 2023, doi: 10.21009/JTP2001.6.
- [18] F. Stephanus Suwita, W. N. Fadli, A. Suhendra, and F. Z. Sulthoni, "Implementation of Waterfall Method in Designing Website-Based Discussion Forums," 2023. [Online]. Available: <https://ojs.unikom.ac.id/index.php/injudes>
- [19] I. I. Hidayat, I. R. Komala, and A. Ruhiat, "Perancangan Sistem Informasi Profil Sekolah Berbasis Website Menggunakan Metode Waterfall (Studi Kasus MTS Al-Hikmah Cicelot)," *Journal Artificial: Informatika dan Sistem Informasi*, vol. 3, no. 2, pp. 213–231, Aug. 2025, doi: 10.54065/artificial.860.
- [20] Y. Anis, A. B. Mukti, and A. N. Rosyid, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Model Waterfall Dalam Pengembangan Sistem Informasi Aset Destinasi Wisata Berbasis Website," *Media Online*, vol. 4, no. 2, pp. 1134–1142, 2023, doi: 10.30865/klik.v4i2.1287.
- [21] A. Jurnal Publikasi *et al.*, "Pembuatan Dan Implementasi Website Profil Sekolah Sebagai Media Informasi Dan Promosi Digital Sdn Genteng Kota Bogor", [Online]. Available: <https://jurnal.portalpublikasi.id/index.php/AJP/index>
- [22] A. E. Sumakul *et al.*, "Web-Based Guest Lecture Information System for Committee and Student Users at FMIPA UNSRAT," vol. 7, no. 2, pp. 88–102, 2026, doi: 10.33365/jatika.v7i2.1608.
- [23] Y. Rahmayati and S. Fadli, "Penerapan Metode Waterfall Dalam Pengembangan Sistem Informasi Pelayanan Pendudukan Berbasis Web (Studi Kasus: Desa Muncan)".
- [24] S. Hamdi *et al.*, "Penerapan Metode Waterfall Dalam Pengembangan Website Usaha Mikro, Kecil Dan Menengah (Studi Kasus: Umkm Pengrajin Pandan Lanji Bersatu)," 2025.
- [25] M. Yusuf *et al.*, "Analisis dan Testing Sistem Aplikasi SiAbon dengan Menggunakan Metode Black Box di Dinas PUPR Provinsi Jambi," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 3, pp. 4074–4079, Aug. 2025, doi: 10.31004/riggs.v4i3.2596.
- [26] M. R. Wayahdi and F. Ruziq, "Pemodelan Sistem Penerimaan Anggota Baru dengan Unified Modeling Language (UML) (Studi Kasus: Programmer Association of Battuta)," *Jurnal Minfo Polgan*, vol. 12, no. 1, pp. 1514–1521, Aug. 2023, doi: 10.33395/jmp.v12i1.12870.
- [27] S. Fadli *et al.*, "Perancangan Sistem Dengan Metode Waterfall Pada Apotek Xyz," Cetak, 2018. [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/misi>

Evaluation of Computer-Based Testing Website Quality Using the DeLone and McLean Information System Success Model

Muhamad Sakinul Wadi¹, Sofiansyah Fadli^{2*}, Mohammad Taufan Asri Zaen³

^{1,3}Department of Information Systems, STMIK Lombok, Praya, Indonesia

²Department of Informatics Engineering, STMIK Lombok, Praya, Indonesia

Email: ¹galaumister46@gmail.com, ²sofiansyah182@gmail.com, ³opanzain@gmail.com

Abstract – The rapid development of information technology has significantly impacted the education sector, particularly in the implementation of computer-based examination systems. This study aims to analyze and evaluate the quality of a Computer-Based Testing (CBT) website at SMK Bangun Bangsa using the DeLone and McLean Information System Success Model. A quantitative approach was employed by collecting primary data through questionnaires distributed to 260 students from grades X and XI of SMK Bangun Bangsa who had experience using the CBT system. The respondents represented various demographic backgrounds in terms of gender, age, and academic majors. The variables analyzed include system quality, information quality, service quality, user satisfaction, and net benefits. Data analysis was conducted using descriptive statistical methods, supported by validity and reliability testing to ensure the accuracy and consistency of the research instrument. The results indicate that the CBT system has generally performed well. System quality obtained a score of 75.4% (good category), indicating that the system is easy to use despite minor technical issues. Information quality achieved 77.7% (very good), reflecting that the system provides accurate, relevant, and understandable information. Service quality recorded the lowest score of 72.2% (good), suggesting the need for improvement in security and technical support. Meanwhile, user satisfaction reached 81.6% and net benefits 80%, both categorized as very good, indicating that the system enhances efficiency and provides a positive user experience. This study contributes to evaluating the effectiveness of CBT systems in vocational education environments and highlights the importance of improving service quality and system reliability to achieve optimal system performance. Future research is recommended to expand the study scope and integrate additional models for deeper analysis.

Keywords – Computer Based Testing, DeLone and McLean, System Quality, User Satisfaction, Information Systems Success

I. INTRODUCTION

The development of information technology in the past few decades has shown a very significant increase and has a wide impact on various sectors, including education [1] [2]. The implementation of information systems has proven capable of improving work effectiveness, time usage efficiency, and resource optimization within an organization [3] [4]. In the educational environment, especially at the secondary school and higher education levels, the use of information systems not only supports administrative processes but also plays an important role in improving the quality of learning and the evaluation of learning outcomes.

One of the forms of learning evaluation that is routinely carried out is the Final Semester Examination (UAS) [5]. This exam serves as an instrument to measure the level of understanding and cognitive abilities of students after undergoing the learning process in a certain period [6]. Conventionally, the implementation of UAS is carried out using the Paper Based Test (PBT) method, which relies on paper media and writing tools [7]. However, this method is considered to have various limitations, such as the potential for errors in correction, the risk of question leakage, and inefficient use of resources [8].

Along with technological advancements, computer-based evaluation systems or Computer Based Testing (CBT) have begun to be adopted as a more modern and efficient alternative. CBT allows exams to be conducted flexibly, both in terms of time and location, and can produce evaluation results more quickly and accurately [9].

In addition, this system can also minimize the potential for cheating through question randomization and improve exam data security [3]. Nevertheless, the implementation of CBT still faces challenges, especially related to the level of user adaptation, where some students are still accustomed to using simple platforms such as Google Forms in conducting exams [10].

SMK Bangun Bangsa is one of the educational institutions that has implemented a computer-based examination system since 2019. The implementation of this system provides various conveniences, including in exam data management, speeding up the assessment process, and increasing the accuracy of evaluation results. In addition, the CBT system also contributes to reducing human error in the correction process and curbing the potential for cheating during exams. However, the success of implementing the system is not only determined by the presence of technology, but also by the quality of the system, the quality of the information produced, and the quality of the service experienced by users.

To measure the success level of an information system, the DeLone and McLean model is one of the evaluation frameworks widely used. This model assesses system success based on several key dimensions, namely system quality, information quality, service quality, usage level, user satisfaction, and net benefits [11] [12]. This model is considered comprehensive because it not only measures technical aspects but also takes into account users' perceptions and experiences in using the system [13].

Several previous studies have shown that the DeLone and McLean model is effective in evaluating the success of



various information systems. Research by [14] shows that system quality affects the level of use of health center management information systems. Furthermore, [15] found that the success of e-learning systems is influenced by ease of use and system display quality. Another study by [16] confirmed that system, information, and service quality contribute to improving the effectiveness of academic information systems. In addition, studies by [17] [18] also showed that most variables in the DeLone and McLean model have a significant effect on satisfaction and the success of information system usage.

Although various studies have been conducted, research that specifically evaluates the quality of computer-based examination systems at the vocational high school level is still relatively limited. In addition, the relationship between system quality, information quality, and service quality in influencing user satisfaction in CBT implementation requires further investigation. Based on the identified gap, this study aims to analyze and evaluate the quality of the Computer-Based Testing (CBT) examination website at SMK Bangun Bangsa using the DeLone and McLean model. This study focuses on examining the influence of system quality, information quality, and service quality on user satisfaction, as well as identifying the net benefits obtained from the implementation of the system.

The results of this study are expected to provide a comprehensive overview of the success level of CBT system implementation, and at the same time serve as a basis for the development and improvement of computer-based examination systems in the future, especially in the vocational secondary education environment.

II. RESEARCH METHODOLOGY

A. Research Design

This study employs a quantitative approach with a descriptive method. The quantitative approach is used to measure the level of system quality and user satisfaction based on numerical data obtained from respondents [19] [20]. The descriptive method aims to provide a systematic and accurate description of the characteristics of the studied variables [20] [21].

B. Population and Sample

The population in this study consists of all students of SMK Bangun Bangsa who have used the Computer-Based Testing (CBT) system. The sample was determined using a sampling technique accidental sampling that is sampling based on students who happen to be available and willing to fill out the questionnaire. The sample size was determined using the Slovin formula:

$$n = \frac{N}{1 + N(e)^2} \quad (1)$$

Where: n = sample size, N = population size, e = margin of error. How to determine the total number of samples needed from a population is determined by applying the Slovin formula. The calculation formula using the population of SMK Bangun Bangsa, which consists of 742 students, is as follows:

$$n = \frac{742}{1 + 742 (0.05)^2}$$

$$n = \frac{742}{1 + 742 \cdot 0.0025}$$

$$n = \frac{742}{1 + 1.855}$$

$$n = \frac{742}{2.855} = 259.8 \text{ rounded to } 260$$

Therefore, the number of samples required is 260 out of 742 students.

C. Data Collection Techniques

The data in this study consists of primary data obtained directly from respondents. The data collection techniques used include: Observation, conducted by directly observing the use of the computer-based exam system in the school environment. Interviews, conducted with school personnel such as principals, teachers, and students to obtain information related to their experience using the system. Literature study, conducted by reviewing literature such as scientific journals, books, and previous relevant research. Questionnaires, as the main research instrument developed based on the indicators of the DeLone and McLean model. The questionnaire uses a 4-point Likert scale, namely: Strongly Agree (4), Agree (3), Disagree (2), and Strongly Disagree (1).

Table 1. Likert scale score weights

Code	Description	Score
SS	Strongly Agree	4
S	Agree	3
TS	Disagree	2
STS	Strongly Disagree	1

D. Research Procedure

This research was carried out through several structured stages, namely: (a) problem formulation based on empirical conditions and gaps in previous research; (b) data collection through observation, interviews, questionnaires, and literature study; (c) literature review as a theoretical foundation; (d) data analysis using statistical software; (e) testing of research instruments; and (f) preparation of the research report systematically.

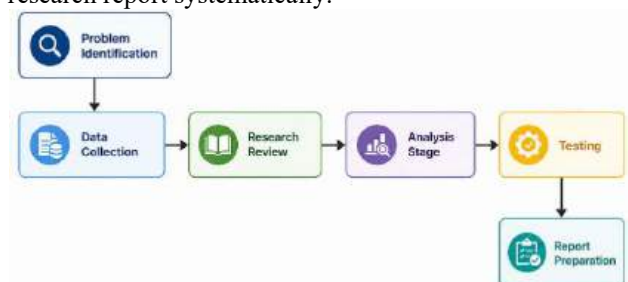


Figure. 1 Research Method Stages.

E. Research Model

This study adopts the [22] information system success model as the analytical framework. This model is used to evaluate information system success through several main constructs, namely system quality, information quality, service quality, user satisfaction, and net benefit. The use variable is not included in this study because the system being examined is mandatory for users.

Each variable is measured using indicators that have been developed in previous studies, such as ease of use, access speed, system reliability, information accuracy, information relevance, service responsiveness, as well as

the impact of system use on user effectiveness and productivity.

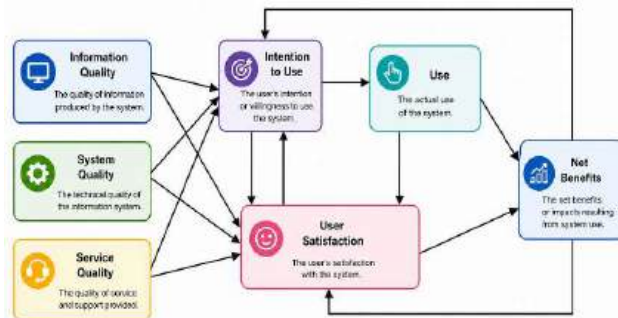


Figure 2. Information system success model of DeLone and McLean

F. Population and Sample

The population in this study consisted of all grade X and XI students at SMK Bangun Bangsa who had used the computer-based examination system, totaling 742 students. These students came from four academic majors, namely Computer and Network Engineering, Office Management, Financial Accounting, and Marketing. The respondents also represented diverse demographic characteristics in terms of gender and age, providing a broad representation of CBT system users within the school. The sampling technique used a non-probability sampling method with a accidental sampling approach, which is the selection of respondents based on certain criteria relevant to the research objectives [23] [24] [25].

The determination of the sample size used Slovin's formula with an error rate of 5% [26] [27], resulting in a sample of 260 respondents. This number is considered representative to describe the overall condition of the population.

G. Data Analysis Techniques

Data analysis was conducted using SPSS version 26 software with several stages as follows:

1) Validity Test

The validity test was conducted to determine the extent to which the research instrument is able to measure the intended variables [28] [29]. The test was carried out by comparing the calculated *r* value with the table *r* value. An item is declared valid if the calculated *r* value is greater than the table *r* value [30].

2) Reliability Test

The reliability test aims to measure the consistency of the research instrument. The test was done using the Cronbach's Alpha coefficient. An instrument is declared reliable if the Cronbach's Alpha value is greater than 0.70 [31] [32] [33].

3) Descriptive Analysis

The collected data was analyzed descriptively by calculating the mean and percentage of each variable. The analysis results are presented in tables to facilitate data interpretation.

H. Research Instrument

The research instrument consists of a questionnaire prepared based on indicators from each variable in the DeLone and McLean model [22]. The indicators used include system quality aspects (ease of use, access speed, reliability), information quality (accuracy, relevance,

completeness), service quality (responsiveness, technical support), user satisfaction, and the net benefits perceived by users.

Table 2. Research Variables and Indicators

Variable	Indicators
System Quality	- Ease of use - Response time - System reliability - Interface design - System security
Information Quality	- Accuracy - Relevance - Completeness - Understandability - Timeliness
Service Quality	- Responsiveness - Technical support - Empathy - Service reliability
Use / Intention to Use	- Frequency of use - Intensity of use - Intention to continue use
Net Benefits	- Improved effectiveness - Improved productivity - Cost/time saving - Better decision making

III. RESULTS AND DISCUSSION

A. Results

This study involved 260 students of SMK Bangun Bangsa who have used the computer-based testing system (Computer Based Testing/CBT). Before further analysis was carried out, instrument testing was conducted to ensure the quality of the data used.

B. Research Instrument Test

The results of the validity test showed that all questionnaire items had a calculated *r* value greater than the *r* table (0.121), so all items were declared valid. In addition, the results of the reliability test showed that all variables had a Cronbach's Alpha value above 0.70, which means the research instrument is consistent and reliable.

Validity testing is carried out to ensure that each statement item in the questionnaire is able to measure the researched variable accurately. This test uses the Pearson Product Moment method by comparing the calculated *r*-value with the table *r*-value. Based on the number of respondents totaling 260, the table *r*-value obtained is 0.121 at a 5% significance level. An item is declared valid if the calculated *r*-value is greater than the table *r*-value. The results of the validity testing for all research variables are presented in Table 3.

Table 3. Validity Test Results

Variable	Item	r-value	r-table (5%)	Criteria
System Quality	1	0.749	0.121	Valid
	2	0.698	0.121	Valid
	3	0.754	0.121	Valid
	4	0.689	0.121	Valid
	5	0.580	0.121	Valid
	6	0.758	0.121	Valid

Information Quality	1	0.804	0.121	Valid
	2	0.865	0.121	Valid
	3	0.818	0.121	Valid
	4	0.838	0.121	Valid
	5	0.824	0.121	Valid
Service Quality	1	0.801	0.121	Valid
	2	0.823	0.121	Valid
	3	0.849	0.121	Valid
User Satisfaction	1	0.875	0.121	Valid
	2	0.904	0.121	Valid
	3	0.845	0.121	Valid
Net enefits	1	0.868	0.121	Valid
	2	0.872	0.121	Valid
	3	0.867	0.121	Valid

Based on Table 3, it can be seen that all statement items for the variables of system quality, information quality, service quality, user satisfaction, and net benefits have r-calculated values greater than the r-table (0.121). Therefore, all questionnaire items are declared valid and suitable for use as data collection instruments. This result indicates that each indicator used has been able to accurately represent the construct being measured.

After the instrument is declared valid, the next stage is to conduct a reliability test to determine the level of consistency of the research instrument. The reliability test is carried out using the Cronbach's Alpha method. A variable is considered reliable if it has a Cronbach's Alpha value greater than 0.70. The results of the reliability test for each research variable are presented in table 4.

Table 4. Reliability Test Results

No	Variable	Cronbach Alpha	Criteria	Description
1	System Quality	0.791	> 0.70	Reliabel
2	Information Quality	0.885		Reliabel
3	Service Quality	0.762		Reliabel
4	User Satisfaction	0.846		Reliabel
5	Net enefits	0.836		Reliabel

Based on Table 4, all variables in this study have a Cronbach's Alpha value above 0.70. This indicates that the research instrument has a good level of consistency and can be trusted (reliable). Therefore, the questionnaire used in this study is deemed suitable for use in the further data analysis process.

C. Respondent Characteristics

The characteristics of respondents in this study are presented in Table 5. Based on the table, the majority of respondents were male, accounting for 58.5%, while females accounted for 41.5%. In terms of age, respondents were dominated by 17-year-olds at 59.6%, indicating that most respondents were within an age range active in technology use. Based on grade level, the majority of respondents were from grade XI at 73.8%, while grade X accounted for 26.2%. This indicates that respondents had sufficient experience in using computer-based examination systems. In terms of majors, most respondents came from the Computer Network and Telecommunications major

(53.1%), followed by Office Management (25%), Financial Accounting (19.2%), and Marketing (2.7%).

Overall, the characteristics of the respondents indicate that the majority of users have a background that is quite relevant to the use of technology, thus enabling them to provide representative assessments of the quality of the computer-based examination system.

To provide a clearer demographic profile of the respondents, this study analyzed several demographic characteristics, including gender, age, grade level, and academic major. These characteristics are important to ensure that the respondents adequately represent the population of CBT system users at SMK Bangun Bangsa.

Table 5. Characteristics of Respondents

Characteristic	Category	Frequency	Percentage (%)
Gender	Male	152	58.5
	Female	108	41.5
Age	16 Years Old	51	19.6
	17 Years Old	155	59.6
	18 Years Old	48	18.5
	19 Years Old	6	2.3
Class	Grade X	68	26.2
	Grade XI	192	73.8
Major	Computer and Network Engineering	138	53.1
	Office Management	65	25
	Financial Accounting	50	19.2
	Marketing	7	2.7
Total		260	100

D. Descriptive Analysis of Variables

Descriptive analysis was conducted to understand respondents' perceptions of the quality of the computer-based examination system based on the DeLone and McLean model. The summary of the average scores for each variable is presented in Table 6.

Table 6. Percentage of variable scores

Variable	Score	Percentage
System Quality	784.3	75.4%
Information Quality	808.4	77.7%
Service Quality	750.7	72.2%
User Satisfaction	848.7	81.6%
Net enefits	832.0	80.0%

Based on Table 6, the results are as follows:

- System Quality received a score of 75.4% (good category). This indicates that the system is relatively easy to use, has clear navigation, and a fairly good interface. However, there are still technical issues such as errors experienced by some users.
- Information Quality received a score of 77.7% (very good category). The information presented is considered accurate, relevant, easy to understand, and timely.
- Service Quality received a score of 72.2% (good category) and is the lowest compared to the other variables. This shows that the service aspect,



- particularly related to system security and technical stability, still needs improvement.
- d. User Satisfaction received the highest score of 81.6% (very good category). This indicates that the CBT system is able to provide a positive experience for users.
 - e. Net Benefits obtained a value of 80% (very good category), which indicates that the system has a positive impact on the efficiency and effectiveness of exam implementation.

To provide a more detailed overview of the contribution of each indicator to the research variables, a summary of the dominant indicators for each variable is presented. This summary aims to identify the most influential aspects as well as aspects that still need to be improved in the implementation of the computer-based examination system. The results of the indicator summary are presented in Table 7.

Table 7. Summary of Indicators per Variable

Variable	Dominant Indicator	Percentage (%)	Category	Notes
System Quality	Ease of use	90.4	Very Good	System is easy to understand
	System stability	59.3	Good	Errors still occur
Information Quality	Ease of understanding	88	Very Good	Information is clear
	Timeliness	84.7	Very Good	Information is up-to-date
Service Quality	Security of use	82	Very Good	Safe to use
	Data privacy	73.5	Good	Needs improvement
User Satisfaction	Ease of use	91.2	Very Good	Easier to use than before
	Effectiveness of use	90.4	Very Good	Process is more efficient
Net Benefits	Time efficiency	89.2	Very Good	Saves time
	Effectiveness of use	88.5	Very Good	Supports outcomes

Based on Table 7, it can be seen that for the system quality variable, the ease of use indicator obtained the highest score, which indicates that the system has been well designed and is easy for users to understand. However, the system stability indicator still received a relatively lower score, which suggests that there are technical issues that need to be addressed. For the information quality variable, all indicators fall into the very good category, particularly in terms of ease of understanding and timeliness. This indicates that the system is able to provide information that meets users' needs optimally. Furthermore, for the service quality variable, the aspect of usage security received a high score, but personal data security still needs improvement. For the user satisfaction and net benefit variable, all indicators show very good scores, indicating that the system provides a positive experience as well as tangible benefits for users.

Overall, Table 7 reinforces the results in Table 6, that although the system has been running well, aspects of system stability and service quality still remain priority areas for further development.

E. Discussion

Based on the research results presented in Table 6 and supported by the indicator details in Table 7, it can be analyzed that the success of implementing a computer-based examination system at SMK Bangun Bangsa is influenced by a combination of system quality, information, services, and users' perceptions of the benefits gained. These findings are in line with the DeLone and McLean information systems success model, which states that system quality, information, and service are the main determinants in shaping user satisfaction and system benefits.

First, in terms of system quality, a score of 75.4% (good category) indicates that the system generally meets user needs. This is reinforced by Table 5, which shows that the ease of use indicator received a very high score (90.4%), indicating that the system is user-friendly and easy to understand. However, the system stability indicator only reached 59.3%, which indicates that there are still technical disruptions. These findings show that although the system has been optimal in terms of design and usability, technical factors such as system reliability still pose challenges that can affect the overall user experience. This result is in line with the research of [34] which asserts that good system quality must be supported by stability and reliability to increase the level of system usage.

Second, the quality of information scored 77.7% (very good category), indicating that the system is capable of providing relevant and high-quality information. Based on Table 5, the indicators of ease of understanding (88%) and timeliness (84.7%) are the dominant factors supporting this assessment. This indicates that users not only receive information, but are also able to understand and utilize it effectively in the context of examinations. This finding is in line with [15] who stated that good information quality will enhance users' positive perceptions of the system, especially in terms of clarity and ease of access to information.

Third, in terms of service quality, the lowest score was obtained, which is 72.2% (good category). Further analysis in Table 5 shows a gap between the perception of usage safety (82%) and personal data security (73.5%). This indicates that although users feel fairly safe when using the system, there are still concerns regarding the protection of personal data. Thus, the service aspect - particularly security and technical support - becomes an area that requires more attention in system development. This finding is supported by research by [16], which emphasizes that service quality, especially in terms of security and technical support, is an important factor in increasing user trust and satisfaction.

Fourth, user satisfaction achieved the highest score of 81.6% (very good category). This is supported by the indicators of ease of use (91.2%) and exam effectiveness (90.4%) in Table 5, which shows that the system is able to provide a better experience compared to conventional exam methods. This high user satisfaction indicates that the system has successfully met user expectations in both functional and operational aspects. These results are in line with the research of [18] [11], which states that user satisfaction is greatly influenced by ease of use and the system's effectiveness in supporting user activities.



Fifth, the net benefits also show a high value of 80% (very good category). Based on Table 5, the indicators of time efficiency (89.2%) and exam implementation effectiveness (88.5%) are the main factors contributing to this value. This indicates that the CBT system not only improves the efficiency of the exam process but also has a positive impact on students' learning outcomes. This finding aligns with the research of [17], which states that the benefits of information systems can be seen from increased productivity, efficiency, and the quality of results obtained by users.

Overall, the analysis in Table 4 and Table 5 shows that the success of a computer-based examination system is not determined by a single factor, but rather by the interaction between system quality, information quality, and service quality, which ultimately affects user satisfaction and perceived benefits. This aligns with the DeLone and McLean model framework, which emphasizes the causal relationships between variables in determining information system success. Although most variables fall into the good to very good category, this study identifies that improving system stability and service quality, particularly in aspects of security and reliability, is a key factor in achieving a more optimal system success.

IV. CONCLUSION

Based on the results of the analysis using the DeLone and McLean model, this study shows that the computer-based testing (CBT) system at SMK Bangun Bangsa has generally been running well and is able to provide satisfaction and benefits to its users. The system quality is in the good category, indicating that the system is easy to use and has a supportive interface, although there are still issues with technical stability. Information quality obtained a very good category, indicating that the information presented is accurate, relevant, and easy to understand. Meanwhile, service quality is the relatively lowest aspect, although still in the good category, indicating the need for improvements in data security and technical support. On the other hand, user satisfaction and net benefits show very good results, indicating that the CBT system is able to improve efficiency, effectiveness, and provide a positive experience for users. Overall, the success of the system is influenced by the interaction between system quality, information, and services, which shape user satisfaction and perceived benefits. For further research, it is recommended to conduct experiments by expanding the research subjects to various educational institutions to obtain broader generalization of results, and to combine the DeLone and McLean model with other approaches such as the Technology Acceptance Model (TAM) or the Unified Theory of Acceptance and Use of Technology (UTAUT) to analyze technology acceptance factors more comprehensively. Additionally, further experiments can also be focused on testing the improvement of system stability and strengthening data security-based services in real-time to continuously enhance the quality of the CBT system.

REFERENCES

- [1] D. Hawkrige, *New information technology in education*. Routledge, 2022.
- [2] S. Timotheou *et al.*, "Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review," *Education and information technologies*, vol. 28, no. 6, pp. 6695-6726, 2023, doi: 10.1007/s10639-022-11431-8.
- [3] I. A. Ariwanta, I. E. Saputra, N. P. E. Apriyanti, I. Gunawan, and G. Indrawan, "Analisis kepuasan pengguna menggunakan metode eucs pada sistem computer based test di institusi pendidikan," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 4, pp. 942-951, 2023, doi: 10.47065/josyc.v4i4.3752.
- [4] P. Diawati, S. S. Gadzali, M. K. N. Abd Aziz, A. M. A. Ausat, and S. Suherlan, "The role of information technology in improving the efficiency and productivity of human resources in the workplace," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 3, pp. 296-302, 2023, doi: 10.47233/jteksis.v5i3.872.
- [5] K. Talman, J. Vierula, T. Karihtala, E. Laakkonen, J. Engblom, and E. Haavisto, "Development and Validity Evaluation of a National Digital Entrance Examination Test for Higher Education Student Selection," *Higher Education Quarterly*, vol. 79, no. 1, p. e70006, 2025, doi: 10.1111/hequ.70006.
- [6] A. D. Augustia, C. N. Agustia, D. Azzahra, L. Hakim, and V. Pratiwi, "Analisis validitas dan reliabilitas soal pilihan ganda dengan menggunakan software Anates pada mata pelajaran perpajakan," *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, vol. 5, no. 1, pp. 250-265, 2025, doi: 10.37481/jmeh.v5i1.1165.
- [7] O. Tananda, M. Nasir, M. Milana, and M. Muslim, "Perbandingan antara sistem computer based test dan paper based test pada hasil pembelajaran mata pelajaran pmkr di smkn 1 sumatera barat," *JTPVI: Jurnal Teknologi Dan Pendidikan Vokasi*



- Indonesia, vol. 1, no. 1, pp. 67-74, 2023, doi: 10.24036/jtpvi.v1i1.10.
- [8] R. Setiawan, "Optimasi Pengalaman Pengguna: Penilaian Otomatis dan Pencegahan Kecurangan Ujian Online," *bit-Tech*, vol. 7, no. 2, pp. 299-306, 2024, doi: 10.32877/bt.v7i2.1758.
- [9] M. Sakban and Y. M. Silaban, "Aplikasi Ujian Computer Based Test (CBT) Berbasis Web," *Jurnal Bisantara Informatika*, vol. 8, no. 2, pp. 65-77, 2024.
- [10] Z. Wahiah, S. M. Prabowo, and H. A. S. Hana, "Eksplorasi Efektivitas Tes Pilihan Ganda Berbasis Komputer Sebagai Evaluasi Pembelajaran: Jurnal Pendidikan," *EDUCATIVO: Jurnal Pendidikan*, vol. 2, no. 2, pp. 342-347, 2023.
- [11] A. M. Kartini, S. Fadli, and H. Fahmi, "The DeLone and McLean model for measuring success hospital management information system case study: Praya Regional Hospital," *JISA (Jurnal Informatika dan Sains)*, vol. 8, no. 1, pp. 66-73, 2025, doi: 10.31326/jisa.v8i1.2203.
- [12] R. Saputri, M. E. Baining, and K. Khairiyani, "Pengaruh kualitas sistem dan kualitas pelayanan terhadap manfaat bersih aplikasi mobile banking dengan variabel intervening kepuasan pengguna," *E-Bisnis: Jurnal Ilmiah Ekonomi Dan Bisnis*, vol. 17, no. 1, pp. 126-138, 2024, doi: 10.51903/e-bisnis.v17i1.1801.
- [13] F. Rozi, S. Fadli, and M. T. A. Zaen, "Application of the DeLone and McLean Success Model to the SEMAIK Website: A Case Study in Central Lombok," *JISA (Jurnal Informatika dan Sains)*, vol. 8, no. 2, pp. 220-227, 2025, doi: 10.31326/jisa.v8i2.2554.
- [14] D. Y. Nofita and J. Veri, "Mengukur keberhasilan penerapan aplikasi Sistem Keuangan Desa (Siskeudes) dengan model kesuksesan Delon and McLean," *Jurnal Ekobistek*, vol. 13, no. 2, pp. 27-35, 2024, doi: 10.35134/ekobistek.v13i2.776.
- [15] R. Pambudi, B. Berlilana, and G. Karyono, "Analisis Kesuksesan Pengguna Website E-Learning Menggunakan Model DeLone & McLean," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 4, no. 3, pp. 948-967, 2024, doi: 10.51454/decode.v4i3.683.
- [16] N. N. Pusparini and A. Sani, "Mengukur keberhasilan penerapan sistem informasi akademik dengan model kesuksesan Delon and Mclean," *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, vol. 4, no. 2, pp. 149-155, 2020, doi: 10.46880/jmika.Vol4No2.pp149-155.
- [17] S. R. Agustini, "Analisis Kesuksesan Aplikasi Noice Terhadap Pengguna DiKota Jambi Menggunakan Metode DeLone and Mclean," *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, vol. 4, no. 1, pp. 883-891, 2024, doi: 10.33998/jakakom.2024.4.1.1660.
- [18] E. Wicaksono, E. M. Safitri, M. A. S. Anam, and R. A. Bimantara, "Analisis kepuasan pengguna pada aplikasi gojek menggunakan pendekatan metode delone-mclean," in *Prosiding Seminar Nasional Teknologi dan Sistem Informasi*, 2023, vol. 3, no. 1, pp. 394-402, doi: 10.33005/sitasi.v3i1.613.
- [19] M. Esmacili Givi, H. Keshavarz, and Z. Kargar Azad, "Quality assessment of E-learning website using asymmetric impact–performance analysis and Kano’s customer satisfaction model: a case study based on WebQual 4.0," *Information Discovery and Delivery*, vol. 51, no. 1, pp. 35-46, 2023, doi: 10.1108/IDD-08-2021-0083.
- [20] A. I. Pauji, R. Rida, and R. H. Emynorane, "The Effect Of Web-Based Academic Information System Use On Student Satisfaction: A Quantitative Study In Islamic Higher Education," *JoIEM (Journal of Islamic Education Management)*, vol. 6, no. 2, pp. 99-105, 2025, doi: 10.30762/joiem.v6i2.6799.
- [21] X. J. Mamakou, P. Zaharias, and M. Milesi, "Measuring customer satisfaction in electronic commerce: the impact of e-service quality and user experience," *International Journal of Quality &*

- Reliability Management*, vol. 41, no. 3, pp. 915-943, 2024, doi: 10.1108/IJQRM-07-2021-0215.
- [22] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: a ten-year update," *Journal of management information systems*, vol. 19, no. 4, pp. 9-30, 2003.
- [23] M. M. Rahman, "Sample size determination for survey research and non-probability sampling techniques: A review and set of recommendations," *Journal of Entrepreneurship, Business and Economics*, vol. 11, no. 1, pp. 42-62, 2023.
- [24] S. J. Stratton, "Population sampling: Probability and non-probability techniques," *Prehospital and Disaster Medicine*, vol. 38, no. 2, pp. 147-148, 2023, doi: 10.1017/S1049023X23000304.
- [25] M. A. Adeoye, "Review of sampling techniques for education," *ASEAN Journal for Science Education*, vol. 2, no. 2, pp. 87-94, 2023.
- [26] N. Nurkholis, Wilarso, P. Sukamto, M. A. Sobarnas, S. Jamaludin, and U. Tsani A, "Digital survey for customer satisfaction of regional drinking water companies (PDAM) using the mWater application and the Slovin formula method," in *BIO Web of Conferences*, 2024, vol. 144: EDP Sciences, p. 03006, doi: 10.1051/bioconf/202414403006.
- [27] M. K. H. Hasan and L. K. Kumar, "Determining adequate sample size for social survey research: sample size for social survey research," *Journal of the Bangladesh Agricultural University*, vol. 22, no. 2, pp. 146-157, 2024, doi: 10.3329/jbau.v22i2.74547.
- [28] T. Tugiman, H. Herman, and A. Yudhana, "Uji validitas dan reliabilitas kuesioner model Utaut untuk evaluasi sistem pendaftaran online rumah sakit," *JATISI*, vol. 9, no. 2, pp. 1621-1630, 2022.
- [29] E. F. Tarigan, S. Nilmarito, K. Islamiyah, A. Darmana, and R. D. Suyanti, "Analisis instrumen tes menggunakan Rasch model dan software SPSS 22.0," *Jurnal Inovasi Pendidikan Kimia*, vol. 16, no. 2, pp. 92-96, 2022, doi: 10.15294/jipk.v16i2.30530.
- [30] R. Mahkotawati, T. Rijanto, and P. W. Rusimamto, "Uji validitas dan reliabilitas instrumen penelitian: Angket pengalaman praktik kerja lapangan (PKL) pada siswa SMK," *JiIP-Jurnal Ilmiah Ilmu Pendidikan*, vol. 8, no. 2, pp. 1830-1835, 2025, doi: 10.54371/jiip.v8i2.6990.
- [31] F. D. P. Anggraini, A. Aprianti, V. A. V. Setyawati, and A. A. Hartanto, "Pembelajaran statistika menggunakan software SPSS untuk uji validitas dan reliabilitas," *Jurnal basicedu*, vol. 6, no. 4, pp. 6491-6504, 2022.
- [32] C. D. A. Istri and R. Fadila, "Uji validitas dan reliabilitas kuesioner pengetahuan masyarakat tentang program JKN," *Jurnal Kesehatan Qamarul Huda*, vol. 11, no. 1, pp. 307-315, 2023.
- [33] E. A. Cahyono and D. Darsini, "Sikap masyarakat terhadap program vaksinasi booster covid-19 (validitas dan reliabilitas kuesioner penelitian)," *Pengembangan Ilmu dan Praktik Kesehatan*, vol. 1, no. 1, pp. 1-21, 2022, doi: 10.56586/pipk.v1i1.185.
- [34] C. J. Viana, E. Selviyanti, A. Deharja, and A. P. Wicaksono, "Analisis Kesuksesan Sistem Informasi Manajemen Puskesmas (SIMPUS) Menggunakan Metode Delone and Mclean di UPT Puskesmas Sananwetan Kota Blitar," *J-REMI: Jurnal Rekam Medik dan Informasi Kesehatan*, vol. 5, no. 2, pp. 120-131, 2024, doi: 10.25047/j-remi.v5i2.4124.

Evaluation of Fresnel Zone in Ubiquiti Wireless Simulation and Implementation Using 802.11ac

Ahmad Tanton¹, Mohammad Taufan Asri Zaen^{2*}

¹Program Studi Teknik Informatika, STMIK Lombok

²Program Studi Sistem Informasi, STMIK Lombok

Email: ¹ahmadtanton@gmail.com, ²opanzain@gmail.com

Abstract – The development of wireless communication technology is increasing along with the need for high-speed internet access, both in urban and rural areas. Point-to-Point (PtP) technology based on the IEEE 802.11ac standard is one of the main solutions for providing long-distance connectivity, particularly in backhaul networks, ISPs, and inter-building communication. However, the network implementation process often relies on link simulations using software such as ISP Design Center or AirLink, which do not always accurately represent field conditions. Wave propagation factors, terrain contours, and obstacles in the Fresnel Zone significantly affect signal quality. The Fresnel Zone is a three-dimensional propagation area that must be at least 60% free of obstacles for optimal transmission. Previous research has shown limitations in studies related to evaluating the accuracy of Fresnel Zone simulations on Ubiquiti 802.11ac devices. Therefore, this research evaluates and compares the parameters of RX Signal Strength, throughput, channel width, and Fresnel Zone clearance between simulation results and direct field testing on a PtP connection. The research results are expected to serve as a technical reference for network practitioners in validating simulation results and providing recommendations for more accurate and efficient PtP connection design. This research also emphasizes the importance of Line of Sight (LOS) and Fresnel Zone optimization in maintaining the stability of outdoor wireless networks based on IEEE 802.11ac

Keywords – *Wireless, Fresnel Zone, Line of Sight; 802.11ac, Ubiquiti*

I. INTRODUCTION

The development of wireless communication technology continues to experience significant improvements, especially in meeting the need for high-speed internet access in both urban and rural areas. (Source: detikInet.com) [1] The number of internet users in Indonesia in 2024 reached 221,563,479 people out of a total population of 278,696,200 Indonesian residents in 2023, meaning the internet penetration rate in Indonesia reached 79.5%, an increase of 1.4% compared to the previous period. One of the technologies widely used to build long-distance connectivity is the Point-to-Point (PtP) system based on the IEEE 802.11ac standard. One example uses Ubiquiti devices that support implementation with the IEEE 802.11ac standard. Using this technology supports high throughput, channel efficiency, and transmission stability, making it widely applied in backhaul networks, internet service providers, RT/RW-Net (community-based internet networks), and communication systems between buildings.

In the process of implementing a wireless network, performing a link simulation using software such as ISP Design Center is often used as a primary reference to determine the height of the tower to be built, how to calculate the link budget, and to estimate the signal quality on the transmission path from the signal transmitter (main1) to the signal receiver (station1). However, the simulation results are often not entirely consistent with the field implementation. This is influenced by many factors such as wave propagation, terrain contours, types of obstacles, and clearance in the Fresnel Zone, which should ideally be completely free of obstructions when wireless technology is used. (Explanation of wireless according to Tanenbaum) [2] Wireless is a computer network

communication technology that uses electromagnetic waves as the transmission medium without using cables. This technology allows devices to connect to each other from short to long distances, with high flexibility and mobility of devices/hardware. Therefore, wireless has become one of the main foundations in the development of modern computer networks. According to Cisco, in the world of wireless [3] 802.11ac technology can transmit at a maximum speed of 1300 Mbps, almost three times faster than 802.11n. The 5 GHz spectrum used in 802.11ac achieves higher speeds for data throughput and provides more bandwidth for additional clients, wireless devices, and Internet of Things (IoT) devices to get online. Using beamforming, 802.11ac also results in less interference during transmission.

The Fresnel Zone is a very important factor in LOS (Line of Sight) communication, because interference can lead to increased noise, decreased RX signal strength resulting in loss of throughput and high latency. Previous research shows that at least 60% of the Fresnel Zone area must be free of obstacles to achieve optimal connection. According to Rappaport [4], The Fresnel Zone is an imaginary elliptical region between the transmitting antenna (main1) and the receiving antenna (station1) that forms the main path of radio wave propagation. This zone is not just a straight line called the Line of Sight (LOS), but also a three-dimensional space that must be kept at least 60% free of obstructions. When the Fresnel Zone is obstructed or blocked by trees or buildings, a decrease in signal strength will occur at the receiving (client) antenna. According to Kumar and Sharma [5], Obstacles in the Fresnel Zone are the main cause of signal quality degradation in wireless communication conducted outdoors. According to Mutmainnah and Hidayat [6] the existence of a clear Line of Sight (LOS) from the



transmitting antenna to the receiving antenna is the main requirement that must be met in the design of outdoor wireless networks so that signal quality can be maintained optimally. Radio communication systems require a completely unobstructed Line of Sight (LOS) condition. In Line of Sight (LOS) propagation, there is an area that must be free from obstacles; this area is called the Fresnel zone [7]. Research evaluating the accuracy of Fresnel Zone simulations against real-world implementations using Ubiquiti 802.11ac devices is still limited and not widely conducted. There are two types of building-to-building connections. The first type is called point-to-point (PtP), and the second is point-to-multipoint (PtMP). A point-to-point connection is a wireless connection between only two buildings and generally uses semi-directional or highly-directional antennas at each end of the connection [8]. In implementing point-to-point outdoor wireless communication, the objective is to determine how much bandwidth will be transmitted between the signal transmitter and the signal receiver, according to Forouzan [9] Bandwidth is the range of frequencies available for transmitting a signal. In digital communication, bandwidth refers to the amount of data (bits) that can be transmitted per second through a transmission medium, measured in bps (bits per second), Kbps, Mbps, and Gbps. In implementing wireless networks, there are obstacles that hinder signal transmission, known as noise floor. Sklar [10] explains that the noise floor determines the quality of the receiver sensitivity. For successful and effective communication, the received signal must be above the noise floor with an adequate SNR (Signal-to-Noise Ratio) margin. (Source: citraweb.com website) [11] the channel width parameter is directly proportional to data throughput, meaning that the wider the channel width, the greater the throughput that can be transmitted. Channel width is likened to a road, while the data passing through is likened to vehicles; meaning the wider the road, the more vehicles can pass through. (Quoted from Stallings) [12], In outdoor wireless communication, TX (Transmit) is the process of sending data from the source device to the destination device, while RX (Receive) is the process of receiving the signal sent by the transmitter.

Previous research has generally focused on link budget calculations, throughput analysis, or general PtP network design, but there has been no specific research on the gap between AirLink simulation results and field realization based on Fresnel Zone parameters. Therefore, a measurable evaluation is needed to determine the accuracy of AirLink simulations compared to real-world implementation in the field, particularly for the Ubiquiti 802.11ac technology currently used in outdoor wireless implementation scenarios.

This research evaluates the Fresnel Zone in simulations and implementations of Ubiquiti wireless networks using the 802.11ac standard by comparing parameters such as RX Signal Strength, throughput, channel width, and Fresnel Zone clearance values between simulation results and direct field testing. The results of this research are expected to serve as a valid technical reference for network practitioners and provide recommendations for more accurate and efficient point-to-point connection designs in the future.

Research conducted by Han [13] focuses on the Application of the Fresnel Zone Model in the Wireless Sensing Field. Previous research has explored the application of the Fresnel Zone in indoor wireless sensing, gesture recognition, and CSI-based monitoring. However, there is no research that evaluates the accuracy of the Fresnel Zone in simulations and field implementations using Ubiquiti 802.11ac devices in an outdoor Point-to-Point (PtP) network scenario. Therefore, this research focuses on analyzing the differences in technical parameters (RX Signal, throughput, Fresnel clearance, channel width, and noise floor) between AirLink simulation results and real-world implementation, with the aim of producing a more precise and applicable evaluation model and recommendations for wireless link design, thereby bridging the existing gap.

Research conducted by Nnadi, Jacob and Ezech [14] on Link Budget Analysis for Line of Sight Wireless Communication Link with Knife Edge Diffraction Obstruction aimed to perform a link budget analysis on a Line of Sight (LOS) wireless communication system with a knife-edge diffraction obstruction, and to determine the effect of the Fresnel Zone clearance percentage on diffraction loss, fade margin, link outage, and the feasibility of the communication link based on mathematical calculations. The conclusion of the previous research was that knife-edge diffraction obstruction was proven to increase diffraction loss and directly reduce the maximum distance achievable by the wireless communication system with acceptable quality of service (QoS). The resulting diffraction loss causes a decrease in received power and a reduction in fade margin, thus limiting the maximum communication distance. The larger the obstruction value (Pc%), the lower the maximum distance that can be achieved with acceptable quality of service. Therefore, Fresnel Zone clearance and link budget analysis are crucial in the design of LOS-based wireless networks. The difference between this research and previous studies is that previous research only conducted theoretical link budget analysis considering knife-edge diffraction and percentage changes in LOS clearance in microwave networks, so the analysis results were more mathematical and simulation-based. In contrast, this research will have a more applied approach by conducting a direct evaluation between AirLink simulations and actual field implementation using Ubiquiti 802.11ac devices. The research focuses on comparing RX Signal Strength, throughput, channel width, and Fresnel Zone clearance levels, thus providing validation of simulation performance and technical recommendations for more accurate PtP connection design under actual field conditions.

Research conducted by Hidayat et al [15] on the Analysis of Power Link Budget in Outdoor Wireless Networks Using Internet Service Provider Design Center: A Case Study of Kutanaganara Village, Garut, aimed to design an outdoor WiFi network at a frequency of 5725 MHz using the Ubiquiti ISP Design Center application as a representation of internet network implementation in Kutanaganara Village. It also analyzed the main network parameters to comply with KOMINFO regulations and TIPHON quality standards. The conclusion of this research is that the calculations and simulations of the Power Link



Budget parameters met the KOMINFO and TIPHON standards, making the outdoor WiFi network design deemed feasible for implementation in Kutanaganara Village. Previous research only focused on calculating and simulating the Power Link Budget using the ISP Design Center application to obtain basic parameters such as EIRP, FSL, FZC, and signal strength as a reference for network design, while this research added an accuracy evaluation stage by comparing the simulation results and direct field implementation based on Fresnel Zone, RX Signal Strength, channel width, and throughput parameters on Ubiquiti 802.11ac devices, resulting in a more comprehensive and applicable gap analysis of real-world conditions.

Research conducted by Ramadhan [16] on the Fresnel Zone Analysis and Link Budget of the WLAN Network from PT. Solo Jala Buana to BTS Karanganyar Using UBNT AIRLINK. The purpose of this research is to analyze Line of Sight (LoS) and Fresnel Zone in the design of a PtP wireless network. It also aims to determine the optimal antenna height and signal performance. The conclusion of this research is that the Radio Mobile simulator shows that with antenna heights of 20 meters (PT. Solo Jala Buana) and 40 meters (Karanganyar BTS), the transmission path has LoS and a Fresnel Zone free from obstacles, resulting in good signal and connection. The difference between this research and previous studies is that this research utilizes the UBNT AirLink and Radio Mobile simulators solely as design tools to determine the feasibility of the Link Budget and antenna height in a Point-to-Point network. This research, however, focuses on evaluating the accuracy of these simulations against real-world implementation conditions. This study specifically tests the validity of the simulation results by comparing parameters such as Throughput, Noise Floor, and RX Signal against the IEEE 802.11ac standard, in order to address the data discrepancies that often occur between the digital planning stage and physical implementation.

The research conducted by Anisa et al [17] on the Analysis of the Implementation of IEEE 802.11ac on the Wireless Backhaul Network of Zettalink Provider, aimed to design a Broadband Wireless Backhaul network based on IEEE 802.11ac, calculate the Power Link Budget for the designed network, including calculating Free Space Loss (FSL) and Received Signal Level (RSL), and analyze the designed network by comparing the theoretical calculation results (Power Link Budget) with actual measurement results using UISP software. The conclusion of this research is that the design, calculation, and analysis of the IEEE 802.11ac wireless backhaul network implementation in rural areas were successful. The theoretical calculation results (FSL = 121.68 dB, RSL = -49.68 dBm) closely matched the actual measurement results (RSL = -49 dBm). The implemented network showed good performance (throughput 480-655.20 Mbps) and was able to meet local internet needs. The difference in this research lies in the validation of the network design by comparing the manual Power Link Budget calculations (such as FSL and RSL) with actual field measurements. Conversely, previous research started from the problem of discrepancies between simulation and real-world implementation, so the focus of the analysis was to measure the accuracy and prediction gap

of the AirLink simulation software, with specific parameters on Fresnel Zone clearance, which is often the main cause of signal quality degradation in the field due to obstacles and terrain.

Research conducted by Wirawan et al [18] in "Analysis of Management Network Using Ubiquiti for Local Office at PT Rizki Inti Madani," aimed to optimize the company's network, which was facing challenges such as connectivity disruptions and uneven bandwidth allocation. The study involved a thorough analysis of the existing network conditions and the design of appropriate solutions using Ubiquiti devices. The conclusion of this research was that technicians and IT staff involved in network management should receive advanced training on the use of Ubiquiti and Cisco devices and efficient network troubleshooting techniques. The implementation of the new network should be done gradually to minimize the risk of operational disruptions. The difference between this research and previous studies is that previous research focused on validating manual calculations (theory) against field measurements (practice) to prove the success of a network design and focused on optimizing the architecture and management of local area networks (LAN/WLAN) in terms of efficiency, security, and centralized administration, with a systemic and managerial analysis. In contrast, the current research evaluates the accuracy of AirLink simulation software, not merely proving the design's success but measuring the gap between simulation predictions and real-world implementation, with the Fresnel Zone as the main variable. The current research is more technical and specific at the link level, evaluating the accuracy of predicting radio wave transmission performance for long-distance connections.

II. RESEARCH METHODOLOGY

The research methodology used is the PPDIOO Method, which is a network lifecycle used by Cisco to design, build, operate, and optimize computer networks to ensure they continue to function according to the organization's business needs. According to Teare [19] The explanation of the PPDIOO stages is as follows:

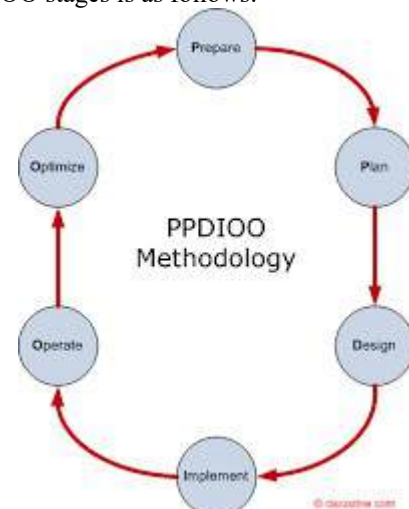


Figure 1. PPDIOO Methodology

A. Prepare

This involves identifying the business needs of an organization that support network development, proposing a reliable conceptual network architecture, and identifying the best technologies that can support the network architecture in the coming years.

B. Plan

The initial network requirements are identified based on technical needs, objectives to be achieved, facilities, user needs, hardware/software requirements, and others. This research involves determining the coordinate locations of the antenna towers to be built, both on the signal transmitting and receiving sides. This planning stage also involves calculating the cost of materials and equipment that will be used in accordance with the infrastructure supporting successful implementation.

C. Design

This involves designing a point-to-point (PtP) network topology simulation using the ispdesign.ui.com application. The main objective of this simulation is to ensure that the planned bandwidth (Mbps) can be successfully transmitted from the signal transmitter to the signal receiver. Determining the appropriate bandwidth (Mbps) requires considering the selection of suitable outdoor wireless hardware to achieve the targeted bandwidth. Various wireless network hardware options offer different throughputs based on the signal transmission distance between the transmitter and receiver. The resulting network design specifications are comprehensive and detailed, meeting both business and technical requirements to support availability, reliability, security, scalability, and performance.

D. Implement

This involves installing antenna towers on both sides, for both the signal transmitter (main1) and the signal receiver (station). It also includes implementing and configuring the outdoor wireless network devices according to the design and hardware used in the previous stage.

E. Operate

The operation phase is the final objective of the conformity in the design and implementation phases that have been carried out. The operation phase involves maintaining the outdoor wireless network through daily activities, including ensuring network availability, monitoring, troubleshooting, maintenance, and reducing costs. This operation phase will evaluate the Fresnel Zone in the simulation and implementation of Ubiquiti wireless using the 802.11ac standard by comparing the RX Signal Strength, throughput, channel width, and Fresnel Zone clearance values between the simulation results and direct field testing.

F. Optimize

The optimization phase involves proactive network management to identify and resolve problems that could affect the business or organization. Detecting and correcting errors (troubleshooting) is necessary when system conditions are unpredictable and can lead to failures. The essence of this optimization phase is to evaluate performance and make improvements if there are problems or new needs.

III. RESULTS AND DISCUSSION

A. Design

In this Design phase, simulations are performed using the web-based application ispdesign.ui.com, which runs in a browser, as follows:

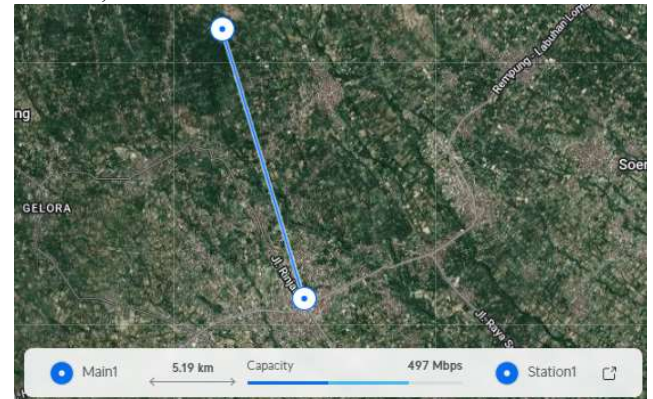


Figure 2. Point-to-Point (PtP) Simulation

Figure 2 shows that the distance between the signal transmitter (main1) and the signal receiver (station1) is 5.19 kilometers, with a simulated bandwidth of 497 Mbps.

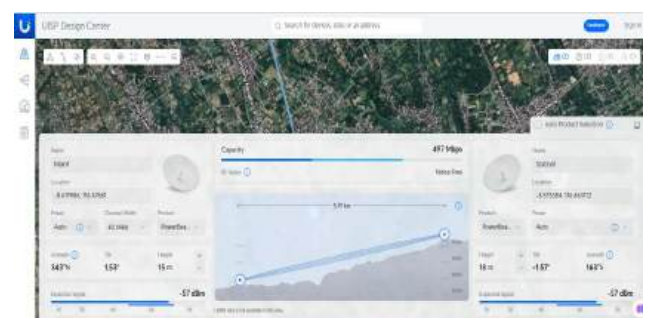


Figure 3. Complete Simulation Results

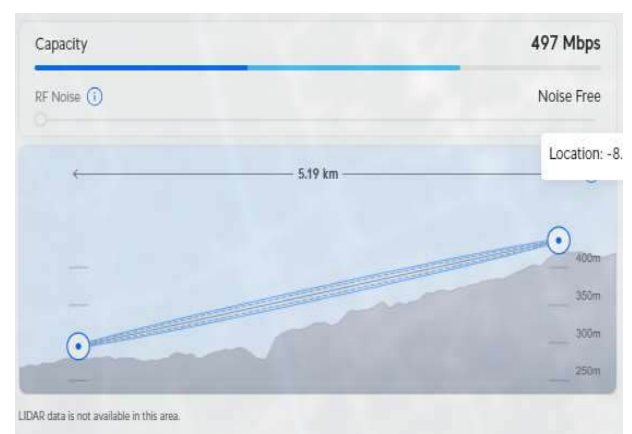


Figure 4. Point-to-Point Results (Side View)

Figure 3 shows the complete simulation results using Ubiquiti Powerbeam 5AC 620 wireless hardware on both sides of the antenna with a distance of 5.19 kilometers. Figure 4 shows the side view, where the signal transmitter (main1) is at an altitude of 300 meters above sea level, while the signal receiver (station1) is at an altitude of 450 meters above sea level.

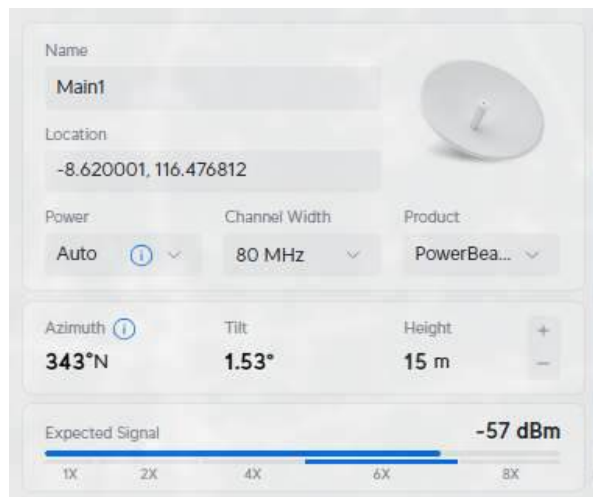


Figure 5. Simulation Results of the Signal Transmitter (main1)

Figure 5 shows the simulation results of the signal transmitter (Main) using Ubiquiti PowerBeam 5AC 620 hardware, with antenna installation coordinates at -8.619986, 116.47681, channel width of 80 MHz, horizontal angle (azimuth) of 343°N, tilt angle of 1.53° (meaning the receiving antenna is directed slightly upwards from the ideal horizontal line (0°)), antenna tower height of 15 meters, expected received signal (RX) of -57 dBm, and expected RX data rate of 6X.

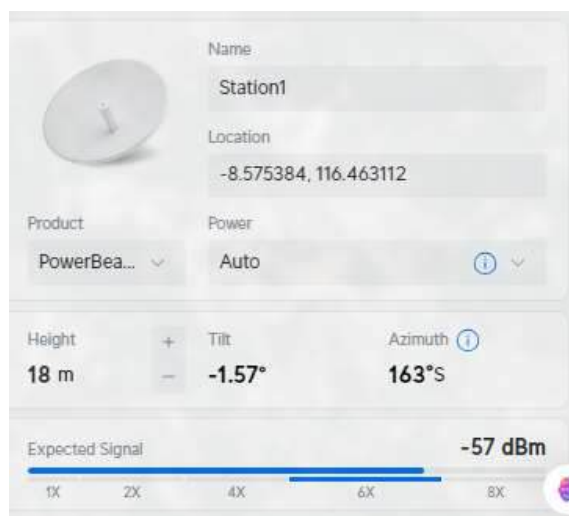


Figure 6. Simulation Results of the Signal Receiver (station1)

Figure 6 shows the simulation results of the signal receiver (Station) with Ubiquiti PowerBeam 5AC 620 hardware, antenna installation coordinates -8.575384, 116.463112, channel width 80 MHz, horizontal angle (azimuth) 163°S, tilt angle -1.57°, which means the receiving antenna is directed slightly downwards from the ideal horizontal line (0°), antenna tower height 18 meters, expected received signal (RX) is -57 dBm and expected RX data rate is 6X.

B. Implement

At the implementation stage, do the following:



Figure 7. Signal Transmitting Antenna (Main1)

Figure 7 shows the construction of a tower and signal transmitter (Main1) with Ubiquiti PowerBeam 5AC 620 hardware, the antenna's coordinates are Lat -8.619986° and Long 116.47681°, located in Masbagik Utara Village, Masbagik District.



Figure 8. Signal Receiving Antenna (station 1)

Figure 8 shows the construction of the tower and signal receiving antenna (station 1) with Ubiquiti PowerBeam 5AC 620 hardware, the coordinates of the antenna construction point are Lat -8.575384° and Long 116.463112°, located in Pringgasela Village, Pringgasela District.



Figure 9. Complete Results of the Actual Implementation

Figure 9 shows the complete real-world implementation results of the Ubiquiti PowerBeam 5AC 620 application over a distance of 5.20 kilometers from the signal transmitter in Masbagik District to the signal receiver in Pringgasela District. At the signal transmitter, a throughput capacity of 484.38 Mbps was achieved, with a TX power of 24 dBm, a Channel Width of 80 MHz at a frequency of 5770 MHz, a received signal strength of -55 dBm, a noise floor of -93 dBm, and a local RX data rate of 6x (64QAM MIMO). Meanwhile, at the signal receiver, a throughput capacity of 336.96 Mbps was achieved, with a TX power of 24 dBm, a Channel Width of 80 MHz at a frequency of 5770 MHz, a received signal strength of -56 dBm, a noise floor of -93 dBm, and a local RX data rate of 6x (64QAM MIMO).

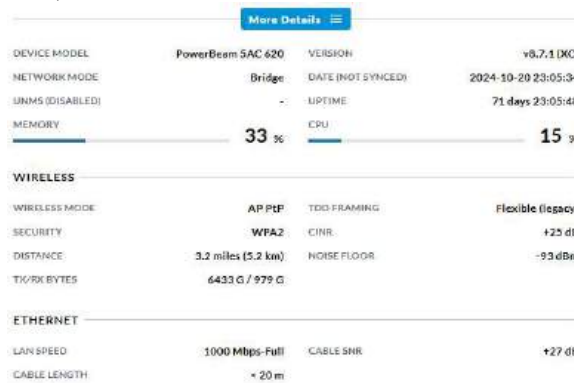


Figure 10. Detailed Information on Signal Transmitters

Figure 10 shows the device model used is PowerBeam 5AC 620, the network mode used is bridge, memory usage is 33%, CPU usage is 15%, the wireless mode used is AP PtP, the wireless security applied is WPA2, the distance is 5.2 kilometers, the noise floor is -93dBm, the LAN speed is 1000 Mbps-full duplex, and the LAN cable length is less than 20 meters.

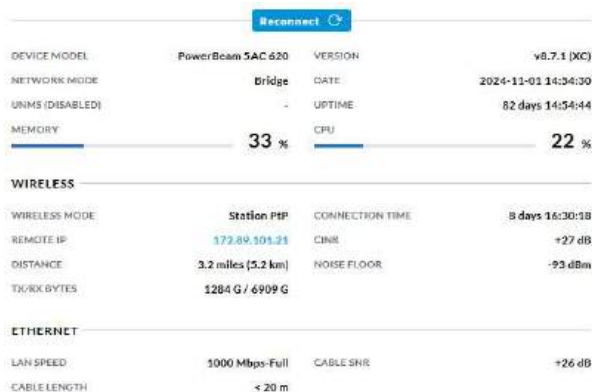


Figure 11. Detailed Information on the Signal Receiver

Figure 11 shows the device model used is PowerBeam 5AC 620, the network mode used is bridge, memory usage is 33%, CPU usage is 22%, the wireless mode used is Station PtP, the wireless security implemented is WPA2, the IP address for remote access is 172.89.101.21, the distance is 5.2 kilometers, the noise floor is -93dBm, the LAN speed is 1000 Mbps-full duplex, and the LAN cable length is less than 20 meters.

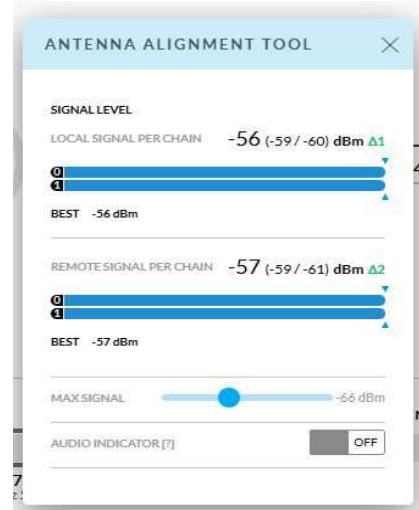


Figure 12. Measuring Signal Alignment Performance

Figure 12 shows the measurement of signal alignment performance from the signal transmitter to the signal receiver or vice versa, where the signal quality at the local signal is -56 dBm (for horizontal -59 dBm and vertical -60 dBm) while the signal quality at the remote signal is -57 dBm (for horizontal -59 dBm and vertical -61 dBm).

C. Operate

During the operation and evaluation phase, testing was conducted by sending duplex data packets (simultaneous download and upload), receive data packets (upload), and transmit data packets (download) for 30 seconds each, as follows:



Figure 13. Duplex Upload and Download Testing

Figure 13 shows the duplex test results over 30 seconds with a ping of 1.06 ms, achieving an average upload speed of 99.96 Mbps and an average download speed of 141.33 Mbps. Furthermore, the highest upload speed achieved was 151.05 Mbps and the highest download speed was 201.57 Mbps.

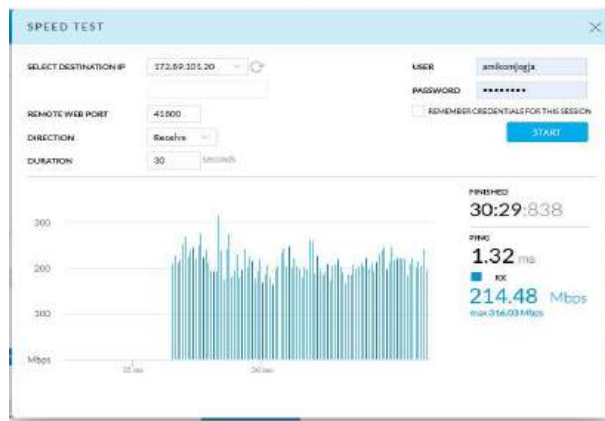


Figure 14. Download Testing

Figure 14 shows a download test lasting 30 seconds with a ping of 1.32 ms, resulting in an average download speed of 214.48 Mbps. Furthermore, the highest download speed achieved was 316.03 Mbps.

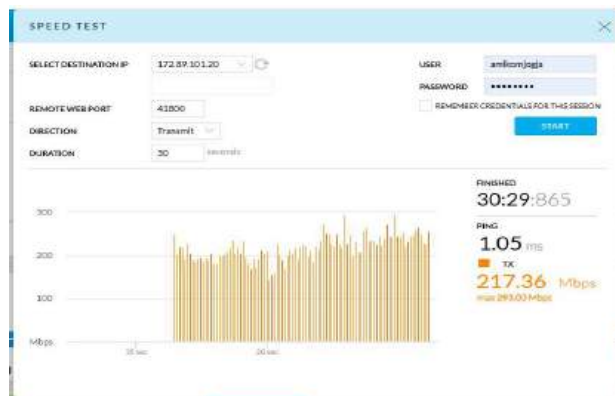


Figure 15. Upload Testing

Figure 15 shows the upload test results over 30 seconds with a ping of 1.05 ms, achieving an average download speed of 217.36 Mbps. In addition, the highest download speed reached 293.03 Mbps.

Table 1. Simulation Testing with Real-World Implementation

Testing	Simulation	Real-world Implementation
Total Capacity Download	-	316,03 Mbps
Total Capacity Upload	-	293,03 Mbps
Total Capacity both (Download & Upload)	497 Mbps	201,57 Mbps
Signal Strength at the TX Transmitter	-57 dBm	-56 dBm
Signal Strength at the RX Receiver	-57 dBm	-57 dBm
Clearance Fresnel Zone	Clear	Clear
Noise Floor	Noise Free	-93 dBm
Frequency	5.8 Ghz	5.8 Ghz
Technology	AirMAX AC	AirMAX AC
Channel Width	80 Mhz	80 Mhz

Table 1 shows a comparison of the network simulation results and the actual implementation in the field. The actual implementation shows fairly good results, namely 316.03 Mbps for download and 293.03 Mbps for upload. When combined (download and upload), the simulation obtained a total capacity of 497 Mbps, while the actual implementation only reached 201.57 Mbps, indicating a decrease in performance when applied in real-world conditions, which is likely influenced by environmental factors and interference.

In terms of signal quality, both the simulation and implementation showed relatively consistent results, namely -57 dBm at the transmitter (TX) and receiver (RX) for the simulation. The real-world implementation also showed results close to ideal, namely -56 dBm for TX and -57 dBm for RX. These values can be categorized as good and stable signals for point-to-point (PtP) communication. In both the simulation and real-world implementation, the Fresnel Zone Clearance was declared Clear, meaning the radio transmission path was not obstructed by physical objects and was in optimal LOS (Line of Sight) conditions. For the Noise Floor parameter, the simulation was in a Noise Free condition, while the real-world implementation recorded a value of -93 dBm, which is still considered good because a lower noise floor value (closer to -100 dBm) indicates a cleaner environment with less interference. Technical parameters such as frequency (5.8 GHz), AirMAX AC technology, and 80 MHz Channel Width were identical in both the simulation and real-world implementation, so the differences in results were more influenced by real-world conditions in the field, not network configuration.

IV. CONCLUSION

This research successfully comprehensively evaluated the accuracy of Fresnel Zone simulations using AirLink software against the actual implementation of a Point-to-Point (PtP) wireless network based on the IEEE 802.11ac standard using Ubiquiti PowerBeam 5AC 620 devices. Based on the test results, it was found that the simulation provided performance predictions that were quite close to field conditions, especially in terms of signal strength (RX) and Fresnel Zone clearance, which remained clear, thus ensuring optimal Line of Sight (LOS). However, there was a significant difference in throughput achieved, where the simulation predicted a capacity of 497 Mbps, while the actual field implementation only reached 201.57 Mbps for duplex testing. This proves that environmental factors, terrain contours, noise floor, and external interference still have a significant impact on network performance that cannot be fully represented by the simulation. Therefore, AirLink simulation is very helpful in the implementation of outdoor wireless networks, but it cannot be used as an absolute reference without field validation. The real-world evaluation proves the importance of optimizing technical parameters such as channel width, antenna height, azimuth/tilt angle, and noise floor monitoring to obtain a stable and efficient link quality. The PPDIOO method used in this research proved effective in ensuring that every stage of network design and optimization proceeded systematically and according to implementation needs. The

limitations of this research include the scope of the location, which only involved one PtP link and did not test performance under more complex topographical conditions. Further research is recommended to test the influence of variations in distance, weather, and the use of MIMO technology to obtain a more precise predictive model. Overall, this research provides an important technical reference for network practitioners in validating simulation results and designing PtP connections more accurately and practically in real-world field conditions.

REFERENCES

- [1] A. T. Haryanto, "APJII: Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang," *detikInet*, 2024. <https://inet.detik.com/cyberlife/d-7169749/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang> (accessed Nov. 11, 2025).
- [2] A. S. Tanenbaum and D. J. Wetherall, *Computer Networks*, 5th editio. Boston: Pearson, 2013.
- [3] cisco.com, "What Is 802.11ac?," <https://www.cisco.com/https://www.cisco.com/site/us/en/learn/topics/networking/what-is-802-11ac.html> (accessed Nov. 11, 2025).
- [4] T. S. Rappaport, *Komunikasi Nirkabel: Prinsip dan Praktik*, Edisi ke-2. Jakarta: Erlangga, 2014.
- [5] Kumar, S. and A. Sharma, *Antena dan Propagasi Gelombang Radio*. Yogyakarta: Andi Publisher, 2019.
- [6] A. Mutmainnah and R. Hidayat, *Teknologi Jaringan Nirkabel*. Yogyakarta: Andi Publisher, 2020.
- [7] Suji, *Transmisi Radio Microwave*. Jakarta Selatan: Nec Pasolink. PT. Alita Praya Mitra, 2013.
- [8] A. A. Pangera, *Menjadi Administrator Jaringan Nirkabel*. Yogyakarta: Andi Publisher, 2008.
- [9] B. A. Forouzan, *Data Communications and Networking*. New York: McGraw-Hill, 2013.
- [10] B. Sklar, *DIGITAL COMMUNICATIONS: Fundamentals and Applications*, 2nd ed. Los Angeles, USA: Prentice Hall, 2001. [Online]. Available: https://www.mangoud.com/EENG373_files/Book-Sklar.pdf
- [11] Citraweb.com, "Channel Width dengan Wireless N Mikrotik," *citraweb.com*, 2025. <https://citraweb.com/artikel/144/> (accessed Nov. 11, 2025).
- [12] W. Stallings, *Data and Computer Communications*, 10th ed. Pearson, 2014.
- [13] J. Han, "Application of Fresnel Zone Model in Wireless Sensing Field," *Highlights Sci. Eng. Technol. EMIS 2024*, vol. 119, pp. 516–523, 2024.
- [14] N. C. Nnadi, H. A. Jacob, and G. N. Ezech, "LINK BUDGET ANALYSIS FOR LINE OF SIGHT WIRELESS COMMUNICATION LINK WITH KNIFE EDGE DIFFRACTION OBSTRUCTION," *Sci. Technol. Publ. (SCI TECH)*, vol. 4, no. 6, pp. 369–374, 2020.
- [15] M. R. Hidayat, T. I. M. Fauzian, E. Alimudin, and H. Yuliana, "Analisis Power Link Budget Pada Jaringan Wireless Outdoor Menggunakan Internet Service Provider Design Center Studi Kasus Desa Kutanagara Garut," *TELEKONTRAN*, vol. 10, no. 2, 2022.
- [16] A. M. Ramadhan and D. W. Chandra, "Analisis Fresnel Zone dan Link Budget Jaringan WLAN dari PT. Solo Jala Buana ke BTS Karanganyar Menggunakan UBNT AIRLINK," Universitas Kristen Satya Wacana, 2024.
- [17] L. H. Anisa, P. Maria, S. Yusnita, and Y. L. Oktapianus, "Analisis Implementasi IEEE 802.11ac pada Jaringan Wireless Backhaul Distribusi Provider Zettalink Analysis of the Implementation of IEEE 802.11ac on the Wireless Backhaul Network of Zettalink Provider," *Electron J. Ilm. Tek. Elektro*, vol. 5, no. November, 2024.
- [18] S. A. Wirawan, A. N. Tahir, and Sutisna, "Analysis of Management Network Using Ubiquiti for Local Office at PT Rizki Inti Madani," *Int. J. Eng. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 243–247, 2025.
- [19] D. Teare, *Designing for Cisco internetwork solutions (DESGN)*. 2008. [Online]. Available: <http://www.loc.gov/catdir/toc/ecip0724/2007032855.html>



Accuracy Comparison of Support Vector Machine, Random Forest, and K-Nearest Neighbors for Sundanese Speech Classification

Laela Nur Rohmah^{1*}, Abdul Halim Anshor², Wahyu Hadikristanto³

¹Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

²Program Studi Sistem Informasi, Fakultas Teknik, Universitas Pelita Bangsa

³Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

Email: ¹laelanurrohmah85@gmail.com, ²abduhalimanshor@pelitabangsa.ac.id,
³wahyu.hadikristanto@pelitabangsa.ac.id

Abstract – To support the preservation of the Sundanese language, speech recognition systems based on machine learning can be developed. This study aims to evaluate and compare the classification performance of Support Vector Machine, Random Forest, and K-Nearest Neighbors which represent margin-based, ensemble-based, and distance-based classification approaches that have been widely applied in speech classification tasks. A secondary dataset consisting of 100 voice recordings was utilized in this research. The study followed the Knowledge Discovery in Database framework, which encompasses data selection, preprocessing, transformation, data mining, and evaluation phases. Feature extraction was performed using the Mel-Frequency Cepstral Coefficients method. Experimental results demonstrate that the Random Forest algorithm achieved superior performance, reaching 100% accuracy and an Area Under Curve (AUC) of 100%. Meanwhile, K-Nearest Neighbors achieved 87% accuracy with an AUC of 100%, and Support Vector Machine yielded the lowest performance with 67% accuracy and an AUC of 72.89%. Although Random Forest achieved the highest metrics, further research is required as a perfect 100% score raises concerns regarding model overfitting. To address this issue, utilizing a large dataset is recommended for future studies. Consequently, K-Nearest Neighbors can be considered a more reliable choice in this study, demonstrating robust and stable performance for MFCC based speech classification on smaller dataset.

Keywords – *Speech Classification, Sundanese Language, Support Vector Machine, Random Forest, K-Nearest Neighbors, MFCC, Machine Learning.*

I. INTRODUCTION

Indonesia is a country rich in linguistic diversity, possessing a total of 718 indigenous languages spread across its region. These local languages serve not only as communication tools but also as vital cultural heritages that reflect regional identity. However, these languages currently face serious threats of extinction. According to the United Nations (UN) and UNESCO, one indigenous language disappears every two weeks due to the declining number of active speakers capable of maintaining its usage [1].

The Sundanese language, a major regional language in West Java, has experienced a notable decline in the number of speakers. According to data from the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek), referring to the 2010-2020 Central Bureau of Statistics (BPS) census, indicates a significant decline in the use of the Sundanese language [2]. Although the Language Development and Fostering Agency has initiated various revitalization programs, their long-term sustainability remains heavily dependent on the community's commitment to daily usage.

Technological advancement offers new opportunities for language preservation. Artificial Intelligence has significantly contributed to various sectors, including healthcare, education, industry, and information systems [3][4]. One of the important implementations of artificial intelligence is speech recognition systems, which enable machines to understand, recognize, and respond to verbal

instructions from users. This technology is widely utilized in virtual assistants, navigation systems, Internet of Things (IoT) devices, and automated service systems [5].

Nevertheless, the development of current speech recognition technology remains focused on major global languages such as English, Mandarin, and Spanish. Local dialects like Sundanese, particularly the West Sundanese and South Sundanese dialects, have received disproportionately less attention in speech recognition development [6]. Current models face limitations in distinguishing acoustic variations between dialects, resulting in low accuracy and a lack of inclusivity in voice-based systems [7][8]. This condition presents a challenge in creating speech recognition systems that can accurately accommodate linguistic diversity.

Machine learning is a branch of artificial intelligence that plays a crucial role in developing voice classification systems. Classification is the process of grouping data based on specific patterns to distinguish characteristics between objects [9][10]. In speech recognition, machine learning is used to identify voice patterns and distinguish dialect characteristics based on features. Various machine learning algorithms can be used in classification, such as Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN). Each algorithm has different characteristics, advantages, and disadvantages in handling voice data.

Support Vector Machine (SVM) is a supervised learning algorithm that functions by identifying the optimal hyperplane to separate data classes [11]. SVM has



advantages in handling complex data and producing high accuracy, especially for high dimensional data[12]. Random forest (RF) is an ensemble learning algorithm that construct multiple decision trees to produce more stable and accurate prediction[13]. It is highly effective in managing large dataset, handling missing values, and performing automated feature selection. Meanwhile, K-Nearest Neighbors 9KNN0 is a non-parametric algorithm that classifies data based on proximity using Euclidean Distance[14][15]. KNN is recognized for its simplicity, ease of implementation, an effectiveness in pattern similarity based classification[16].

In speech classification research, features extraction is an important stage. One of the most widely adopted methods is Mel-Frequency Cepstral Coefficients (MFCC), which is a technique that transform raw audio signals into numerical representations relevant to human auditory perception[17]. MFCC effectively captures essential voice signal characteristics making it highly suitable for machine learning based speech recognition[18].

Previous studies have explored the classification of Sundanese speech using machine learning. Anshor and Wiyatno[19] compared Random Forest and SVM for Sundanese dialect classification using MFCC features and ROC-AUC evaluation. Similarly, Shandy et al [20] utilized KNN for MFCC based Sundanese dialect classification. However, this studies were limited to the evaluation of only one or two algorithm and lacked a comprehensive comparison between SVM, Random Forest, and KNN on the same dataset

Based on this background, this study aims to perform a accuracy comparison of Support Vector Machine, Random Forest, and K-Nearest Neighbors for Sundanese speech classification. The research utilizes a public dataset titled “Suara Dialek Sunda”, containing recordings of West and South Sundanese dialect. Following feature extraction via MFCC, the three algorithms are tested and compared based on accuracy, precision, recall, f1-score, and ROC-AUC metrics.

The result of this research are expected to contribute to identifying the most effective classification method for Sundanese speech recognition and support the development of indigenous language-based speech technology. In addition, this study is also expected to become one of the efforts to utilized technology in preserving local culture in the digital era.

II. RESEARCH METHODOLOGY

The research methodology follows a data mining approach based on the Knowledge Discovery in Database (KDD) framework. This method was selected for its systematic stages in data processing, beginning with data selection and proceeding through preprocessing, transformation, modeling, to the evaluation of classification results[21]. The KDD framework is particularly suitable for this research, as raw audio data requires a comprehensive feature extraction process before it can be effectively utilized by machine learning models.

This research aims to compare the performance of three classification algorithm, Support Vector Machine (SVM), Random Forest, and K-nearest Neighbors (KNN), in

classifying Sundanese speech based on the West Sundanese and South Sundanese dialect. the main focus is to identify which classification method yields the highest accuracy for Sundanese speech recognition.

The research is use a public dataset titled “Suara Dialek Sunda” which is available on the Kaggle platform. The dataset contains 100 voice sample in .flac format representing two dialect classes, West Sundanese dialect and South Sundanese dialect. this data was selected because it is relevant to the research objective of identifying differences in voice characteristics based on dialect.

The overall research flow can be illustrated through the flowchart in Figure 1. Each step is designed to ensure the research process runs systematically and objectively, in orde to obtain accurate and accountable results.

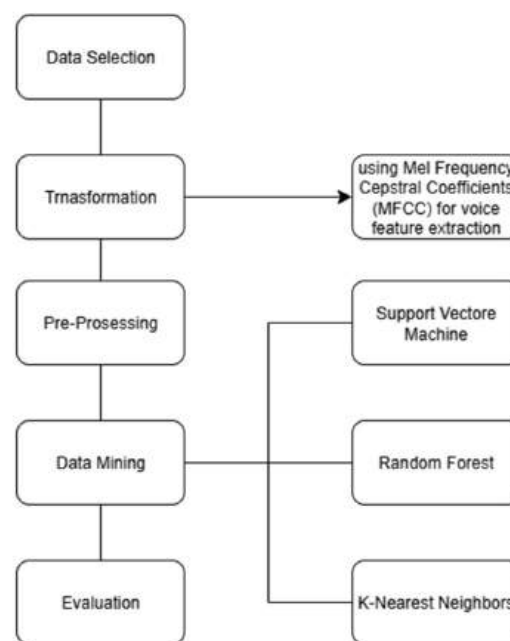


Figure 1. Research flow diagram

The detailed explanation for each stage in the research flowchart is as follows:

1. **Data selection:** This first stage is selection and collection data of Sundanese voice recording. The dataset comprises 100 audio files obtained from a public dataset on the Kaggle platform.
2. **Transformation (Feature Extraction):** This stage is conducted to transform the data into a format suitable for the research requirements. In this research, raw audio data is converted into a numerical representations. The Mel-Frequency Cepstral Coefficients (MFCC) method is utilized at this stage due to its capability to capture the essential characteristics of speech signal. The output of this transformation will be divided into training data and testing data, which are subsequently processed in the machine learning classification phase. The MFCC feature extraction process consists of several sequential stages, as follows[18]:

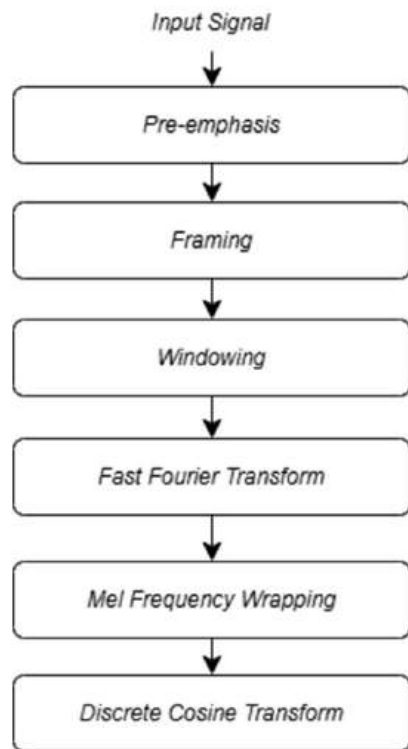


Figure 2. MFCC process flowchart

- a) **Pre-emphasis:** is a filtering process aimed at smoothing the spectral form of the speech signal and reducing noise interference during signal acquisition. This process is defined in equation (1).

$$y[n] = x[n] - ax[n - 1] \quad (1)$$

Description:

$y[n]$ = sinyal output

$x[n]$ = sinyal input

a = konstanta fitur dengan karakteristik $0.9 < a < 1.0$

- b) **Farming:** Framing, or Frame Blocking, is the process of segmenting the continuous signal into small, overlapping sections. The speech signal is converted into several short-duration frames to ensure the signal remains quasi-stationary.

- c) **Windowing:** Windowing involves applying a weighting function to each frame generated in the previous stage. This process aims to minimize signal discontinuities at the beginning and end of each frame caused by frame blocking. The equation used for the windowing stage is as follows (2).

$$x(n) = x_i(n)w(n), \quad n = 0, 1, 2, \dots, N \quad (2)$$

Description:

$x(n)$ = Signal value after windowing

x_i = Signal value from the i -th frame

$w(n)$ = Windowing function (typically Hamming window)

N = sample index within each frame

- d) **Fast Fourier Transform (FFT)**

FFT is widely used in digital signal processing to convert a signal from the time domain to the frequency domain. The mathematical formula used is as follows (3).

$$S(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N} \quad (3)$$

Description:

$S(k)$ = k -th FFT calculation result

$x(n)$ = n -th windowing output

k = Frequency index $(0, 1, 2, \dots, N)$

N = Number of samples to be processed

- e) **Mel-Frequency Wrapping:** This stage involves a filtering process to determine the energy magnitude of specific frequency bands within the speech signal. The power spectrum is passed through a Mel filter bank, which consists of several triangular filters distributed non-linearly along the Mel frequency scale. The Mel scale is a logarithmic frequency scale that more closely aligns with human auditory perception.

- f) **Discrete Cosine Transform (DCT):** DCT is the final step in MFCC feature extraction. The fundamental concept of DCT is to decorrelate the Mel spectrum to produce a robust representation of local spectral properties. In this stage, the Mel spectrum values in the frequency domain are converted back to the time domain to obtain the coefficients. DCT is applied to the log-energy to compress the information into a set of cepstral coefficients. Typically, the first 12–13 coefficients are used for the final MFCC representation. The DCT process can be expressed in Equation (4).

$$c(k) = 2 \sum_{n=0}^{N-1} x(n) \cos \frac{\pi(2n+1)k}{2N} \quad (4)$$

Description:

$x(n)$ = Final MFCC coefficients

N = Number of desired coefficients

3. Pre-Processing

The objective of this stage is to improve data quality and produce a more accurate and efficient analytical model. In this study, pre-processing is conducted to handle inaccurate, incomplete, or inconsistent data, such as addressing missing values, reducing noise, and resolving data inconsistencies.

4. Data Mining

At this stage, classification models are training using three machine learning algorithms: Support Vector Machine, Random Forest, and K-Nearest Neighbors. Prior to the training process, the

preprocessed dataset is partitioned into training data and testing data with a 70:30 ratio. The training data is utilized to construct the model and learn the characteristic patterns of the sundanese dialect, while the testing data is used to evaluate the model classification performance on unseen data that was not involved in the training phases. This data splitting is implemented to measure the algorithms capability to recognize and distinguish sundanese dialect.

These three algorithms employed in this study utilize different classification approaches. Support Vector Machine (SVM) is a learning algorithm that functions by finding an optimal hyperplane with a maximum margin between classes, enabling effective classification in high-dimensional feature spaces. Random Forest is an ensemble classification algorithm that construct multiple decision tree from randomly selected subsets of data and feature. The final Random Forest prediction is determined through a majority voting mechanism, which contribute to enhancing model stability and reducing the risk of overfitting. Lastly, K-Nearest Neighbors (KNN) is an instance-based algorithm that perform classification based on the labels of the nearest neighbors using the euclidean distance metric. This three algorithm were selected to compare the performance of margin-based, ensemble-based, and distance-based classification approaches in identifying sundanese speech dialect based on acoustic feature extracted through the Mel-Frequency Cepstral Coefficients (MFCC) method.

5. Evaluation

This final stage is conducted to analyze and assess the performance of the constructed models. Each model is evaluated using the test data based on accuracy as the primary metric, while also considering precision, recall, and the F1-score[22]. Furthermore, this study employs the Receiver Operating Characteristic – Area Under Curve (ROC-AUC) to provide a detailed evaluation of the models' ability to distinguish between the two classes. The AUC value ranges from 0 to 1, where a value closer to 1 indicates superior classification performance.

The evaluation process is implemented using Google Colab, a cloud-based programming environment that supports Python for machine learning[23]. Google Colab is utilized for model training, classification testing, and the visualization of confusion matrices, classification reports, and ROC-AUC analysis. The results from all three algorithms are then compared to determine the most optimal method for Sundanese speech classification. The final conclusion is drawn from these experimental results to support the development of more accurate and efficient indigenous language-based speech recognition systems.

III. RESULTS AND DISCUSSION

Based on the classification models implemented using Google Collaboratory (Google Colab), model performance was analyzed using several evaluation metrics, including accuracy, precision, recall, f1-score, ROC-AUC, and confusion matrix analysis. Accuracy was utilized to measure the proportion of correctly classified samples, while precision, recall, and f1-score were employed to assess the classification quality for each dialect class. Additionally, ROC-AUC was used to evaluate the model capability to differentiate between dialect classes across various decision threshold. The comparative performance results of the three machine learning algorithm are presented in Table 1.

Table 1. Comparison of Machine Learning Algorithm Performance

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
SVM	67%	69%	60%	64%	72.89%
RF	100%	100%	100%	100%	100%
KNN	87%	100%	73%	85%	100%

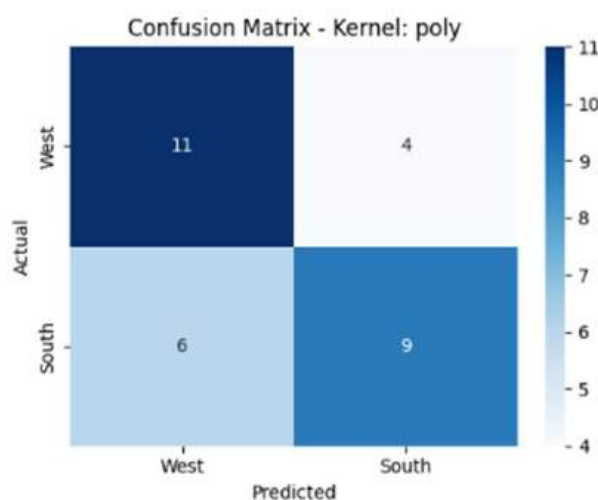


Figure 3. Confusion Matrix of Support Vector Machine

In figure 3, model SVM accurately classified 11 West Sundanese dialect samples correctly. However, 4 sample were misclassified as South Sundanese dialect. regarding the South dialect, model successfully identified 9 instances correctly, while 6 samples were misclassified as the West dialect. The confusion matrix results indicate that while the SVM model is capable of correctly identifying a portion of the dialect samples, significant misclassifications still occur across both classes. This finding demonstrates the model's limited capability to differentiate between Western and Southern Sundanese dialects, which aligns with its lower accuracy and ROC-AUC scores compared to those of Random Forest and K-Nearest Neighbors (KNN).

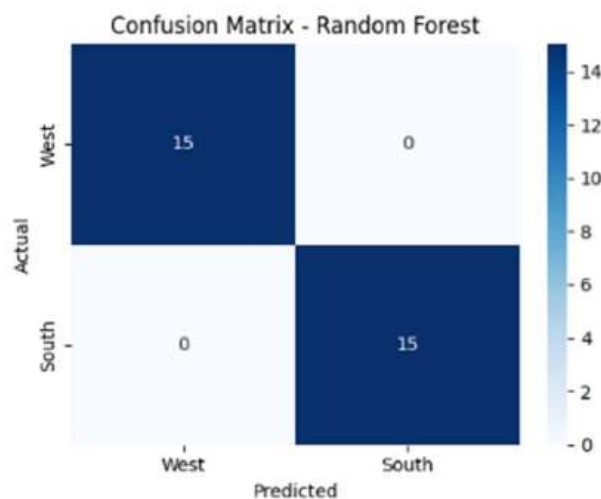


Figure 4. Confusion Matrix of Random Forest

In figure 4, model RF show that all data were successfully predicted correctly. Out of the 15 class data sample, all were accurately classified as west dialect. similarly for the 15 south class data sample, all were successfully predicted as south. These results indicate that the model's excellent capability to differentiate between Western and Southern Sundanese dialects based on the extracted MFCC features. Such flawless classification performance also aligns with the obtained accuracy and ROC-AUC scores, which are higher than those of SVM and K-Nearest Neighbors (KNN).

The following method is K-Nearest Neighbors. This algorithm works by classifying data based on the nearest distance between features. Within the KNN framework, the following steps are performed after the data splitting process:

1. Determining the k-value: There are no fixed rules for determining the k-value, in this research the k-value is set to 10.
2. Euclidean Distance Calculation: This step involves calculating the distance between the training data and the testing data. In this research, 15 training data and 1 testing data sample are utilized as a sample.

Table 2. Euclidean Distance Calculation

Data training	Euclidean Distance Calculation
D86	$\sqrt{((-254.00989) - (-163.97807))^2 + ((88.732315) - (34.690945))^2 + ((-40.110630) - (-27.213692))^2 + ((30.947512) - (22.054520))^2 + ((-1.890005) - (-13.354515))^2 + ((-2.493251) - (14.781217))^2 + ((7.844913) - (-12.342719))^2 + ((-10.974412) - (1.776014))^2 + ((1.200290) - (-10.514988))^2 + ((-7.606719) - (0.159154))^2 + ((-2.069135) - (-0.354056))^2 + ((1.658952) - (3.970661))^2 + ((-5.120046) - (-7.359786))^2} = 111.724$
D54	$\sqrt{((-326.57794) - (-163.97807))^2 + ((51.159830) - (34.690945))^2 + ((-29.254915) - (-27.213692))^2 + ((45.570140) - (22.054520))^2 +$

	$((-6.977783) - (-13.354515))^2 + ((1.615259) - (14.781217))^2 + ((8.016255) - (-12.342719))^2 + ((-11.957951) - (1.776014))^2 + ((8.280593) - (-10.514988))^2 + ((-6.925526) - (0.159154))^2 + ((-1.302201) - (-0.354056))^2 + ((4.698872) - (3.970661))^2 + ((-4.145867) - (-7.359786))^2 = 168.817$
....	
D92	$\sqrt{((-266.93310) - (-163.97807))^2 + ((74.309430) - (34.690945))^2 + ((-31.016413) - (-27.213692))^2 + ((42.155600) - (22.054520))^2 + ((-12.889493) - (-13.354515))^2 + ((3.149026) - (14.781217))^2 + ((12.862194) - (-12.342719))^2 + ((-11.102824) - (1.776014))^2 + ((7.251139) - (-10.514988))^2 + ((-3.911420) - (0.159154))^2 + ((-2.561542) - (-0.354056))^2 + ((-1.291445) - (3.970661))^2 + ((-1.718913) - (-7.359786))^2} = 117.987886$

Subsequently, the result of euclidean distance calculation are sorted in ascending order from the shortest to the farthest distance.

Table 3. Ascending Result of Euclidean Distance

Rank	X_train_index	Result of Euclidean Distance	Actual_Dialect_Name	K=10
1	33	9.520783	West	1
2	31	10.680736	West	2
3	39	19.087149	West	3
4	12	83.056314	West	4
5	62	90.482413	South	5
6	65	91.426549	South	6
7	11	95.208741	West	7
8	64	98.675702	South	8
9	9	100.898593	West	9
10	6	103.368218	West	10
11	86	111.724792	South	
12	83	114.864435	South	
13	92	117.987880	South	
14	72	137.177135	South	
15	48	144.635470	West	

Based on table 3, which present the prediction results from the calculation of one testing data sample using a value of k=10, identified 7 of the west dialect and 3 of the south dialect among the nearest neighbors. This indicates that the prediction result is consistent with the actual label of testing data, which is the west dialect.

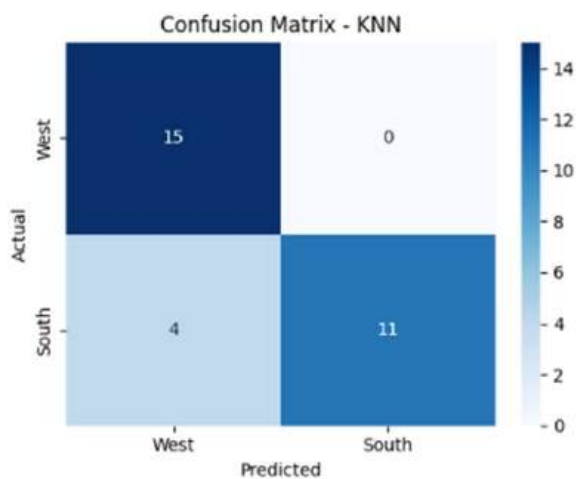


Figure 5. Confusion Matrix of K-Nearest Neighbors

In figure 5, model KNN show that successfully classified all 15 west dialect sample correctly. However for the south dialect, the model succesfully identified only 11 sample correctly. These results indicate that the KNN model has excellent performance in recognizing the west class, but it still encounters classification errors within in a porpotion of the south class data.

IV. CONCLUSION

Based on the experimental result of the three algorithms, random forest achieved the highest performance with accuracy and ROC-AUC of 100%. However achieved perfect accuracy and ROC-AUC score, the random forest model need to be analyzed futher since the dataset were used in this research relatively small. Such verification is essential to ensure that the model does not suffer from overfitting and maintains its generalizability to new unseen data.

In the second position the KNN model yielded an accuracy of 87% with an AUC of 100%. Although some misclassification occurred dureng testing, the AUC value suggest that the model has very strong performance because ROC-AUC evaluates not only the classification result at a single decision threshold but also the overall performance across various thresholds.

Finally, the SVM algorithm showed the lowest performance with an accuracy of 67% and an ROC-AUC of 72.89%. This indicates that the polynomial kernel utilized in this research was not yet optimal in mapping the MFCC features to effectively separate the classes. Overall, the comparison of these three algorithms demonstrates that the choice of algorithm significantly influences the success rate of speech classification models.

For the future research, it is recommended to use a larger speech dataset to more comprehensively evaluate the model generalizability to unseen data. The application of additional validation techniques is also recommended to obtain more reliable evaluation results and minimize the potential risk of overfitting. Futhermore, future studies could explore alternative speech feature extraction methods or compare the performance of the three algorithms used in this study with other classification model for sundanese dialect classification.

REFERENCES

- [1] E. Rusyana and R. U. N. Rohmah, "Interferensi bahasa Indonesia terhadap bahasa Sunda dalam karangan berbahasa Sunda siswa SMP," *Diglosia J. Kaji. Bahasa, Sastra, dan Pengajarannya*, vol. 7, no. 2, pp. 237–246, 2024, doi: 10.30872/diglosia.v7i2.954.
- [2] Pun, "UNESCO: Setiap Dua Minggu, Satu Bahasa Daerah Puna di Dunia," *Berita*. p. <https://jabarprov.go.id/berita/unesco-setiap-dua-m>, 2024.
- [3] M. Azhar and H. F. Pardede, "Klasifikasi Dialek Pengujar Bahasa Inggris Menggunakan Random Forest," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 439, 2021, doi: 10.30865/mib.v5i2.2754.
- [4] A. M. Afida, "Klasifikasi Jenis Burung Berdasarkan Suara Menggunakan Algoritme Support Vector Machine," *MATHunesa J. Ilm. Mat.*, vol. 8, no. 1, pp. 1–6, 2020, doi: 10.26740/mathunesa.v8n1.p1-6.
- [5] U. B. Nusantara, "SPEECH RECOGNITION DENGAN WHISPER DALAM BAHASA INDONESIA," vol. 9, no. 2, pp. 386–397, 2025.
- [6] M. A. As Sarofi, I. Irhamah, and A. Mukarromah, "Identifikasi Genre Musik dengan Menggunakan Metode Random Forest," *J. Sains dan Seni ITS*, vol. 9, no. 1, pp. 79–86, 2020, doi: 10.12962/j23373520.v9i1.51311.
- [7] G. Drougkas, E. M. Bakker, and M. Spruit, "Multimodal machine learning for language and speech markers identification in mental health," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, 2024, doi: 10.1186/s12911-024-02772-0.
- [8] T. Kessler, D. Schneider, and M. Shadaydeh, "A Comparison of Machine Learning Algorithms and Feature Sets for Automatic Vocal Emotion Recognition in Speech," 2022.
- [9] Desi Permatasari and Suliadi, "Algoritma Pemilihan Variabel untuk Klasifikasi dan Penerapannya pada Klasifikasi Desa-Kelurahan," *J. Ris. Stat.*, pp. 49–56, 2024, doi:

- 10.29313/jrs.v4i1.3876.
- [10] P. Rosyani, S. Saprudin, and R. Amalia, "Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 132, 2021, doi: 10.26418/justin.v9i2.44120.
- [11] Y. Supit, "Penerapan Algoritma Support Vector Machine (SVM) untuk Deteksi Penyakit Mata Katarak," *Indones. J. Data Sci.*, vol. 2, no. 2, pp. 87–96, 2025, doi: 10.30989/ijds.v2i2.1449.
- [12] S. N. Khairil Lukman, "Komparasi Algoritma Naïve Bayes dan SVM untuk Identifikasi Cyberbullying Selebriti di Media Sosial Twitter.pdf." 2025. doi: 10.33364/algoritma/v.22-1.2196.
- [13] N. I. Yaman, A. R. Juwita, S. A. P. Lestari, and S. Faisal, "Perbandingan Kinerja Algoritma Decision Tree dan Random Forest untuk Klasifikasi Nutrisi pada Makanan Cepat Saji," *Jurnal Algoritma*, vol. 21, no. 2, pp. 184–196, 2024. doi: 10.33364/algoritma/v.21-2.1649.
- [14] Widodo and Supatman, "Klasifikasi Transaksi Obat Puskesmas Bagelen Menggunakan Algoritma KNN," *J. Zetroem*, vol. 6, no. 1, pp. 79–85, 2024, doi: 10.36526/ztr.v6i1.3540.
- [15] A. Benita Suciani, G. Gunawansyah, and K. Maulana Adhinugraha, "Klasifikasi Jenis Batik Menggunakan Algoritma K-Nearest Neighbors (Knn)," *Infotronik J. Teknol. Inf. dan Elektron.*, vol. 9, no. 2, pp. 83–89, 2024, doi: 10.32897/infotronik.2024.9.2.3846.
- [16] V. Artanti, M. Faisal, and F. Kurniawan, "Klasifikasi Cardiovascular Diseases Menggunakan Algoritma K-Nearest Neighbors (KNN) Classification of Cardiovascular Diseases using K-Nearest Neighbors (KNN) Algorithm," *Techno.COM*, vol. 23, no. 2, pp. 467–479, 2024.
- [17] M. Oktaviani, T. E. Sutanto, and M. Mahmudi, "Klasifikasi Usia Berdasarkan Suara Dengan Ekstraksi Ciri Mel Frequency Cepstral Coefficients Menggunakan Support Vector Machine," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 4, no. 4, pp. 901–907, 2023, [Online]. Available: [https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/240%0Afiles/5264/Oktaviani et al. - 2023 - Klasifikasi Usia Berdasarkan Suara Dengan Ekstraksi.pdf](https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/240%0Afiles/5264/Oktaviani%20et%20al.%20-%20Klasifikasi%20Usia%20Berdasarkan%20Suara%20Dengan%20Ekstraksi.pdf)
- [18] R. Hidayat and A. Winursito, "A Modified MFCC for Improved Wavelet-Based Denoising on Robust Speech Recognition," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 1, pp. 12–21, 2020, doi: 10.22266/IJIES2021.0228.02.
- [19] A. H. Anshor and T. N. Wiyatno, "Comparative Analysis of Random Forest and Support Vector Machine for Sundanese Dialect Classification Using Speech Recognition Features," vol. 14, pp. 269–276, 2025.
- [20] E. Shandy, A. H. Anshor, and D. Ardiatma, "Implementation of Data Mining for Speech Recognition Classification of Sundanese Dialect Using KNN Method with MFCC Feature Extraction," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 6, no. 3, pp. 1145–1158, 2024, doi: 10.47709/cnahpc.v6i3.4226.
- [21] P. Angga, R. Herdiana, and D. Solihudin, "Klasifikasi menggunakan metode Naive Bayes.," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 8, pp. 456–460, 2023.
- [22] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [23] R. Andarsyah and A. Yanuar, "Sentimen Analisis Aplikasi Posaja Pada Google Playstore Untuk Peningkatan Pospay Superapp Menggunakan Support Vector Machine," *J. Tek. Inform.*, vol. 16, no. 2, pp. 1–7, 2024, [Online]. Available: <https://ejurnal.ulbi.ac.id/index.php/informatika/article/view/3533>

Usability Evaluation of the Yellow Cart Feature in the TikTok Shop Application Using the System Usability Scale (SUS)

Hikmah Sani Nadia^{1*}, Evy Nurmiati², Nuryasin³

¹Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta

²Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta

³Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta

Email: ¹hikmah.sani24@mhs.uinjkt.ac.id, ²evy.nurmiati@uinjkt.ac.id, ³nuryasin@uinjkt.ac.id

Abstract – TikTok Shop is a rapidly growing e-commerce platform in Indonesia, particularly through its Yellow Basket feature, which allows content creators to embed product links directly in videos. This feature has become essential within Indonesia's social commerce ecosystem. However, limited research specifically evaluates the usability of this feature from a general user perspective. This study aims to measure and evaluate the usability of TikTok Shop's Yellow Basket feature using the System Usability Scale (SUS) method. This method was selected because it is a widely validated and reliable instrument for assessing the perceived usability of interactive systems, providing a standardized measure that facilitates comparison across studies. Testing involved 50 active TikTok Shop users from various age groups, within four predefined age categories (17–21, 22–35, 36–45, and above 45 years). Data were collected through an online questionnaire consisting of 10 statements with a Likert scale and then analyzed using the standard SUS scoring formula. The evaluation yielded a mean SUS score of 71.30, categorized as “Good” according to the adjective rating scale. Dimension-level analysis identified efficiency and errors as the weakest dimensions, attributed to cognitive load induced by a concurrently active background video during catalog navigation. Thematic analysis of open-ended responses revealed three dominant concerns: pricing transparency issues, a rigid and inflexible user interface, and a perceived lack of practicality. Additionally, some of respondents reported usability difficulties suggesting an age-related usability gap not captured by the aggregate SUS score. Design recommendations are proposed to address the identified usability gaps and improve the overall user experience.

Keywords – Usability Testing, System Usability Scale, TikTok Shop, Yellow Basket, User Experience, Human Computer Interaction

I. INTRODUCTION

The rapid advancement of digital technology has fundamentally altered shopping behavior among Indonesian consumers. Social media platform have evolved by making features that provide business transactions [1]. These platforms make a new digital ecosystem, referred as social commerce. TikTok, one of the world's most widely used applications, has embraced this trend through TikTok Shop a feature that allows businesses to market their own product by seamlessly integrates short-form video content with real-time purchasing activity [2].

For their customers, TikTok Shops offers engaging and also interactive shopping experience [3]. One of feature from TikTok Shop ecosystem is the Yellow Cart feature, identifiable by a distinctive yellow shopping cart icon embedded within video and live streaming content. This feature allows users to add a product to their own cart directly while watching a short creative content without the need to search manually, thereby reducing customer journey and purchasing process [4]. Potential customer will find the transaction feels fun and this feature expected to be an easy feature and make it more convenient for customer [5].

TikTok Shop is one of the most known social media platform nowadays, much more in the younger generation [3]. Although is popularity, evaluating the usability of this feature becomes important to ensure user satisfaction and effectiveness. A good feature does not make people frustrated and tired while using the functionalities [6]. Usability is a measure of the quality of a user's experience interacting with a product or system, whether a website, software application, mobile technology, or other device.

According to ISO 9241-11 in 1998, usability is defined as the extent to which a product can be used by a specific user to achieve a specific goal with effectiveness, efficiency, and satisfaction in a specific usage context [7][8]. Nevertheless, the success of an application feature cannot be assessed solely by adoption rate. It must also be evaluated in terms of how effectively, efficiently, and satisfactorily it is used dimensions collectively defined as usability within the field of Human Computer Interaction (HCI). Nielsen [9] identified five core components of usability: learnability, efficiency, memorability, errors, and satisfaction. Evaluating these dimensions provides actionable insights for user centered design improvement [8].

The usability of mobile-based e-commerce features is increasingly evaluated using the System Usability Scale (SUS), a 10-item standardized questionnaire developed by Brooke in 1996 [10], that measures users' perceived ease of use through a 5-point Likert scale. SUS has been widely validated as a reliable, efficient, and technology-agnostic instrument applicable across diverse digital systems, including mobile applications, websites, and social commerce platforms [11]. Its low administration cost and straightforward scoring formula make it particularly suited for evaluating feature-level usability in consumer-facing applications. Purwandani [2] confirmed this applicability by applying SUS directly to the TikTok Shop platform, obtaining a score of 79.49, and demonstrating that SUS can effectively capture usability perceptions in a social commerce context. Building on this precedent, the present study applies SUS at a more granular level, specifically targeting the Yellow Cart feature as a distinct micro-level interaction point.



Several prior studies have employed the System Usability Scale (SUS) to evaluate social commerce and e-commerce platforms. Reynald and Harry [12] assessed the usability of the e-commerce checkout system using SUS and usability testing with total of 35 user as respondents, obtaining average SUS score of 69.57 (“Good”) indicating there some issue with its usability. Purwadani [2] applied SUS to perceived usability in TikTok Shop platform with total of 49 respondents, yielding a mean score of 70,49. These studies said TikTok grade scale it is quite good but furthermore need room of improvement, that shown with one of the instruments said the application is quite complicated. Both studies establish a relevant empirical baseline for social commerce usability evaluation using SUS.

A review of existing literature reveals, however, that usability evaluations of TikTok Shop have predominantly been conducted at the macro or general application level. No prior study has specifically examined the micro-level interaction of the Yellow Cart feature in particular, the cognitive load experienced by users when navigating the product pop-up catalog while a short form video continues to play in the background. In this dual-attention scenario, users must simultaneously manage continuous audio-visual stimuli and process product details (color, size, price) within a half-screen catalog interface. This represents a distinctive and under-examined usability challenge in the social commerce context.

This study addresses this gap by evaluating the usability of the Yellow Cart feature using the SUS method with 50 general TikTok Shop users. The findings aim to provide an objective, data driven basis for targeted design improvements to the feature.

II. RESEARCH METHODOLOGY

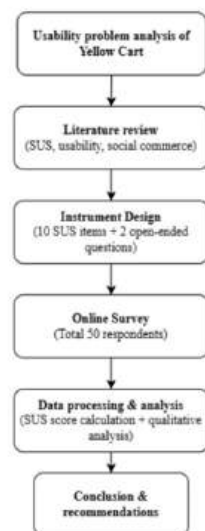


Figure 1. Flowchart Research Phase

The research initiated to analyze the perceived usability problems of the TikTok Yellow Basket feature, followed by a comprehensive literature review focusing on the System Usability Scale. In the third phase, the research instrument was designed, using the standard 10 item SUS questionnaire combine with 2 qualitative open-ended questions. Data collection was conducted through google

form by gathering responses from total 50 active TikTok Shop users with the following requirement. The collected data then analyze by using quantitative SUS scoring and by using qualitative analysis for the open-ended question. Finally, the study concluded with the evaluation of the features usability level and the formulation of actionable improvement recommendations

This study adopted a quantitative descriptive approach using a survey method. The primary evaluation instrument was the System Usability Scale (SUS), developed by Brooke [10]. SUS is a technology-agnostic, 10-item questionnaire rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). It has been widely validated as a reliable and efficient tool for usability measurement across diverse systems and applications.

For odd-numbered items (Q1, Q3, Q5, Q7, Q9), the contribution score is the respondent’s rating minus 1. For even numbered items (Q2, Q4, Q6, Q8, Q10), the contribution score is 5 minus the respondent’s rating. All 10 contribution scores are summed and multiplied by 2.5, yielding a final SUS score between 0 and 100. Interpretation follows the adjective rating scale by Bangor [13], presented in Table 1 [14].

Table 1. SUS Score Interpretation

SUS Score Range	Adjective Rating
> 90.9	Best Imaginable
80.8 – 90.9	Excellent
68.0 – 80.3	Good
63	Ok
51.0 – 67.9	Poor
< 51.0	Awful

The population consisted of TikTok Shop active user and had used the Yellow Cart at least once. A sample of 50 respondents was selected through purposive sampling based on the following criteria: (1) Active TikTok Shop user age above 17 years old; (2) had at least once click or used Yellow Cart feature; and last (3) willing to complete questionnaire. Data were collected via google form questionnaire distributed with link. Then open questions were given as a complementary instrument to the SUS questionnaire. Question was given for all of respondents ages above 17 years old, so we can determine the usability for different ages.

The SUS questionnaire administered to all respondents consisted of the following 10 standardized items adapted to the context of the Yellow Cart feature:

Table 2. SUS Questionnaire Items

No	Statement
Q1	I think I will be using this Yellow Basket feature a lot to search for items.
Q2	I find the flow from clicking the cart to the checkout page complicated.
Q3	In my opinion, this Yellow Basket feature is very easy to use.
Q4	I think I need someone else's help or to watch a tutorial first to be able to shop smoothly using this feature.
Q5	I feel this feature performs its function very well (from clicking the icon to the product details appearing)
Q6	I feel the shopping flow in the Yellow Basket is inconsistent.

Q7	I think other people will be able to learn to use this Yellow Basket feature very quickly.
Q8	I found the Yellow Basket feature confusing when I first tried it.
Q9	I feel very confident and have no obstacles when using this feature.
Q10	I feel like I have to get used to it or learn a lot of things first before I can shop smoothly using this feature.

Respondents also given an open-ended question, consist of 2 question that respondents could answer voluntary. (1) “What is the main problem or difficulty when using Yellow Cart feature?”. (2) “What improvement would you suggest to make Yellow Cart easier or more convenient to use?”. The inclusion of open-ended question and SUS instrument was intended to provide qualitative depth that explain why certain usability score lower.

The sample size of 50 respondents was determined based on established recommendations in the usability literature. Tullis and Stetson [15] demonstrated empirically that SUS achieves reliable conclusions in over 90% of cases with as few as 12–14 participants, establishing a clear minimum threshold. Lewis [16] further confirmed that SUS produces stable and consistent reliability estimates across varying sample sizes, making it one of the most robust usability instruments available. With 50 respondents, this study substantially exceeds the recommended minimum, ensuring that the SUS results are statistically stable and representative of the general Yellow Cart user population.

III. RESULTS AND DISCUSSION

From total of 50 respondents participated in this study. Majority were aged between 17-22 years old with a percentage of 70%. Followed by the age of 22-35 years old for 14%, people >45 for 10%, and last people ages around 36-45 years old for 6%. All of the respondents confirmed having click or used Yellow Cart feature at least once. This demographic profile is consistent with TikTok user base in Indonesia, affirming that the sample is representative of the general Yellow Cart user population.

With all the data gathered for SUS item across 50 respondents, together with the corresponding contribution scores computed using the standard SUS formula. The sum of all contribution scores and the resulting final SUS score are shown at the bottom of the table.

Table 3. SUS Result for Usability of Yellow Cart in TikTok Shop

Respondent	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	SUS Score
R1	3	3	4	3	3	3	4	3	4	3	82,5
R2	2	4	4	4	4	4	4	4	4	4	95
R3	3	4	4	4	3	2	2	3	4	4	82,5
R4	2	3	4	4	3	4	4	4	3	4	87,5
R5	3	2	3	2	3	1	3	3	3	3	65
R6	2	3	3	3	3	2	3	3	3	3	70
R7	2	1	4	1	2	3	0	3	2	1	47,5
R8	0	0	0	0	2	0	2	0	2	1	17,5
R9	4	4	4	4	4	4	4	4	4	4	100
...	...	...	...	...	...	...	...	...	...	...	...
R50	3	3	3	4	3	2	3	3	3	3	75
Total Average											71,3

With all the response from 50 participant, the mean SUS score of 71,3 falls within the “Good” category (68,0-80,7) and still in the “Acceptable” range according to Bangor [13]. This means that the Yellow Cart feature in TikTok Shop actually meets the baseline threshold of usability and it generally functional for the majority of users. However, the score also signal that the feature has not reach the optimum yet. That mean there is a possible area for improvement.

The SUS Score of 71,3 obtained in this study is closely comparable to existing findings in the social commerce usability literature. Purwadani [2] reported a score of 79,49 for the TikTok Shop application. Notably, however, the present study differs from both prior works in its level of analysis. Rather than evaluating an entire application, this study isolates a single, high-traffic feature and its specific micro-level interaction context. In other side, qualitative responses further revealed two categories of concern: interface design limitations and pricing transparency issues. The latter, while not directly measured by SUS, significantly impacts the satisfaction dimension, as discrepancies between advertised and actual prices erode user trust and increase perceived risk of purchase.

The most frequently reported theme was pricing transparency. Respondent describe discrepancies between displayed price and the actual price appearing in catalog pop-up. Further design recommendation a dynamic price synchronization mechanism should be implemented to ensure consistency between the price displayed in the product catalog and the final checkout price. By using real time database with the merchant pricing database, it can minimize discrepancies and also increase user trust. Also, the second theme reported that the user interface was rigid and inflexible. Several noted difficult to scrolling through product felt difficult and make some of the prefer to use another application. In addition, several participants found the interface visually cluttered, with excessive information increasing cognitive load, particularly for older or less digitally experienced users. Based on these findings, TikTok could further optimize the existing interface by improving navigation fluidity and simplifying information presentation. A clearer visual hierarchy and a less cluttered product display may reduce cognitive load and make the Yellow Basket feature more intuitive for users with diverse levels of digital experience

IV. CONCLUSION

This study evaluated the usability of the Yellow Cart feature in TikTok and TikTok Shop using the System Usability Scale (SUS). With the data from 50 general TikTok and TikTok Shop users as a respondent, the evaluation produced a mean SUS score of 71,3 placing the feature in the “Good” category within the “Acceptable” range. This result indicates that the feature in Yellow Cart is completely functional and satisfactory for general users, while simultaneously identifying concrete area for improvement.

Dimension-level analysis identified *errors* and *efficiency* as the two usability dimensions most in need of attention. These weaknesses are closely tied to the dual-attention cognitive load generated when users must



navigate the product catalog while the background video remains active a micro-level usability challenge that distinguishes this study from prior macro-level evaluations of TikTok Shop. *Satisfaction* and *learnability* were identified as the strongest dimensions, affirming the feature's core intuitiveness for its target demographic.

Beyond the quantitative SUS findings, qualitative responses highlighted two additional concerns with implications for the feature's broader usability. First, several respondents reported difficulties related to the user interface design of the catalog pop-up, citing cluttered layout and inconsistent visual elements as sources of confusion. Future studies are therefore recommended to conduct age-stratified usability analyses to determine whether the Yellow Cart feature meets inclusive design standards across a broader demographic spectrum. Experimental evaluations of interface simplification strategies could further identify design improvements that enhance navigation and overall user experience. Furthermore, future work should focus on developing a high-fidelity UI/UX prototype based on the recommendations proposed in this study.

REFERENCES

- [1] N. Putri, Y. Prasetya, P. W. Handayani, and H. Fitriani, "TikTok Shop: How trust and privacy influence generation Z's purchasing behaviors," *Cogent Soc. Sci.*, vol. 10, no. 1, 2023, doi: 10.1080/23311886.2023.2292759.
- [2] I. Purwandani, N. O. Syamsiah, and S. Nurwahyuni, "Perceived Usability Evaluation of TikTok Shop Platform Using the System Usability Scale," *Sinkron*, vol. 8, no. 3, pp. 1389–1399, 2023, doi: 10.33395/sinkron.v8i3.12473.
- [3] S. Dahniar, W. Anugra, A. Sakinah, W. Febrianti, and M. Hasan, "Utilization of TikTok Shop Interactive Features and Their Impact on Consumer Purchasing Decisions," *Int. J. Asian Bus. Manag.*, vol. 2, no. 6, pp. 947–960, 2023, doi: 10.55927/ijabm.v2i6.6891.
- [4] D. M. Tan, K. R. Alexia, and U. Bunda, "the Influence of Tiktok Affiliate Content Quality and Credibility on Purchase Decisions Via the Yellow Basket Feature," *JUMDER J. Bisnis Digit. dan Ekon. Kreat.*, vol. 1, no. 2, pp. 58–79, 2025.
- [5] S. Shasya and W. J. A, "Yellow Cart Feature on Tiktok as a Medium to Attract Purchase Interest," *J. Indones. Sos. Sains*, vol. 5, no. 07, pp. 1846–1853, 2024, doi: 10.59141/jiss.v5i07.1195.
- [6] A. Baj-Rogowska and M. Sikorski, "Exploring the usability and user experience of social media apps through a text mining approach," *Eng. Manag. Prod. Serv.*, vol. 15, no. 1, pp. 86–105, 2023, doi: 10.2478/emj-2023-0007.
- [7] I. Standard and T. S. Preview, "Iso 9241-11," vol. 1998, 1998.
- [8] M. Kurosu, "Usability, quality in use and the model of quality characteristics," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9169, pp. 227–237, 2015, doi: 10.1007/978-3-319-20901-2_21.
- [9] J. Nielsen, *Usability engineering*. Morgan Kaufmann, 1994.
- [10] J. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," *Usability Eval. Ind.*, pp. 207–212, 2020, doi: 10.1201/9781498710411-35.
- [11] A. M. Deshmukh and R. Chalmeta, "Validation of system usability scale as a usability metric to evaluate voice user interfaces," *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/peerj-cs.1918.
- [12] R. D. Pahlevi and H. B. Santoso, "A Usability Evaluation of an E-Commerce Checkout System Using SUS and Usability Testing," vol. 7, no. 1, pp. 160–173, 2026, doi: 10.59395/ijadis.v7i1.1498.
- [13] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean; adding an adjective rating," *J. usability Stud.*, vol. 4, no. 3, pp. 114–23, 2009.
- [14] H. R. Saeidnia, M. Karajizadeh, Z. Mohammadzadeh, S. Abdoli, and M. Hassanzadeh, "Usability Evaluation of the Mask Mobile Application: The Official Application of the Iranian Government," *Iran. J. Med. Microbiol.*, vol. 16, no. 1, pp. 49–55, 2022, doi: 10.30699/ijmm.16.1.49.
- [15] T. S. Tullis and J. N. Stetson, "A Comparison of Questionnaires for Assessing Website Usability



ABSTRACT : Introduction,” *Usability Prof. Assoc. Conf.*, pp. 1–12, 2004, [Online]. Available: <http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf>

- [16] J. R. Lewis and J. Sauro, “The Factor Structure of the System Usability Scale. In: Kurosu, M. (eds) *Human Centered Design. Lecture Notes in Computer Science.*,” *Lncs*, vol. 5619, pp. 94–103, 2009.



An AHP–TOPSIS-Based Decision Support System for Objective Pencak Silat Athlete Selection at KONI Lamongan

Ahmat Rendi Saputra¹ Brina Miftahurrohmah^{2(*)}

^{1,2}Departemen Sistem Informatika, Semen Indonesia International University (UISI)

Email: ¹ahmat.saputra22@student.uisi.ac.id ²brina.miftahurrohmah@uisi.ac.id

Abstract – Sports achievement development requires an objective and measurable athlete selection mechanism, particularly in regional sports organizations preparing for competitive events. However, the athlete selection process conducted by Komite Olahraga Nasional Indonesia (KONI) Kabupaten Lamongan for the Pencak Silat branch in preparation for PORPROV Jawa Timur IX 2025 still relied on subjective weighting, namely 40% physical test and 60% achievement certificate assessment, potentially causing decision bias. Therefore, this study developed a Decision Support System (DSS) integrating Analytic Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The combination was chosen because AHP determines consistent criterion weights from expert judgments, while TOPSIS effectively ranks alternatives based on their proximity to ideal solutions. The study involved 26 athlete candidates, nine physical performance criteria, and seven expert respondents. The results showed that Sprint 30m, Illinois Agility, Stork Stand, Standing Broad Jump, and Beep Test obtained the highest priority weights (14.94%). The Consistency Ratio (CR) was 0.0018, indicating reliable judgments. TOPSIS identified Sastro Hadi Saputro as the top-ranked athlete (Ci = 0.8954). System rankings also showed strong agreement with coach evaluations (Spearman = 0.9945; Pearson = 0.9863). These findings demonstrate that the AHP–TOPSIS framework improves transparency and reduces subjectivity in athlete selection.

Keywords – Decision Support System, AHP, TOPSIS, Athlete Selection, Pencak Silat, Multi-Criteria Decision Making.

I. INTRODUCTION

Sports achievement has become an important indicator in regional development, not only in relation to institutional prestige but also as part of long-term human resource development through structured athlete coaching [1];[2]. In this context, Komite Olahraga Nasional Indonesia (KONI) Kabupaten Lamongan carries a strategic responsibility in preparing competitive contingents for the East Java Provincial Sports Week (PORPROV) IX 2025. Among the prioritized sports, pencak silat occupies a significant position due to its strong athlete regeneration system and considerable medal potential [3]. The success of competitive sports development is closely associated with the quality of athlete selection, since the selection stage determines the composition of athletes who will represent the region in high-level competitions [4];[5].

Despite its strategic role, the athlete selection process conducted by KONI Kabupaten Lamongan still encounters several issues related to assessment objectivity and consistency. The current mechanism relies on two primary assessment components, namely physical performance tests weighted at 40% and achievement certificates weighted at 60%. However, the determination of these proportions remains largely based on managerial assumptions rather than on scientifically validated weighting procedures. This condition becomes increasingly problematic considering the diversity of athlete profiles involved in the selection process. The candidate pool consists of 25 athletes with varying levels of physical capability, competition experience, and achievement records, creating a complex decision-making environment that cannot be adequately addressed through conventional judgment alone.

Preliminary findings obtained through interviews with the Vice Secretary II of KONI Kabupaten Lamongan on 12

November 2025 revealed a fundamental weakness within the current selection mechanism. The dominance of achievement certificates in the assessment structure tends to disadvantage newcomer athletes who demonstrate outstanding physical performance but possess limited competitive experience [6]. Consequently, athletes with declining physical conditions may still receive higher evaluations solely because of previously accumulated certificates [7];[8]. Such conditions indicate that the existing weighting mechanism has not fully represented the actual performance potential required for current competitive demands. The issue of weighting imbalance therefore becomes an urgent concern, particularly in ensuring that the selection process remains adaptive, fair, and performance-oriented.

From an academic perspective, determining criterion weights within athlete selection is not merely a technical issue but also a substantial problem in decision-making and sports management studies [9];[10]. In multi-criteria decision environments, inappropriate weighting structures may generate biased recommendations and reduce the validity of final decisions [11]; [12]. This issue becomes increasingly critical in sports selection systems where assessment outcomes directly influence athlete opportunities and organizational strategies. Previous studies have also emphasized that subjectivity frequently emerges in athlete evaluations due to the large number of participants and the complexity of performance indicators, thereby requiring decision support mechanisms capable of improving consistency and transparency in the assessment process [13]; [14].

Several previous studies have explored the implementation of Decision Support Systems (DSS) using Multi-Criteria Decision Making (MCDM) approaches for athlete selection [15];[16]. Research demonstrated that the integration of Analytic Hierarchy Process (AHP) and



Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) was capable of improving objectivity in pencak silat athlete selection [17];[18]. Similar findings were also reported in studies utilizing TOPSIS for ranking athlete candidates in competitive sports environments [4]. Nevertheless, most previous studies were conducted in broader competition settings and did not specifically address the weighting imbalance between physical performance and achievement certificates within regional athlete development systems. In addition, earlier studies generally focused on technical ranking procedures without sufficiently accommodating the practical needs of local sports organizations facing disparities between experienced athletes and physically superior newcomers [10];[19]. These limitations indicate that further investigation is still required to develop a more adaptive and context-relevant weighting model for athlete selection at the regional level.

To address these issues, this study proposes the implementation of a Decision Support System based on the Multi-Criteria Decision Making (MCDM) approach through the integration of Analytic Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). AHP was selected due to its capability in decomposing complex decision problems into hierarchical structures and transforming expert judgments into measurable criterion weights. Previous studies have shown that AHP contributes to more controlled managerial processes and improves the quality of decision-making compared with conventional approaches [20]. Furthermore, TOPSIS was employed to rank athlete candidates based on their relative distances from positive and negative ideal solutions. The method is particularly suitable for competitive selection environments because it identifies alternatives with the closest proximity to ideal conditions while simultaneously maximizing their distance from undesirable conditions [17].

Accordingly, this study aims to develop a Decision Support System for pencak silat athlete selection in the preparation of PORPROV Jawa Timur IX 2025 by integrating AHP and TOPSIS methods. The AHP method is utilized to determine criterion weights systematically based on expert evaluations, while TOPSIS is employed to rank athlete candidates according to their performance proximity to ideal solutions. Through this approach, the study is expected to provide a more objective, transparent, and measurable athlete selection mechanism that can support strategic decision-making within KONI Kabupaten Lamongan.

II. RESEARCH METHODOLOGY

This study employed a mixed-methods approach using an exploratory sequential design to integrate qualitative expert judgment with quantitative computational analysis [21];[22]. The qualitative phase was conducted to identify relevant selection criteria and understand practical considerations within the athlete selection process at Komite Olahraga Nasional Indonesia (KONI) Kabupaten Lamongan. Subsequently, the quantitative phase was performed to calculate criterion

weights and rank athlete candidates using the integration of Analytic Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). This approach was selected to ensure that the resulting decision model was not only mathematically measurable but also empirically aligned with real conditions in regional athlete development.

The object of this study focused on the selection process of pencak silat athletes prepared for PORPROV Jawa Timur IX 2025. The alternatives evaluated in the system consisted of 25 athlete candidates from IPSI KONI Kabupaten Lamongan. In addition, seven experts were purposively selected as respondents in the weighting process, comprising representatives from KONI management, senior coaches, and the Monitoring and Evaluation (MONEV) team. The involvement of multiple expert groups was intended to obtain balanced judgments reflecting managerial, technical, and performance evaluation perspectives.

Data collection involved both primary and secondary data sources. Primary data were obtained through semi-structured interviews and pairwise comparison questionnaires based on the Saaty scale [23];[24]. Interviews were conducted to identify relevant selection criteria and to examine weaknesses in the existing selection mechanism [25];[26]. Meanwhile, the pairwise comparison questionnaires were distributed to experts to assess the relative importance among criteria and sub-criteria [27]. Secondary data consisted of athlete performance records, including physical test results and achievement certificate documents obtained from IPSI KONI Kabupaten Lamongan. The physical assessment criteria included strength, speed, power, endurance, agility, coordination, balance, and aerobic capacity, which were considered essential indicators in pencak silat performance evaluation. The criteria and sub-criteria used in the athlete selection process were determined based on expert discussions and physical assessment standards applied by IPSI KONI Kabupaten Lamongan. The details of the evaluation criteria are presented in Table 1 [28].

Table 1. Criteria and Attributes Used in Athlete Selection

Criteria	Sub-Criteria	Attribute	Measurement Scale	Data Source
Physical Test	Strength	Benefit	Strength score	Physical test results
Physical Test	Speed	Benefit	Time (seconds)	Physical test results
Physical Test	Power	Benefit	Power score	Physical test results
Physical Test	Endurance	Benefit	Endurance score	Physical test results
Achievement	Competition Certificate	Benefit	Competition level	IPSI/KONI documents

The identified criteria were subsequently processed using the AHP method to determine criterion weights before the TOPSIS ranking stage. The overall research procedure implemented in this study is illustrated in Figure 1.





Figure. 1 Research Framework of AHP-TOPSIS Athlete Selection System

Based on the research framework shown in Figure 1, the study began with problem identification and data collection, followed by the determination of criteria and sub-criteria for athlete evaluation. The AHP method was subsequently applied to obtain criterion weights, while TOPSIS was employed to generate athlete rankings and recommendation results.

The Analytical Hierarchy Process (AHP) method was applied to determine criterion weights systematically. The process began with the construction of pairwise comparison matrices based on expert judgments using the Saaty scale ranging from 1 to 9[29];[30]. Since multiple experts were involved, the judgments were aggregated using the geometric mean method to obtain a collective assessment. The matrices were subsequently normalized to calculate the priority weights of each criterion and sub-criterion [31]. To ensure logical consistency in expert judgments, consistency testing was conducted using the Consistency Index (CI) and Consistency Ratio (CR)[32]. The comparison matrix was considered acceptable when the CR value was less than or equal to 0.10, indicating that the judgments were sufficiently consistent for decision-making purposes [33]; [34]; [35].

After obtaining the criterion weights from AHP, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method was implemented to rank athlete candidates. The TOPSIS procedure began with the construction of a decision matrix representing athlete performance scores across all evaluation criteria. The matrix was then normalized to eliminate scale differences among criteria before being multiplied by the corresponding AHP weights to produce a weighted normalized matrix. Based on this matrix, positive ideal solutions and negative ideal solutions were identified to represent the best and worst possible conditions for each criterion [36]; [37]. Euclidean distance calculations were subsequently performed to measure the relative proximity of each athlete candidate to both ideal solutions. The final preference value was calculated to determine athlete rankings, where higher preference values indicated greater eligibility for recommendation in the PORPROV Jawa Timur IX 2025 preparation process.

The integrated AHP-TOPSIS framework was implemented within a web-based Decision Support System to facilitate data processing and recommendation generation. System evaluation was conducted by comparing the recommendation results produced by the

proposed model with the conventional selection mechanism previously applied by KONI Kabupaten Lamongan. This evaluation aimed to assess the extent to which the proposed system improved objectivity and consistency in athlete selection.

III. RESULTS AND DISCUSSION

This study implemented an integrated Analytic Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) framework to support the athlete selection process conducted by Komite Olahraga Nasional Indonesia (KONI) Kabupaten Lamongan in preparation for PORPROV Jawa Timur IX 2025. The proposed system was developed to overcome the limitations of the conventional athlete selection mechanism, which predominantly relied on subjective assessment and static weighting assumptions. Through the integration of Multi-Criteria Decision Making (MCDM) methods, qualitative expert judgment was transformed into a measurable and systematic computational model capable of generating more transparent and accountable selection recommendations.

A total of 26 athlete candidates consisting of 13 male athletes and 13 female athletes participated in the evaluation process. Athlete performance data were obtained from standardized physical tests conducted by KONI Lamongan, covering nine primary criteria, namely Push up (PU), Sit up (SU), Sprint 30m (S30), Illinois Agility (IA), Core Stability (CS), Stork Stand (SS), Standing Board Jump (SBJ), Handwall Toss (HT), and Beep Test (BT) [38]. These criteria were selected because they represent the dominant physical components required in pencak silat competitions, including muscular endurance, agility, balance, explosive power, coordination, speed, and aerobic endurance.

The first stage of analysis involved determining criterion weights using the AHP method. Seven expert respondents consisting of KONI management representatives, senior coaches, and Monitoring and Evaluation (MONEV) personnel participated in the pairwise comparison assessment process. The resulting judgments were aggregated using the geometric mean approach to obtain a collective organizational perspective regarding the strategic importance of each physical criterion. The pairwise comparison matrix generated from expert assessments is presented in Table 2 [38].



Table 2. Criteria and Attributes Used in Athlete Selection

Criteria	PU	SU	S30	IA	CS	SS	SBJ	HT	BT
Push up (PU)	1	1	0,33	0,33	0,5	0,33	0,33	0,5	0,33
Sit up (SU)	1	1	0,33	0,33	0,5	0,33	0,33	0,5	0,33
Sprint 30m (S30)	3	3	1	1	2	1	1	2	1
Illinois Agility (IA)	3	3	1	1	2	1	1	2	1
Core Stability (CS)	2	2	0,5	0,5	1	0,5	0,5	1	0,5
Stork Stand (SS)	3	3	1	1	2	1	1	2	1
Stand Board Jump (SBJ)	3	3	1	1	2	1	1	2	1
Handwall Toss (HT)	2	2	0,5	0,5	1	0,5	0,5	1	0,5
Beep Test (BT)	3	3	1	1	2	1	1	2	1

The matrix analysis demonstrates a strong tendency among experts to prioritize dynamic physical performance indicators over static muscular endurance variables. Criteria associated with speed, agility, balance, explosive power, and aerobic endurance consistently received higher relative importance values compared with Pushup and Situp. This finding indicates that pencak silat athlete performance is perceived to depend primarily on explosive movement capability, rapid directional transitions, cardiovascular endurance, and body stability rather than isolated upper-body strength alone.

From a sports science perspective, these findings are highly relevant to the characteristics of pencak silat competitions, which require athletes to maintain mobility, reaction speed, and endurance during continuous combat situations. Consequently, the weighting structure generated through expert judgment reflects the practical demands of competitive performance rather than merely emphasizing general physical fitness indicators [29];[39]. Based on the pairwise comparison matrix, normalization and eigenvector calculations were subsequently performed to determine the final priority weights of each criterion. The resulting criterion weights are presented in Table 3 [40].

Table 3. Priority Weights of Athlete Selection Criteria

Numb	Criteria	Weight (W)	Persentase (%)
1	Sprint 30m (/s)	0,1494	14,94%
2	Illinois Agility (/s)	0,1494	14,94%
3	Stork Stand	0,1494	14,94%
4	Standing Board Jump (cm)	0,1494	14,94%
5	Beep Test	0,1494	14,94%
6	Core Stability (/s)	0,0797	7,97%
7	Handwall Toss	0,0797	7,97%
8	Push up	0,0467	4,67%
9	Sit up	0,0467	4,67%
Total		1	100%

The results in Table 3 indicate that Sprint 30m, Illinois Agility, Stork Stand, Standing Board Jump, and Beep Test obtained the highest priority values, each contributing 14.94% to the overall evaluation model. The dominance of these criteria confirms that the experts considered speed, agility, balance, explosive power, and cardiovascular endurance as the primary determinants of athlete competitiveness within the KONI Lamongan pencak silat selection framework.

In contrast, Push up and Sit up received the lowest weights at 4.67%, suggesting that muscular endurance of the upper body and abdominal region was considered less decisive in comparison with dynamic movement-oriented variables. Scientifically, this weighting structure demonstrates that athlete evaluation in combat sports requires multidimensional physical analysis emphasizing movement efficiency and physiological sustainability rather than isolated muscular capacity alone.

The validity of the weighting process was further evaluated through consistency testing. The AHP computation produced a maximum eigenvalue (λ_{max}) of 9.0205, a Consistency Index (CI) value of 0.0026, and a Consistency Ratio (CR) value of 0.0018. Since the resulting CR value was substantially lower than the acceptable threshold of 0.10, the expert judgments were considered logically consistent and mathematically reliable. This result is important because inconsistency in pairwise comparisons may reduce the stability and credibility of decision-support recommendations. Therefore, the low inconsistency value obtained in this study strengthens the validity of the proposed weighting model for athlete selection applications.

After obtaining the criterion weights, the TOPSIS method was applied to evaluate and rank all athlete candidates. TOPSIS was selected because of its capability to identify alternatives possessing the shortest distance from the positive ideal solution and the farthest distance from the negative ideal solution simultaneously. Within the context of athlete selection, this approach enables the system to identify athletes demonstrating balanced overall performance rather than athletes who excel only in isolated criteria. The TOPSIS ranking results for the ten highest-performing athlete candidates are presented in Table 4.

Table 4. TOPSIS Ranking Results of Athlete Candidates

Numb	Athlete's Name	Gender	D+	D-	Preference Value Ci	Rank
1	Sastro Hadi Saputro	Male	0,002	0,0175	0,8954	1
2	Hegar Marga Birawa	Male	0,0022	0,0173	0,8886	2
3	Bagus Prayetno	Male	0,0024	0,017	0,8749	3
4	M. Shofar Habibur Rochman	Male	0,0041	0,0156	0,794	4
5	Famay Ahmad Al-Silmi	Male	0,0047	0,0148	0,7586	5
6	Mohammad Waliyullah	Male	0,0048	0,0147	0,7523	6
7	Ahmat Rendi Saputra	Male	0,0055	0,0139	0,7157	7
8	Mohammad Zidan Rizki	Male	0,0068	0,0129	0,655	8
9	Khulafaur Rosidin	Male	0,0074	0,0122	0,6209	9
10	Muktar Efendi	Male	0,008	0,0116	0,5933	10

The results presented in Table 4 indicate that Sastro Hadi Saputro achieved the highest preference value ($C_i = 0.8954$), indicating that this athlete possessed the closest proximity to the positive ideal solution and the greatest overall performance balance among all evaluated candidates. Similar performance patterns were also observed for Hegar Marga Birawa and Bagus Prayetno, who consistently demonstrated strong performance across dominant criteria such as speed, agility, explosive power, and endurance.

An important finding emerging from the ranking analysis is that the proposed AHP-TOPSIS framework generated noticeable ranking shifts compared with the conventional selection mechanism previously implemented by KONI Lamongan. Several athletes with strong physical performance but relatively limited achievement records experienced ranking improvements within the proposed framework. Conversely, athletes relying predominantly on historical achievements but demonstrating lower physical readiness tended to experience ranking declines.

This finding indicates that the reconstruction of criterion weighting significantly influences fairness within athlete selection systems. From a practical perspective, the proposed framework provides greater opportunities for newcomer athletes to compete objectively based on current performance quality rather than historical administrative advantages alone. Such findings are particularly relevant for regional sports development systems where athlete regeneration frequently becomes constrained by the dominance of senior athletes possessing accumulated competition certificates.

The validity of the proposed system was further examined by comparing computational ranking outcomes with manual evaluations performed by senior coaches. The comparison produced a Spearman correlation coefficient

of 0.9945 and a Pearson correlation coefficient of 0.9863, indicating an extremely high level of agreement between the system-generated rankings and expert evaluation results. In addition, the ranking similarity accuracy reached 84.62%, demonstrating that the proposed system was capable of reproducing professional decision-making patterns while simultaneously minimizing subjective bias. Compared with previous studies, the present research contributes additional insight regarding the strategic importance of weighting reconstruction within MCDM-based athlete selection systems. Earlier studies involving AHP and TOPSIS integration generally focused on technical implementation and ranking optimization without specifically examining the imbalance between physical readiness and historical achievement indicators. In contrast, this study demonstrates that weighting structure itself represents a critical variable capable of influencing athlete opportunity distribution and organizational fairness.

From a managerial perspective, the implementation of the proposed Decision Support System contributes to more transparent and accountable athlete management practices. Coaches and management personnel are able to evaluate athlete performance using measurable indicators supported by mathematical computation rather than relying predominantly on intuition or personal preference. Furthermore, the flexibility of the AHP framework enables organizational priorities to be reconstructed dynamically according to evolving competition requirements and athlete conditions.

Despite these promising findings, several limitations remain within the present study. The evaluation framework primarily focused on physical performance variables and did not incorporate psychological readiness, tactical intelligence, injury history, or behavioral discipline factors, all of which may also influence athlete competitiveness. Additionally, the study was conducted within a single regional sports organization involving a relatively limited number of athlete candidates. Future studies may therefore expand the evaluation framework by integrating additional performance dimensions and comparing alternative Multi-Criteria Decision Making methods to further strengthen the robustness and generalizability of athlete selection systems in competitive sports environments.

IV. CONCLUSION

This study successfully implemented an integrated Analytic Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) framework to support the athlete selection process for PORPROV Jawa Timur IX 2025 within Komite Olahraga Nasional Indonesia (KONI) Kabupaten Lamongan. The proposed Decision Support System was capable of transforming subjective expert judgments into measurable computational recommendations, thereby improving transparency, consistency, and objectivity in athlete evaluation.

The AHP analysis demonstrated that speed, agility, balance, explosive power, and aerobic endurance constituted the most influential criteria in determining athlete competitiveness. The resulting weighting structure



reflected the practical characteristics of pencak silat competition, which prioritizes dynamic physical readiness over isolated muscular endurance indicators. Furthermore, the consistency test produced a Consistency Ratio (CR) value of 0.0018, indicating that the expert judgments used in the weighting process were logically reliable and mathematically consistent.

The implementation of the TOPSIS method successfully generated athlete rankings based on proximity to ideal performance conditions. The ranking results revealed that athletes with balanced physical performance across multiple criteria obtained higher preference values compared with athletes relying predominantly on historical achievement records alone. In addition, the comparison between system-generated rankings and manual coach evaluations demonstrated a very high level of agreement, as indicated by Spearman and Pearson correlation values of 0.9945 and 0.9863, respectively. These findings confirm that the proposed framework is capable of supporting professional decision-making while simultaneously reducing subjective bias within the athlete selection process.

Overall, the integration of AHP and TOPSIS provides a more adaptive and scientifically grounded athlete selection mechanism for regional sports organizations. The proposed framework not only improves computational accuracy but also contributes to fairer athlete opportunity distribution by emphasizing current performance quality rather than static administrative achievements. Future studies are recommended to incorporate additional evaluation dimensions such as psychological readiness, tactical intelligence, and injury history to further strengthen the comprehensiveness of athlete selection systems in competitive sports environments.

References

- [1] Z. Zhai, M. Du, and H. Meng, "The sports industry and its effect on the quality of life and regional development," *Soc. Sci. Q.*, vol. 105, no. 3, pp. 744–759, May 2024, doi: 10.1111/ssqu.13357.
- [2] T. V. Kopylova, "ECONOMIC IMPACT OF MAJOR SPORTING EVENTS ON REGIONAL DEVELOPMENT," *EKONOMIKA I UPRAVLINIE: PROBLEMY, RESHENIYA*, vol. 10/9, no. 151, pp. 96–102, 2024, doi: 10.36871/ek.up.p.r.2024.10.09.011.
- [3] P. Keliat and A. Al Munawar, "Analysis of the Impact of the Pencak Silat Coach Training Program on Athlete Development in North Sumatra in 2024," *International Journal Of Education, Social Studies, And Management (IJESSM)*, vol. 4, no. 3, pp. 1351–1357, Jan. 2025, doi: 10.52121/ijessm.v4i3.579.
- [4] M. Rizaldy, A. Muchayan, and M. Kamisutara, "The Decision Support System For The Acceptance of Pencak Silat Athletes in Pra-PON and PUSLATDA Team Selection Using Technique For Other Preference By Similarity To Ideal Solution Method," *IJEEIT International Journal of Electrical Engineering and Information Technology*, vol. 3, no. 2, pp. 8–19, Jan. 2021, doi: 10.29138/ijeeit.v3i2.1297.
- [5] K. Ramadhan, S. Hendarto, and B. Satyawan, "Management of pencak silat achievement in IPSI Klaten Regency in 2015-2019," *PHEDHERAL*, vol. 17, no. 2, p. 1, May 2021, doi: 10.20961/phduns.v17i2.51444.
- [6] F. Thompson, F. Rongen, I. Cowburn, and K. Till, "The Impacts of Sports Schools on Holistic Athlete Development: A Mixed Methods Systematic Review," *Sports Medicine*, vol. 52, no. 8, pp. 1879–1917, Aug. 2022, doi: 10.1007/s40279-022-01664-5.
- [7] P. Song, "Development of a fuzzy comprehensive evaluation system for the effect evaluation of sprint training - based on AHP and entropy weight method," *Sci. Rep.*, vol. 16, no. 1, p. 3301, Dec. 2025, doi: 10.1038/s41598-025-33088-2.
- [8] A. Weise *et al.*, "Pre-Participation Evaluation of Recreational and Competitive Athletes – A Systematic Review of Guidelines and Consensus Statements," *Sports Med. Open*, vol. 11, no. 1, p. 33, Apr. 2025, doi: 10.1186/s40798-025-00837-6.
- [9] R. Vavrek, "An Analysis of Usage of a Multi-Criteria Approach in an Athlete Evaluation: An Evidence of NHL Attackers," *Mathematics*, vol. 9, no. 12, p. 1399, Jun. 2021, doi: 10.3390/math9121399.
- [10] S. ESEN, "Talent Selection in Sport With Multi-Criteria Decision Making Methods: TOPSIS and VIKOR Comparison," *The Online Journal of Recreation and Sports*, vol. 12, no. 4, pp. 818–823, Oct. 2023, doi: 10.22282/tojras.1351345.
- [11] J. Baker, J. Schorer, and N. Wattie, "Compromising Talent: Issues in Identifying and Selecting Talent in Sport," *Quest*, vol. 70, no. 1, pp. 48–63, Jan. 2018, doi: 10.1080/00336297.2017.1333438.
- [12] K. Johnston, L. Farah, and J. Baker, "Storm Clouds on the Horizon: On the Emerging Need to Tighten Selection Policies," *Front. Sports Act. Living*, vol. 3, Nov. 2021, doi: 10.3389/fspor.2021.772181.
- [13] Eko Setyawan, Muhammad Fairuzabadi, and Puji Handayani Putri, "DECISION SUPPORT SYSTEM FOR SELECTING ATHLETE CANDIDATES IN PSTI BANTUL USING PROFILE MATCHING METHOD BASED ON WEB," *JTH: Journal of Technology and Health*, vol. 1, no. 2, pp. 119–130, Dec. 2023, doi: 10.61677/jth.v1i2.104.
- [14] X. Schelling, E. Alonso-Perez-Chao, and S. Robertson, "Implementation of a Decision Support System to Enhance Movement Proficiency Assessment in Sport," *J. Funct. Morphol. Kinesiol.*, vol. 10, no. 1, p. 86, Mar. 2025, doi: 10.3390/jfmk10010086.
- [15] X. Schelling and S. Robertson, "A development framework for decision support systems in high-performance sport," *Int. J. Comput. Sci. Sport*, vol. 19, no. 1, pp. 1–23, Jul. 2020, doi: 10.2478/ijcss-2020-0001.
- [16] T. Hovorushchenko, Y. Hnatchuk, A. Hnatchuk, and O. Fehyr, "Decision Support System For The Coaching Staff in the Process of Preliminary



- Selection of Players,” in *2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT)*, IEEE, Oct. 2023, pp. 1–4. doi: 10.1109/CSIT61576.2023.10324119.
- [17] T. Parmonangan and M. A. Hasan, “DECISION SUPPORT SYSTEM FOR THE SELECTION OF PENCAK SILAT ATHLETES USING AHP - TOPSIS,” *ZONAsi: Jurnal Sistem Informasi*, vol. 4, no. 1, pp. 117–127, Jun. 2022, doi: 10.31849/zn.v4i1.9320.
- [18] S. S. Putro, W. R. Santoso, Husni, and A. Rachmad, “Candidate Selection of Athletes Sparring for Boy Category Pencak Silat Using TOPSIS: Case Study in PSHT Pencak Silat,” *J. Phys. Conf. Ser.*, vol. 1477, no. 2, p. 022019, Mar. 2020, doi: 10.1088/1742-6596/1477/2/022019.
- [19] A. Akmaludin, S. Sidik, N. Iriadi, A. Arfian, and A. D. Surianto, “Selection of the Best Swimming Athletes using MCDM-AHP and VIKOR Methods,” *Sinkron*, vol. 6, no. 1, pp. 44–52, Oct. 2021, doi: 10.33395/sinkron.v6i1.10998.
- [20] I. Canco, D. Kruja, and T. Iancu, “AHP, a Reliable Method for Quality Decision Making: A Case Study in Business,” *Sustainability*, vol. 13, no. 24, p. 13932, Dec. 2021, doi: 10.3390/su132413932.
- [21] R. Gharizadeh Beiragh *et al.*, “An integrated Multi-Criteria Decision Making Model for Sustainability Performance Assessment for Insurance Companies,” *Sustainability*, vol. 12, no. 3, p. 789, Jan. 2020, doi: 10.3390/su12030789.
- [22] R. Aulia, “Enhancing Decision Making through Professional Judgment and Comprehensive Information Analyses,” *Advances in Managerial Auditing Research*, vol. 1, no. 3, Sep. 2023, doi: 10.60079/amar.v1i3.230.
- [23] O. A. Adeoye-Olatunde and N. L. Olenik, “Research and scholarly methods: Semi-structured interviews,” *JACCP: JOURNAL OF THE AMERICAN COLLEGE OF CLINICAL PHARMACY*, vol. 4, no. 10, pp. 1358–1367, Oct. 2021, doi: 10.1002/jac5.1441.
- [24] H. Kallio, A. Pietilä, M. Johnson, and M. Kangasniemi, “Systematic methodological review: developing a framework for a qualitative semi-structured interview guide,” *J. Adv. Nurs.*, vol. 72, no. 12, pp. 2954–2965, Dec. 2016, doi: 10.1111/jan.13031.
- [25] F. M. Theunissen, C. N. Bezuidenhout, and S. Alam, “Exploring the shortcomings in formal criteria selection for multicriteria decision making based inventory classification models: a systematic review and future directions,” *Int. J. Prod. Res.*, vol. 62, no. 19, pp. 7279–7299, Oct. 2024, doi: 10.1080/00207543.2024.2320680.
- [26] A. Rasmussen, H. Sabic, S. Saha, and I. E. Nielsen, “Supplier selection for aerospace & defense industry through MCDM methods,” *Clean. Eng. Technol.*, vol. 12, p. 100590, Feb. 2023, doi: 10.1016/j.clet.2022.100590.
- [27] W.-C. Yang, J.-B. Ri, J.-Y. Yang, and J.-S. Kim, “Materials selection criteria weighting method using analytic hierarchy process (AHP) with simplest questionnaire and modifying method of inconsistent pairwise comparison matrix,” *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, vol. 236, no. 1, pp. 69–85, Jan. 2022, doi: 10.1177/14644207211039912.
- [28] S. Muniroglu, “An Overview of the Important Points of Talent Selection in Sports,” *Research & Investigations in Sports Medicine*, vol. 3, no. 3, Jun. 2018, doi: 10.31031/RISM.2018.03.000561.
- [29] A. R. Colson and R. M. Cooke, “Cross validation for the classical model of structured expert judgment,” *Reliab. Eng. Syst. Saf.*, vol. 163, pp. 109–120, Jul. 2017, doi: 10.1016/j.res.2017.02.003.
- [30] E. Triantaphyllou and J. Yanase, “The use of pairwise comparisons for decision making may lead to grossly inaccurate results,” *Comput. Ind. Eng.*, vol. 198, p. 110653, Dec. 2024, doi: 10.1016/j.cie.2024.110653.
- [31] J. Krejčí and J. Stoklasa, “Aggregation in the analytic hierarchy process: Why weighted geometric mean should be used instead of weighted arithmetic mean,” *Expert Syst. Appl.*, vol. 114, pp. 97–106, Dec. 2018, doi: 10.1016/j.eswa.2018.06.060.
- [32] V. Lukinskiy, V. Lukinskiy, B. Sokolov, and D. Bazhina, “An Empirical Examination of the Consistency Ratio in the Analytic Hierarchy Process (AHP),” 2021, pp. 477–485. doi: 10.1007/978-3-030-85914-5_51.
- [33] Y. Liu, C. M. Eckert, and C. Earl, “A review of fuzzy AHP methods for decision-making with subjective judgements,” *Expert Syst. Appl.*, vol. 161, p. 113738, Dec. 2020, doi: 10.1016/j.eswa.2020.113738.
- [34] L. V. Cremades and A. Ponsich, “Simple and objective determination of criteria weights for evaluating alternatives when using the Analytic Hierarchy Process,” *International Journal of the Analytic Hierarchy Process*, vol. 16, no. 3, Jan. 2025, doi: 10.13033/ijahp.v16i3.1177.
- [35] M. Jagtap and P. Karande, “The m-polar fuzzy set ELECTRE-I with revised Simos’ and AHP weight calculation methods for selection of non-traditional machining processes,” *Decision Making: Applications in Management and Engineering*, vol. 6, no. 1, pp. 240–281, Apr. 2023, doi: 10.31181/dmame060129022023j.
- [36] T. H. Dinh and T.-T.-D. Nguyen, “An integrated multi-criteria decision-making approach for life cycle approach in road construction projects,” *Journal of Infrastructure Policy and Development*, vol. 8, no. 9, p. 6926, Sep. 2024, doi: 10.24294/jipd.v8i9.6926.
- [37] R. Ishak, R. Suryaditya, S. Nur Khasanah, and P. Handayani, “Penentuan Mahasiswa Terbaik Dengan Menggunakan Metode Topsis (Teknik Urutan Preferensi Berdasarkan Kesamaan dengan Solusi Ideal),” *Infotek: Jurnal Informatika dan*



- Teknologi*, vol. 8, no. 2, pp. 554–562, Jul. 2025, doi: 10.29408/jit.v8i2.30601.
- [38] R. Syahputra, O. Oktarifaldi, L. P. Putri, and S. Bakhtiar, “Physical Performance Characteristic of West Sumatera Athletes in Combat and Game Sports,” *International Journal of Human Movement and Sports Sciences*, vol. 10, no. 4, pp. 732–747, Aug. 2022, doi: 10.13189/saj.2022.100414.
- [39] R. M. Cooke, D. Marti, and T. Mazzuchi, “Expert forecasting with and without uncertainty quantification and weighting: What do the data say?,” *Int. J. Forecast.*, vol. 37, no. 1, pp. 378–387, Jan. 2021, doi: 10.1016/j.ijforecast.2020.06.007.
- [40] Z. Liang, J. Nie, Q. Ou, and J. Wu, “A Dual - Weighting TOPSIS Approach to Evaluating Olympic Sports,” *Advances in Engineering Technology Research*, vol. 13, no. 1, p. 1359, Mar. 2025, doi: 10.56028/aetr.13.1.1359.2025.



Comparison of LDA and BERTopic in Identifying Public Issues in the MBG Program

Nur Hayati¹, Saikin^{2*}, Hairul Fahmi³

^{1,2}Information Systems Study Program, STMIK Lombok, Praya, Indonesia

³Informatics Engineering Study Program, STMIK Lombok, Praya, Indonesia

Email: ¹nurhayaati83@gmail.com, ²eken.apache@gmail.com, ³iroel.ami@gmail.com

Abstract –The Free Nutritious Food Program (MBG) is a government policy that has generated various public responses and opinions on social media. The large amount of unstructured text data. This study aims to compare the performance of the Latent Dirichlet Allocation (LDA) and BERTopic methods in identifying public issues related to the MBG program on TikTok data. The dataset used amounted to 13,538 data obtained through a scraping process based on keywords related to MBG. The research stages include text preprocessing, bigram and trigram formation, text representation using TF-IDF and embedding, topic modeling, and evaluation using coherence score and topic diversity. The results showed that the LDA method produced better evaluation performance with a coherence score of 0.5098 and a topic diversity of 0.9000. Meanwhile, BERTopic produced a coherence score of 0.4133 and a topic diversity of 0.7667, but was able to produce topics that were more contextual and semantically representative. Based on these results, LDA is superior in terms of the stability and quality of word associations between topics, while BERTopic is more effective in understanding the context of issues in short and unstructured social media data.

Keywords – *Topic Modeling, LDA, BERTopic, Social Media, Free Nutritious Food.*

I. INTRODUCTION

The Free Nutritious Food Program (MBG) is a government initiative aimed at improving the nutritional quality of the community and supporting the development of healthier and more productive human resources. As a public policy that directly impacts the community, the implementation of this program not only impacts health aspects but also triggers various responses, perceptions, and discussions among the public. Differences in social and economic backgrounds and levels of public understanding of the program have led to the emergence of diverse views, both supportive and critical. In this context, understanding how public opinion is formed and developed is crucial for policy evaluation. Along with the increasing use of digital platforms, various public responses to the MBG program are now widely recorded and disseminated through social media, thus opening up opportunities for further analysis as a rich source of information on emerging issues in society [1][2].

The large amount of data generated from social media related to the MBG program presents unique challenges in the analysis process, especially because the data is unstructured, dynamic, and contains various language variations. To overcome this, approaches based on text mining and natural language processing (NLP) are effective solutions in extracting important information from large collections of text. One technique widely used in this context is topic modeling, which aims to identify patterns of topics or key issues hidden within text data without requiring prior labels. Methods such as Latent Dirichlet Allocation (LDA) have been widely used due to their ability to cluster documents based on word distribution, although they still have limitations in capturing deeper semantic context, especially in short texts such as social media. As technology advances, more modern approaches such as BERTopic are being developed that utilize embedding-based representations, thus being

able to produce more contextual and coherent topics. Therefore, selecting the right method is a crucial factor in producing accurate and relevant public issue analysis [1][3][4][5].

Although various topic modeling methods have been widely used to analyze text data, selecting the right model remains a significant challenge, especially in the context of short and unstructured social media data. Classical methods such as Latent Dirichlet Allocation (LDA) have proven effective in identifying latent topics in various types of documents, but tend to be less than optimal in capturing complex semantic relationships in short texts. On the other hand, newer approaches such as BERTopic offer advantages through the use of transformer-based embeddings that are able to produce more contextual and coherent topics. Several recent studies have shown differences in performance between the two methods, where BERTopic tends to excel in understanding context, while LDA still has advantages in interpretability and computational efficiency. However, the results of this comparison still depend on the characteristics of the data used, so further studies are needed to evaluate the performance of both methods in specific contexts, such as public issue analysis on social media data related to the MBG program [6][7][8][9][10].

Several previous studies have examined the application of topic modeling in text data analysis, particularly in the context of social media and digital documents. A study by Grootendorst (2022) introduced BERTopic as an embedding-based approach capable of generating topics with a higher degree of coherence than traditional methods [1]. Furthermore, Egger and Yu (2022) compared LDA and BERTopic on Twitter data, showing that BERTopic was superior in capturing semantic context, while LDA remained relevant in terms of model interpretability [6]. Another study by Bianchi et al. (2021) also emphasized that the use of contextualized embeddings can improve the quality of the generated topics [3]. On the other hand, a



study by Pratiwi and Tania (2025) showed that a combination of several topic modeling methods such as LDA, BERTopic, and NMF can provide more comprehensive insights in the information extraction process [7]. In addition, Liu (2024) revealed that the performance of each method is highly dependent on the characteristics of the data used, so further evaluation in different data contexts is needed [4]. However, most of these studies have not specifically examined the comparison of the two methods in the context of public issue analysis of government programs such as MBG, thus opening up opportunities for further research.

Based on the background and review of previous research, this study aims to analyze and compare the performance of the Latent Dirichlet Allocation (LDA) and BERTopic methods in identifying public issues related to the Free Nutritious Food (MBG) program in social media data. This study proposes a structured analysis approach through the stages of data collection, preprocessing, text representation, topic modeling, to coherence score-based evaluation and interpretation of results. The main contribution of this study lies in the comparative application of two different topic modeling methods, namely the probabilistic method and embedding-based methods, in the context of dynamic and unstructured social media data. In addition, this study also provides added value through analysis that not only focuses on the quality of the resulting topics, but also on the ability of each method to represent public issues in a more contextual and relevant manner. Thus, the results of this study are expected to provide methodological recommendations in selecting the appropriate topic modeling model and serve as a reference for further research and policy makers in understanding the dynamics of public opinion towards the MBG program.

Research on topic modeling has grown rapidly in text data analysis, particularly in social media. Embedding-based approaches such as BERTopic have been introduced as methods capable of generating topics with a better level of coherence than traditional methods [1][11]. Several studies have shown that BERTopic has advantages in understanding the semantic context of text data, especially on social media platforms such as Twitter [8]. In addition, the use of contextualized embeddings has also been shown to improve the quality of the generated topics [3].

On the other hand, classical methods such as Latent Dirichlet Allocation (LDA) remain widely used due to their ability to effectively identify latent topic structures [12]. In the evaluation process, the coherence score is used as one of the main metrics to measure topic quality [13]. The development of embedding-based methods is also continuously carried out to improve topic modeling performance [14]. However, the performance of each method is highly dependent on the characteristics of the data used [12][15]. Therefore, a comparative study between LDA and BERTopic is needed in specific contexts, such as public issue analysis in the MBG program.

II. RESEARCH METHODOLOGY

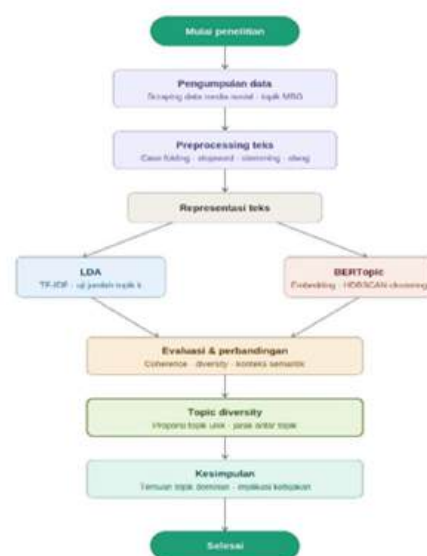


Figure 1. Research Flow

The research began with identifying problems related to analyzing public issues related to the Free Nutritious Food (MBG) program on social media. At this stage, the research objective was determined, namely to compare the performance of the LDA and BERTopic methods in identifying topics.

2.1 Data Collection

In the data collection stage, this study utilized the TikTok platform as the primary source to obtain data related to the Free Nutritious Food (MBG) program issue. Data were collected using scraping techniques based on relevant keywords, resulting in 13,538 data points. The resulting dataset consists of several important attributes, including text (caption or comment content), diggCount (number of likes), replyCommentTotal (number of comment replies), createTimeISO (upload time), uniqueId (user identity), and videoWebUrl and source_file as data sources. In addition, a preprocessing process was carried out that produced derived features such as text_clean, tokens, token_count, tokens_nostop, and text_clean_advanced. The dataset is also equipped with additional information such as text_length, word_count, and temporal attributes (date, month, hour) to support further analysis.

2.2 Text Preprocessing

Text preprocessing is the initial stage in text analysis, which aims to clean and normalize data so that it is ready for further processing. This stage includes case folding, tokenizing, stopword removal, and stemming. In social media data such as TikTok, this process also includes slang normalization to address unstructured language variations. Good preprocessing will improve the quality of text representation and the accuracy of the model used [16]. The collected data is then processed through preprocessing stages to improve data quality. These stages include:

- a. Folding case

The case folding stage is the process of converting all text to lowercase. This process is carried out as an initial part of text preprocessing to standardize the word formatting in the dataset. Technically, the system will read each character in the document and then convert all capital letters to lowercase using a string transformation function. At this stage, alphabetic characters such as "A" to "Z" are converted to "a" to "z," while non-letter characters can be retained or processed further in the next preprocessing stage. The main purpose of case folding is to avoid differences in word representation due to variations in the use of uppercase and lowercase letters. In text analysis, words such as "Data," "DATA," and "data" actually have the same meaning, but the computer will treat them as different tokens if not normalized. This condition can unnecessarily increase the vocabulary size and cause redundancy in the text feature representation.

b. Tokenizing

The tokenizing stage is the process of breaking down text into smaller units of words or tokens so that the data can be analyzed in a structured manner. At this stage, the system reads each sentence and then separates words based on certain characters such as spaces, punctuation, symbols, or other delimiters. The result of the tokenizing process is a list of tokens that represent each word in the document. For example, the sentence "The MBG program is very helpful for the community" will be converted into tokens such as ["program", "mbg", "very", "helping", "community"]. Technically, tokenizing is a crucial stage in preprocessing because most text mining and natural language processing algorithms cannot directly process raw text in the form of complete sentences. By breaking text into tokens, the system can perform further processes such as filtering, stopword removal, stemming, word weighting using TF-IDF, and even feature formation in topic modeling. In addition, tokenizing also helps the system calculate the frequency of word occurrences and recognize word distribution patterns in documents. Therefore, the quality of the process

c. Stopword removal

```
custom_stopwords = {
    'ya', 'yaa', 'yaaa', 'yg', 'yang', 'di', 'dari', 'dan', 'untuk', 'ini', 'it',
    'dengan', 'pada', 'ke', 'adalah', 'juga', 'akan', 'atau', 'tersebut', 'bisa',
    'tidak', 'ada', 'sudah', 'saya', 'kita', 'mereka', 'dia', 'kami', 'kalo',
    'kalau', 'nih', 'nah', 'sih', 'dong', 'deh', 'lho', 'loh', 'tuh',
    'pak', 'bu', 'bapak', 'ibu', 'mbg', 'bgn', 'ag', 'aja', 'saja', 'banget',
    'banget', 'sangat', 'sudah', 'masih', 'udah', 'akan', 'lagi', 'ttp', 'tetap',
    'jd', 'jadi', 'utk', 'dgn', 'krn', 'karena', 'jd', 'jdi', 'sdh',
    'bs', 'bisa', 'ga', 'gak', 'nggak', 'nggak', 'gimana', 'gimana', 'Gimana',
    'gimana', 'dong', 'dongs', 'kak', 'kakak', 'bang', 'teh', 'tuh', 'ni',
    'emang', 'emangnya', 'bener', 'betul', 'sy', 'aku', 'ku', 'mu', 'kau',
    'yaudah', 'ya', 'yaa', 'yok', 'ayo', 'yuk', 'gitu', 'begitu', 'gtu',
    'dm', 'dm', 'gpp', 'gakpapa', 'ok', 'oke', 'sip', 'sipp', 'siapp',
    'please', 'pls', 'tolong', 'tig', 'thx', 'thanks', 'thank', 'you',
    'wkwk', 'wkwkwk', 'haha', 'hahaha', 'lol', 'lmao', 'btw', 'rt',
    'via', 'amp', 'the', 'and', 'of', 'to', 'is', 'in', 'for', 'on',
    'jd', 'jika', 'bila', 'apabila', 'bahwa', 'sejak', 'sejak', 'antara'
}
```

Figure 2. Stopword Removal

At this stage, a stopword removal process is carried out using a list of custom stopwords compiled according to the characteristics of the social media data. The stopword list includes common Indonesian words such as "dan," "yang," "di," "dari," and "untuk," as well as non-standard and slang words that frequently appear in social media conversations such as "gak," "gag," "aja," "banget," "wkwk," "haha," and

"lol." Furthermore, several common English words such as "the," "and," "of," and "is" are also removed because they are deemed not to provide important information to the topic analysis process. Based on the list in the figure, the custom stopwords also include greetings, abbreviations, emotional expressions, and repeated words that frequently appear in social media comments. Technically, the stopword removal process is carried out by comparing each tokenized token with a predetermined list of custom stopwords. If a token is found in the stopword list, it will be removed from the document. Conversely, tokens not included in the list will be retained as important features. The use of custom stopwords aims to make the text cleaning process more relevant to the context of TikTok data, which tends to be informal and unstructured. By eliminating words that do not have important meaning, the text representation becomes more focused on words that represent the main information so that the quality of topic modeling analysis can be improved.

d. Stemming

The stemming stage is carried out to convert each token resulting from the stopword removal process into its basic word form using a stemmer. This process is carried out by iterating on each token, then applying the stemming function to obtain the basic word form so that variations of words with the same meaning can be standardized. If an error occurs in a token during the stemming process, the token is kept unchanged to avoid losing information in the data. The stemming results are then stored as a token list and recombined into complete text for use in the next analysis stage. In addition, a progress indicator is added to monitor the amount of data that has been successfully processed during the process. The stemming stage aims to simplify word variations, reduce redundancy, and increase the consistency of text representation so that the analysis and topic modeling processes can be carried out more effectively.

e. Normalization of slang

In the slang normalization stage, non-standard words, abbreviations, and informal language are replaced with standard word forms using a pre-compiled slang dictionary. The dictionary consists of 423 word entries covering various variations of everyday language commonly used on social media, such as "gk," "gak," and "gag" which are changed to "tidak," and "yg" to "yang". This stage aims to standardize word forms and reduce spelling variations in text data. Technically, the normalization process is carried out by matching each preprocessed token to the slang dictionary. The system will iterate over all tokens, then check whether the token is included in the slang list. If found, the token will be replaced with the standard word form, while tokens not in the dictionary will be retained in their original form. This process is carried out automatically on all documents. Slang normalization is important because social media data generally contains a lot of informal language and abbreviations that can increase noise in the data. By standardizing words that have the same meaning, text representation becomes more consistent, thus improving the quality of analysis and model performance in the next stage[21][22]

f. Bigram/trigram extraction

N-grams are a text processing technique used to capture

word sequences within a document. Bigrams and trigrams are able to maintain local context that unigrams cannot. For example, the phrase “nutritious food” will be more meaningful if treated as a single unit. Therefore, the use of n-grams can improve the quality of topic modeling results, especially in short texts such as social media [17]. Bigram and trigram extraction uses the Phrases model to combine words that frequently appear in sequence. The data used comes from normalized tokens, then processed to form two-word (bigram) and three-word (trigram) combinations with a certain minimum occurrence limit. The results show that from 13,574 data, 203 unique bigrams and 75 unique trigrams were formed. Some of the dominant bigrams include “bapak_prabowo”, “sehat_selalu”, and “terima_kasih”, while the trigrams that frequently appear are “di_bawa_pulang” and “terima_kasih_bapak_prabowo”. In addition, a visualization is performed to display the top 20 word combinations based on their frequency of occurrence. This process aims to capture a more specific word context so that the analysis becomes more representative.

g. Text Representation

Text representation is the process of converting text data into numerical form so that it can be processed by an algorithm. The LDA method uses the TF-IDF (Term Frequency–Inverse Document Frequency) approach to measure the importance of words in a document. Meanwhile, an embedding-based approach such as BERT is used in BERTopic to capture semantic context more deeply. This model is able to understand the relationship between words in a sentence contextually [15].

1. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is a probabilistic-based topic modeling method used to discover latent topics in a collection of documents. LDA assumes that each document consists of a distribution of several topics, while each topic consists of a distribution of words [18]. This method is widely used because it has high interpretability and good computational efficiency, although it still has limitations in capturing semantic context in short texts [19][20]. At this stage, the text data is first converted into a numerical representation using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The TF-IDF method is used to measure the level of importance of a word in a document to the entire collection of documents, where TF indicates the frequency of the word's appearance in the document and IDF is used to reduce the weight of words that appear too often in many documents. The TF-IDF representation produces a numerical matrix that is used as input in the topic modeling process using the LDA method.

$$TF - IDF(t, d) = TF(t, d) \times \log\left(\frac{N}{df(t)}\right) \quad (1)$$

Information:

- $TF(t, d)$ = word frequency in document
- N = number of documents
- $df(t)$ = number of documents containing the word

2. BERTopic (Embedding)

BERTopic is a modern topic modeling method that combines transformer embeddings, clustering, and c-TF-IDF. This approach enables the formation of more coherent

and contextual topics than traditional methods [9][18]. BERTopic is very effective for use on social media data such as TikTok, which is short, dynamic, and unstructured [1], [10][19]. In this approach, each document is first converted into a numeric vector using sentence embedding, then automatically grouped into topics using a density-based clustering algorithm such as HDBSCAN, without the need to specify the number of topics in advance. Mathematically, the embedding representation process can be expressed as:

$$V_d = f(d) \quad (2)$$

Information:

- V_d = document embedding vector d
- $f(d)$ = embedding function (e.g. transformer-based model like BERT)

Next, the clustering process is carried out by grouping documents based on the proximity of the distance in the vector space, which is generally calculated using cosine similarity:

$$\text{sim}(x, y) = \frac{x \cdot y}{||x|| ||y||} \quad (3)$$

Once the clusters are formed, each topic is represented using class-based TF-IDF (c-TF-IDF) to extract the most representative words in each topic:

$$c_TFIDF_{t,c} = \frac{f_{t,c}}{\sum f_{t,c}} x \log\left(\frac{N}{df_t}\right) \quad (4)$$

Information:

- $f_{t,c}$ = frequency of word t in cluster c
- N = total number of documents
- df_t = number of clusters containing word t

h. Evaluation and Comparison

Evaluation is done to measure the quality of the topic using several metrics:

1. Coherence Score

$$C = \frac{1}{M} \sum \log\left(\frac{D(w_i, w_j) + 1}{D(w_j)}\right) \quad (4)$$

2. Topic Diversity

The topics generated by the topic modeling model. This metric aims to ensure that each topic has a unique word representation and does not experience excessive word overlap. The higher the topic diversity value, the more diverse the words used in each topic, thus improving the quality of topic separation. Conversely, a low topic diversity value indicates that many words appear repeatedly across several topics, so the resulting topics tend to be less specific and similar to each other. Mathematically, topic diversity is calculated based on the proportion of unique words to the total words from all topics generated by the model. In the formula, represents the set of words in the i -th topic, k indicates the number of topics, and n is the number of top words retrieved from each topic. The union operator is used to calculate the total unique words from all topics. This value is then divided by the product of the number of topics and the number of top words in each topic ($k \times n$). Thus, topic diversity can be



used to evaluate the extent to which the model is able to generate diverse, representative topics with minimal overlap between topics. In general, topic diversity is calculated based on the proportion of unique words to the total words appearing in a number of top topics, which can be formulated as: $W_i \cup_1^k = 1W_i$

$$TD = \frac{|U_i^k = W_i|}{k \times n} \quad (5)$$

Information:

- TD = topic diversity
- W_i = set of words on topic i
- k = number of topics
- n = number of top words on each topic

A high topic diversity value indicates that the generated topics have more diverse words, thus indicating better topic modeling quality.

III. RESULTS AND DISCUSSION

3.1 LDA(TF-IDF)

In the topic modeling stage using the Latent Dirichlet Allocation method, the text data is first converted into a Document-Term Matrix using CountVectorizer with parameters `max_df=0.95`, `min_df=2`, and `max_features=10000`. The transformation results in a matrix of size (6787, 4599), which shows that there are 6,787 documents and 4,599 unique word features used in the topic modeling process.

```
Training LDA dengan k=5...
Perplexity: 764.99
Coherence (C_v): 0.5098
Log-Likelihood: -606265.69
Topic Diversity: 0.9000
Top 5 kata topik 0: gizi, pangan, pelayanan, mokol, tahu

Training LDA dengan k=8...
Perplexity: 799.22
Coherence (C_v): 0.5098
Log-Likelihood: -610262.79
Topic Diversity: 0.9000
Top 5 kata topik 0: tahu, min, bpk_prabowo, omon, masya_allah

Training LDA dengan k=10...
Perplexity: 828.40
Coherence (C_v): 0.4782
Log-Likelihood: -613536.58
Topic Diversity: 0.9100
Top 5 kata topik 0: tahu, pemimpin, wo, kapan, jangan
```

Figure 3. Coherence Score Matrix

Next, LDA model training was conducted with varying topic counts of 5, 8, 10, 12, 15, and 20 topics to find the optimal model. Evaluation was performed using perplexity, coherence score (C_v), log-likelihood, and topic diversity metrics. The test results showed that the model with 8 topics produced the highest coherence score of 0.5098 with a topic diversity of 0.9000, so it was selected as the best model because it was able to produce more coherent and representative topics compared to other topic counts.

	k	perplexity	coherence	log_likelihood	topic_diversity
0	5	764.987803	0.506777	-606265.686932	0.900000
1	8	799.220142	0.509753	-610262.791483	0.900000
2	10	828.395877	0.478227	-613536.578687	0.910000
3	12	851.365677	0.465593	-616033.878700	0.933333
4	15	876.960178	0.475153	-618738.377923	0.886667
5	20	909.348790	0.430043	-622049.821984	0.955000

Table 1. LDA Model Evaluation Results Based on Variations in the Number of Topics

The evaluation results of the Latent Dirichlet Allocation model were then stored in a dataframe to facilitate analysis and comparison between models. Based on the evaluation results, the model with 8 topics produced the highest coherence score of 0.5098 with a topic diversity of 0.9000. Meanwhile, the model with 5 topics had the lowest perplexity score of 764.99. However, as the number of topics increased to 20, the perplexity value tended to increase and the coherence score decreased. This indicates that too many topics can reduce the quality of semantic relationships between words in the topic. Therefore, the model with 8 topics was chosen as the optimal model because it provides the best balance between topic quality and data representation.

3.2 Topic (Embedding)

In the topic modeling stage using the BERTopic method, this study uses the multilingual SentenceTransformers paraphrase-multilingual-MiniLM-L12-v2 embedding model to represent text into semantic vectors. The modeling process is carried out with the parameters `language='indonesian'`, `nr_topics='auto'`, `top_n_words=15`, and `calculate_probabilities=True`. The stages start from document transformation into embedding, dimensionality reduction using UMAP, density-based clustering, to topic extraction using c-TF-IDF. The modeling results show that BERTopic successfully reduced 101 initial topics to 25 topics, with 24 main topics after removing outliers, and produced 2,345 outlier documents. These results indicate that BERTopic is able to identify public issues related to the MBG program more contextually and representatively in short and unstructured social media data.

```
BERTopic Coherence (C_v): 0.4133
BERTopic Topic Diversity: 0.7667
```

Figure 4. BERTopic Coherence and Topic Diversity Values

Based on the evaluation results, the BERTopic method produced a coherence score (C_v) of 0.4133 and a topic diversity score of 0.7667. The coherence score indicates that the resulting topics have a fairly good semantic relationship, while the topic diversity score indicates a relatively high diversity of words between topics. These results indicate that BERTopic is capable of producing fairly representative and contextual topics in identifying public issues related to the MBG program in social media data.

3.3 Model Evaluation

	Metrik	LDA	BERTopic
0	Coherence Score (C_v)	0.5098	0.4133
1	Perplexity	799.22	N/A
2	Topic Diversity	0.9000	0.7667
3	Silhouette Score	N/A	-0.0510
4	Jumlah Topik	8	24
5	Outlier (%)	0.0%	34.6%
6	Jumlah Dokumen	6787	6787

Table 2. Model Evaluation



Based on the evaluation results, the LDA method produced a coherence score of 0.5098, higher than BERTopic's 0.4133, indicating that the topics in LDA have better word relationships. In the topic diversity metric, LDA also obtained a higher score of 0.9000 compared to BERTopic's 0.7667, thus indicating better word diversity between topics. Meanwhile, BERTopic produced 24 topics with outliers of 34.6%, while LDA produced 8 topics without outliers. In addition, BERTopic obtained a silhouette score of -0.0510, indicating that the quality of cluster separation is still less than optimal. Overall, LDA showed more stable performance based on the evaluation metrics, while BERTopic was superior in generating contextual topics in social media data.

3.4 Topic Diversity

Topic diversity was implemented by calculating the number of unique words across all topics and then dividing them by the total number of words used. Furthermore, a similarity analysis was performed between topics to determine the extent to which the words appearing in each topic overlap. A higher topic diversity value indicates that the resulting topics have more diverse and representative words, thereby reducing overlap between topics. Based on the evaluation results, the LDA method produced a higher topic diversity value than BERTopic. In the LDA model, the formula topic diversity value increased to around 0.92 at $k = 12$ and tended to be stable at subsequent topic counts, while the pairwise topic diversity value approached 1.0, indicating a very good level of word difference between topics. Meanwhile, BERTopic produced a topic diversity value of 0.7667, indicating that the word diversity between topics was still below LDA. These results indicate that LDA is more capable of producing diverse topics with minimal overlap, while BERTopic is superior in producing contextual topics in social media data.

	Model	Jumlah Topik	TD Formula	TD Pairwise	Unique Words
0	LDA (k=5)	5	0.760000	0.895533	38
1	LDA (k=8)	8	0.875000	0.976817	70
2	LDA (k=10)	10	0.890000	0.983863	89
3	LDA (k=12)	12	0.925000	0.990076	111
4	LDA (k=15)	15	0.906667	0.988577	136
5	LDA (k=20)	20	0.890000	0.988681	178
6	BERTopic (80 topics)	80	0.823750	0.995835	659

Table 3. Comparison of Topic Diversity

Based on the results of the topic diversity comparison, the LDA and BERTopic methods show different characteristics in producing topic diversity. In the LDA model, the TD Formula value increases with the increase in the number of topics, starting from 0.7600 at $k = 5$ to reach the highest value of 0.9250 at $k = 12$ with 111 unique words. After that, the diversity value tends to decrease slightly at $k = 15$ and $k = 20$. Meanwhile, the Pairwise TD value in LDA also increases from 0.8955 to close to 1.0, indicating that the topics have a very good level of word differentiation and minimal overlap. On the other hand, BERTopic with 80 topics produces a TD Formula value of 0.8238 and a TD Pairwise value of 0.9958 with a total of 659 unique words. These results indicate that BERTopic is

able to produce a much larger number of unique words and has a very high difference between topics, but the TD Formula value is still below the optimal LDA at $k = 12$. Overall, LDA produces a more stable and balanced topic diversity, while BERTopic produces more and highly semantically diverse topics in social media data.

IV. CONCLUSION

Based on the research results, it can be concluded that the Latent Dirichlet Allocation (LDA) and BERTopic methods have distinct characteristics and advantages in identifying public issues related to the Free Nutritious Food (MBG) program in TikTok social media data. The analysis process was carried out through data collection, text preprocessing, text representation, topic modeling, and evaluation using coherence scores, topic diversity, and manual interpretation.

1. The results showed that the LDA method produced better evaluation performance than BERTopic based on coherence score and topic diversity. The optimal LDA model with 8 topics obtained a coherence score of 0.5098 and a topic diversity of 0.9000, indicating that the resulting topics had good word associations and high topic diversity. Furthermore, LDA was able to produce more stable topics with minimal overlap between topics.
2. On the other hand, BERTopic demonstrated superiority in generating more contextual and semantic topics due to its use of transformer-based embedding. BERTopic successfully identified 24 key topics from short, unstructured social media data. Although the coherence score and topic diversity obtained were lower than LDA, at 0.4133 and 0.7667, respectively, this method was able to capture the context of public issues more deeply and representatively.
3. Overall, this study shows that LDA is superior in terms of model stability, word association quality, and topic diversity, while BERTopic is more effective in understanding semantic context in social media data. Therefore, the choice of topic modeling method needs to be tailored to the analysis objectives and the characteristics of the data used. The results of this study are expected to serve as a reference in the development of social media-based public issue analysis and provide methodological recommendations for further research in the fields of topic modeling and natural language processing.

Future research is recommended to use a larger dataset from various social media platforms to provide a more representative public issue analysis. Furthermore, this research can be developed by comparing it with other topic modeling methods, such as Top2Vec or NMF, to achieve more optimal results. Preprocessing that is more adaptive to social media slang and language also needs to be improved. Future research could also incorporate sentiment analysis to gain a more in-depth and comprehensive understanding of public opinion trends toward the MBG program.

REFERENCES

- [1] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2020.
- [2] P. Koochemeshkian and N. Bouguila, "Integration of Neural Embeddings and Probabilistic Models in Topic Modeling Integration of Neural Embeddings and Probabilistic Models in Topic Modeling," *Appl. Artif. Intel.*, vol. 38, no. 1, 2024, doi: 10.1080/08839514.2024.2403904.
- [3] F. Bianchi, S. Terragni, and D. Hovy, "Pre-training is a Hot Topic : Contextualized Document Embeddings Improve Topic Coherence," pp. 759–766, 2021.
- [4] AB Dieng and DM Blei, "Topic Modeling in Embedding Spaces," 2017.
- [5] S. Koltcov, A. Surkov, V. Filippov, and V. Ignatenko, "Topic models with elements of neural networks : investigation of stability, coherence, and determining the optimal number of topics," pp. 1–41, 2024, doi: 10.7717/peerj-cs.1758.
- [6] L. Yijia, "Comparison of LDA and BERTopic in News Topic Modeling : A Case Study of The New York Times' Reports on China," vol. 7, no. March, pp. 47–51, 2024, doi: 10.55014/pij.v7i3.616.
- [7] MD Pratiwi, KD Tania, S. Informasi, and U. Sriwijaya, "Knowledge Discovery Through Topic Modeling on GoPartner User Reviews Using BERTopic, LDA, and NMF," vol. 9, no. 1, pp. 1–7, 2025.
- [8] Egger, J. Yu, and J. Yu, "A Topic Modeling Comparison Between LDA , NMF , Top2Vec , and BERTopic to Demystify Twitter Posts Making Sense of Social Media Using," vol. 7, no. May, pp. 1–16, 2022, doi: 10.3389/fsoc.2022.886498
- [9] Luiz and M. Owa, "Identification of Topics from Scientific Papers through Topic Modeling," pp. 541–548, 2021, doi: 10.4236/ojapps.2021.114038.
- [10] CB Pavithra and J. Savitha, "Topic Modeling for Evolving Textual Data Using LDA , HDP , NMF , BERTOPIC , and DTM With a Focus on Research Papers," vol. 5, no. 2, 2024, doi: 10.37802/joti.v5i2.618.
- [11] P. Resnik, "Improving Neural Topic Models using Knowledge Distillation," pp. 1752–1771, 2020.
- [12] A. Farea, S. Tripathi, G. Glazko, and F. Emmertstreib, "Engineering Applications of Artificial Intelligence Investigating the optimal number of topics by advanced text-mining techniques : Sustainable energy research," *Eng. Appl. Artif. Intel.*, vol. 136, no. PA, p. 108877, 2024, doi: 10.1016/j.engappai.2024.108877.
- [13] X. Wu, T. Nguyen, and AT Luu, "and challenges," *Artif. Intel. Rev.*, vol. 57, no. 2, pp. 1–30, 2024, doi: 10.1007/s10462-023-10661-7.
- [14] MR Haas and A. Heijn, "Experiments on Generalizability of BERTopic on Multi-Domain Short Text," pp. 1–3, 2020.
- [15] MC Kenton, L. Kristina, and J. Devlin, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," no. MLM, 1953.
- [16] E. Taheri and JL Junkins, "How Many Impulses Redux".
- [17] M. Röder and A. Hinneburg, "Exploring the Space of Topic Coherence Measures".
- [18] P. Aprilio, PS Nugraha, and H. Fahmi, "Hybrid Feature Combination of TF-IDF and BERT for Enhanced Information Retrieval Accuracy," vol. 08, no. 01, pp. 8–15, 2025.
- [19] DM Blei, AY Ng, and MI Jordan, "Latent Dirichlet Allocation," vol. 3, pp. 993–1022, 2003.
- [20] I. Computer and T. Hofmann, "Probabilistic Latent Semantic Indexing," pp. 50–57.
- [21] COO Optimization, "Topic Modeling as Multi-Objective Contrastive Optimization," pp. 1–20, 2024.
- [22] PMSArdinata, AAJ Permana, and INSW Wijaya, "IDENTIFICATION AND NORMALIZATION OF SLANG TEXT WITH," vol. 21, no. 1, 2024.
- [23] RLNurdiansyah and KE Dewi, "KOMPUTA: Scientific Journal of Computers and



Informatics THE EFFECT OF INFORMATION
GAIN AND WORD NORMALIZATION ON
ASPECT-BASED SENTIMENT ANALYSIS
KOMPUTA: Scientific Journal of Computers and
Informatics,” vol. 12, no. 2, 2023.

- [24] R. Li, F. González-pizarro, L. Xing, G. Murray, and G. Carenini, “Diversity-Aware Coherence Loss for Improving Neural Topic Models,” vol. 2, pp. 1710–1722, 2023.
- [25] R. Li, F. González-pizarro, L. Xing, G. Murray, and G. Carenini, “Diversity-Aware Coherence Loss for Improving Neural Topic Models,” vol. 2, pp. 1710–1722, 2023.

Deployment of a Machine Learning Based Indoor Air Quality Predictor on NodeMCU ESP32 Using EloquentTinyML

Yaddarabullah^{1*}, Umar Al Faruq²

¹Department of Informatics, Universitas Trilogi

²Department of Information System, Universitas Trilogi

Email: ¹yaddarabullah@trilogi.ac.id, ²faruq@trilogi.ac.id

Abstract Indoor air quality monitoring increasingly requires local intelligence because indoor exposure conditions can change faster than centralized systems can respond. However, many machine learning based indoor air quality predictors remain difficult to deploy on microcontrollers due to memory limits, computation constraints, network dependence, and the lack of a reproducible edge deployment workflow. This study develops an end-to-end TinyML framework for deploying a lightweight indoor air quality predictor on a NodeMCU ESP32 using EloquentTinyML and TensorFlow Lite Micro. A synthetic IAQ dataset was generated from eight environmental variables, namely temperature, humidity, PM2.5, PM10, NO2, SO2, CO, and room type, and the indoor air quality score was derived from pollutant weighted features before balancing comfort labels using SMOTE. The proposed compact MLP contains eight inputs, one hidden layer with twelve neurons, ReLU activation, dropout regularization, and a single regression output. Five-fold validation produced an average root mean square error of 5.794, mean absolute error of 4.394, and R^2 of 0.877, while the converted model required only 121 trainable parameters. These results indicate that compact TinyML deployment can provide a feasible proof-of-concept for local indoor air quality estimation, although physical sensor validation remains necessary.

Keywords - Indoor air quality, TinyML, NodeMCU ESP32, EloquentTinyML, TensorFlow Lite Micro, embedded machine learning, SMOTE

I. INTRODUCTION

Indoor air quality (IAQ) is a critical component of healthy learning spaces, offices, laboratories, and public facilities because indoor exposure to particulate matter, gaseous pollutants, temperature, and humidity directly affects comfort, cognitive performance, respiratory risk, and the perceived safety of occupants. As indoor environments are increasingly instrumented with low-cost sensors, IAQ monitoring is shifting from occasional measurement to continuous observation, enabling data-driven estimation of air condition and air comfort states at finer temporal resolution [1]-[3].

Despite this progress, IAQ prediction systems face a fundamental system constraint when they are moved from laboratory computers or cloud platforms to embedded devices. Microcontrollers such as the NodeMCU ESP32 provide affordability, portability, and wireless connectivity, but they have limited RAM, flash memory, computational throughput, and power budget. Therefore, an IAQ predictor that performs well on a computer may not be directly suitable for on-device inference unless the model is compact, the runtime is compatible with microcontroller execution, and the deployment pipeline is reproducible.

Recent literature has addressed different parts of this problem. Low-cost IAQ sensing studies emphasize sensor calibration, bias correction, and data reliability in indoor or urban environments [2], [3], [15]. IoT-based air quality studies generally focus on data acquisition, dashboard visualization, or server-side analytics, whereas TinyML studies demonstrate that neural networks can be compressed and executed on highly constrained devices using TensorFlow Lite Micro or similar runtimes [4]-[6]. In parallel, embedded AI platforms such as

Edge Impulse and MLPerf Tiny have shown that systematic benchmarking, model conversion, and deployment automation are important for bringing machine learning closer to physical sensing devices [11], [12].

However, existing methods still show several limitations for practical IAQ deployment. Cloud-based prediction introduces network dependence, possible latency, and additional communication energy. High-capacity machine learning models such as ensemble regressors can produce accurate estimates but are often less convenient to convert into a microcontroller-native inference graph. Conversely, extremely small models can be deployed more easily but may lose predictive flexibility when sensor patterns become nonlinear or when environmental context needs to be included in the input representation.

The research gap addressed in this study is therefore not only the prediction of IAQ score, but the construction of a complete and transparent path from IAQ data generation, preprocessing, model training, model conversion, C-array embedding, and firmware-level inference on NodeMCU ESP32. This gap is particularly relevant for educational and early-stage research settings where physical sensor data may not yet be available, but a reproducible proof-of-concept is needed before field instrumentation and long-term validation are conducted.

This study positions a compact multilayer perceptron (MLP) as a TinyML-oriented regression model. The MLP is selected not because it is expected to outperform every baseline on a synthetic formula-based dataset, but because it can be represented as a TensorFlow Lite Micro model, embedded into Arduino firmware, and executed through EloquentTinyML while still retaining nonlinear modeling capacity for future real sensor data. This positioning makes the proposed method a



deployment framework rather than merely an offline regression experiment.

The proposed framework integrates synthetic IAQ dataset generation, comfort-label balancing using SMOTE, feature standardization, compact MLP training, cross-validation, TensorFlow Lite conversion, C-array model embedding, and serial-monitor feedback on predicted IAQ, comfort level, inference time, free heap memory, and estimated CPU usage. The main contributions are an end-to-end TinyML deployment workflow for IAQ prediction, an analysis of the accuracy-efficiency trade-off between compact and larger models, an ablation study of preprocessing and input components, and an embedded implementation structure that can be extended to calibrated physical sensors.

The significance of this work lies in its role as a bridge between IAQ machine learning and microcontroller-based environmental intelligence. By demonstrating what can and cannot be claimed from a synthetic proof-of-concept, the study provides a careful foundation for future field deployment, sensor calibration, and long-term indoor exposure monitoring. The rest of this paper is organized as follows. Section II describes the methodology, dataset, preprocessing, model, training configuration, and evaluation metrics. Section III presents the results. Section IV discusses the findings, novelty, practical meaning, and limitations. Section V concludes the paper and outlines future work.

II. RESEARCH METHODOLOGY

A. Flowchart of Methodology

The methodology was designed as a traceable engineering pipeline rather than a single offline prediction experiment. Figure 1 summarizes the workflow from controlled dataset generation to embedded deployment. Each step has a direct implementation artifact: a synthetic dataset, a preprocessed training matrix, a compact model, cross-validation metrics, a TensorFlow Lite file, C-array model files, and Arduino firmware for on-device inference.

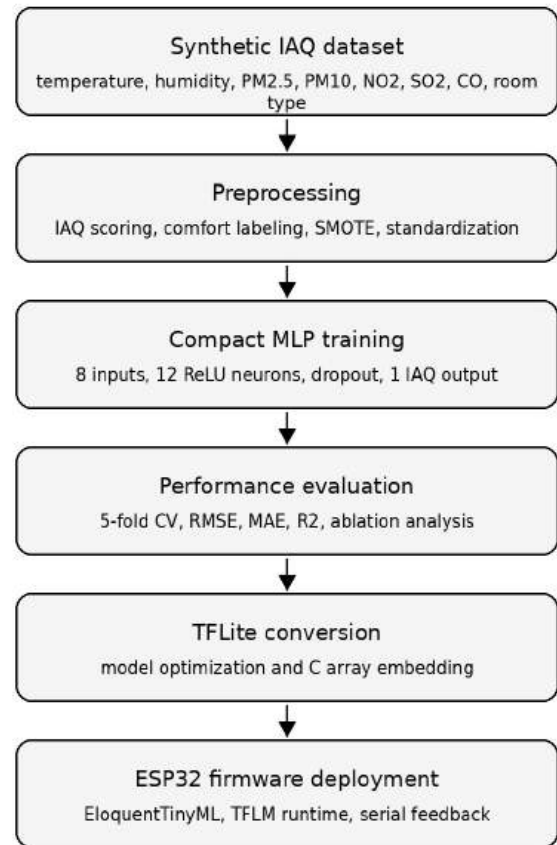


Figure. 1 Methodology flowchart of the indoor air quality TinyML deployment pipeline.

B. Dataset

The dataset was generated synthetically to reproduce the structure of an IAQ sensing problem before physical sensors are installed. The input variables were temperature, relative humidity, PM2.5, PM10, NO2, SO2, CO, and room type. The first two variables represent thermal comfort, five variables represent particulate and gaseous pollutants, and room type represents a simple spatial-context variable. The IAQ score was computed using a pollutant-weighted rule and clipped to the interval between 0 and 100 as shown in Equation (1).

$$\text{IAQ} = \text{clip} \left(100 - (0.5 \text{ PM}_{2.5} + 0.3 \text{ PM}_{10} + 0.7 \text{ NO}_2 + 3 \text{ CO}), 0, 100 \right) \quad (1)$$

Table 1. Descriptive characteristics of the synthetic IAQ dataset

Feature	Mean	Std.	Min.	Max.
temp	26.135	2.965	16.276	37.558
hum	49.936	10.054	19.805	89.262
PM2.5	29.280	14.995	-17.651	76.694
PM10	64.866	31.003	-45.651	170.872
NO2	22.025	7.908	-9.379	49.019
SO2	6.923	2.978	-3.127	17.133
CO	1.529	0.800	-0.966	4.243
room_type	2.500	1.141	1.000	4.000
IAQ	45.895	13.469	4.052	86.362

Table 1 shows that the synthetic generator creates a wide but controlled environmental space. The pollutant variables have sufficient variance to produce a broad IAQ range, while the temperature and humidity variables preserve the comfort context used in the labeling rule. Some generated pollutant values are negative because unconstrained normal distributions were used

in the original material; this is acceptable for pipeline testing but should be replaced by bounded or sensor-calibrated distributions in field validation.

C. Preprocessing

Preprocessing consisted of comfort label construction, class balancing, feature scaling, and train-test partitioning through cross-validation. Comfort labels were generated from the combination of IAQ score, temperature, and humidity. Class 0 represents very comfortable air, class 1 comfortable air, class 2 slightly uncomfortable air, and class 3 uncomfortable air. Because the deterministic rule produced highly imbalanced classes, SMOTE was applied to generate synthetic minority samples in the feature space before regression training [9].

Table 2. Comfort-label distribution before and after SMOTE

Class	Meaning	Before	After
0	Very comfortable	7	1103
1	Comfortable	167	1103
2	Slightly uncomfortable	723	1103
3	Uncomfortable	1103	1103

Table 2 indicates that the original dataset was dominated by the uncomfortable class, while the very comfortable class was extremely rare. SMOTE equalized all comfort labels to 1,103 samples each. The technical implication is that the model is trained on a more balanced representation of comfort states, although the synthetic minority samples must later be checked against realistic sensor behavior.

D. Proposed Model

The proposed model is a compact MLP designed for TensorFlow Lite Micro compatibility. It accepts eight environmental inputs and produces one continuous IAQ score. The hidden layer contains twelve neurons with ReLU activation, followed by dropout regularization in the Keras implementation and a linear output neuron. This architecture intentionally prioritizes a small memory footprint and simple inference graph over maximum offline accuracy.

Table 3 clarifies the main architectural advantage of the proposed model. With only 121 parameters, the model can be stored as a compact C array and executed with a small tensor arena. The trade-off is that the hidden representation is intentionally narrow, which can reduce accuracy compared with larger neural networks or ensemble models on the training computer.

Table 3. Proposed compact MLP architecture

Layer	Configuration	Output
Input layer	8 environmental variables	8
Dense hidden layer	12 neurons, ReLU	12
Dropout	Rate 0.1 during training	12
Output layer	1 linear neuron for IAQ regression	1
Trainable parameters	Approximately 121 weights and biases	-

E. Training Configuration

Training was performed using five-fold cross-validation with shuffled folds and a fixed random seed to improve reproducibility. StandardScaler was used before MLP training so

that features with different physical units did not dominate the optimization process. For comparison, several conventional regressors were evaluated on the same balanced dataset, including linear regression, random forest, gradient boosting, XGBoost, a compact MLP, and a larger MLP. The comparison is used to interpret accuracy-efficiency trade-offs rather than to claim that the smallest model is always the most accurate model.

Table 4. Training and deployment configuration

Component	Configuration
Validation strategy	5-fold cross-validation, shuffled, random state = 42
Training target	Continuous IAQ score
Primary model	Compact MLP with 12 hidden neurons
Optimizer/loss	Adam optimizer and MSE loss in the original Keras workflow
Feature scaling	StandardScaler fitted within each training fold
Deployment conversion	Keras model to TensorFlow Lite, then C array
Runtime target	EloquentTinyML with TensorFlow Lite Micro on ESP32

F. Performance Evaluation

Model accuracy was evaluated using RMSE, MAE, and coefficient of determination. RMSE emphasizes larger prediction errors, MAE provides an interpretable average absolute deviation in IAQ-score units, and R2 measures the proportion of target variance explained by the model. For the embedded stage, the firmware design also reports predicted IAQ, comfort level, inference time, free heap memory, and estimated CPU usage through the serial monitor. In this paper, embedded metrics are treated as implementation outputs because the uploaded material still uses simulated sensor input rather than calibrated physical sensors.

III. RESULTS

A. Dataset Characteristics

The synthetic dataset contains 2,000 initial samples and 4,412 samples after SMOTE balancing. The key dataset characteristic is the strong imbalance produced by the comfort rule: only seven initial samples belonged to the very comfortable class, while 1,103 samples belonged to the uncomfortable class. This pattern is technically meaningful because IAQ comfort rules often create rare boundary conditions, and a model trained without balancing tends to underrepresent those rare conditions.

The IAQ target had a mean of 45.895, a standard deviation of 13.469, and a range from 4.052 to 86.362. This distribution shows that the generated environment is dominated by moderate to low IAQ values rather than very clean air. In the research context, this makes the dataset useful for testing whether a lightweight model can estimate degraded air conditions, but it also shows why field validation is required before the system can be used for health-related decisions.

Table 5 should be interpreted with the characteristics of the synthetic target in mind. Linear regression achieved nearly perfect reconstruction because the IAQ score was generated from a linear pollutant-weighted formula with clipping. This result is an internal consistency check for the synthetic generator rather than evidence that a linear model will be sufficient for real indoor environments. Tree ensembles and boosting models

produced strong accuracy, but they are less straightforward to deploy through TensorFlow Lite Micro and EloquentTinyML than a compact neural network.

B. Model Performance Comparison

Table 5. Model performance comparison

Model	RMSE	MAE	R2	Complexity	TinyML readiness
Linear regression	0.000	0.000	1.000	9	Very high
Random forest	2.254	1.459	0.981	Tree ensemble	Low
Gradient boosting	2.064	1.484	0.984	Tree ensemble	Low
XGBoost	2.405	1.769	0.979	Tree ensemble	Low
Larger MLP, 32-16 neurons	0.301	0.157	1.000	833	Medium
Compact MLP, 12 neurons (proposed)	5.794	4.394	0.877	121	High

The proposed compact MLP produced an average RMSE of 5.794, MAE of 4.394, and R2 of 0.877. Its main advantage is not absolute dominance in offline accuracy, but the balance between acceptable regression performance, small parameter count, nonlinear capacity, and direct compatibility with the TFLite-based deployment path. The larger MLP achieved substantially higher accuracy, but it increased parameter count from 121 to about 833, illustrating the trade-off between predictive capacity and embedded simplicity.

C. Analysis of Key Advantage of the Proposed Model

Table 6 links the quantitative model design to the practical IAQ context. A model with 121 parameters is small enough for embedded experimentation, while the output structure remains useful for occupants because the IAQ score is translated into a comfort label. This design also supports future extension: once calibrated sensor data are available, the same deployment path can be reused with updated training data and a revised model.

Table 6. Key advantages of the proposed model

Aspect	Finding	Technical implication
Model compactness	121 parameters	Small memory footprint and simple C-array storage
Deployment path	Keras to TFLite to model_data.cc/h	Traceable conversion from Python to Arduino
Runtime compatibility	EloquentTinyML and TFLM	Microcontroller-oriented inference interface
Output interpretability	IAQ score and comfort label	Combines numerical prediction and human-readable feedback
Extensibility	Eight environmental inputs	Can be connected to real IAQ sensors in future work

D. Ablation Study Results

Table 7 provides a deeper explanation of how preprocessing and input selection affect the compact model. Removing SMOTE reduced R2 from 0.877 to 0.608, showing that balancing the comfort-conditioned samples helps the model learn across a broader IAQ-comfort space. Removing feature

scaling also degraded performance, confirming that optimization becomes less stable when variables with different units and magnitudes are provided directly to a small MLP.

Table 7. Ablation study

Configuration	RMSE	MAE	R2
Full compact MLP with SMOTE and scaling	5.794	4.394	0.877
Without SMOTE	8.392	6.853	0.608
Without contextual room_type	6.753	5.080	0.834
Pollutant-only inputs	4.558	3.392	0.924
Without feature scaling	6.354	4.565	0.852

The pollutant-only configuration achieved the highest ablation R2 because the synthetic IAQ formula depends directly on PM2.5, PM10, NO2, and CO, while temperature, humidity, SO2, and room type do not directly enter Equation (1). This finding is not a reason to remove contextual variables from the final framework. Instead, it reveals the trade-off between formula-specific accuracy and context-aware deployment. In real IAQ systems, thermal variables and room context are important for comfort interpretation, sensor placement, and model generalization.

E. Additional Metrics Results

Table 8 shows that the proposed compact model is relatively stable across validation folds. RMSE varies only from 5.705 to 5.956, and R2 remains between 0.869 and 0.886. This narrow spread indicates that the result is not dominated by a single favorable split. For an embedded proof-of-concept, such stability is important because the model is expected to operate repeatedly on streaming sensor input rather than on a single static test subset.

Table 8. Fold-level metrics of the proposed compact MLP

Fold	RMSE	MAE	R2
1	5.764	4.322	0.886
2	5.705	4.433	0.877
3	5.739	4.306	0.884
4	5.805	4.388	0.869
5	5.956	4.521	0.870
Average	5.794	4.394	0.877

Figure 2 confirms the same pattern visually. Predictions follow the ideal diagonal trend, although the compact model compresses some high-variance regions because it has only one narrow hidden layer.

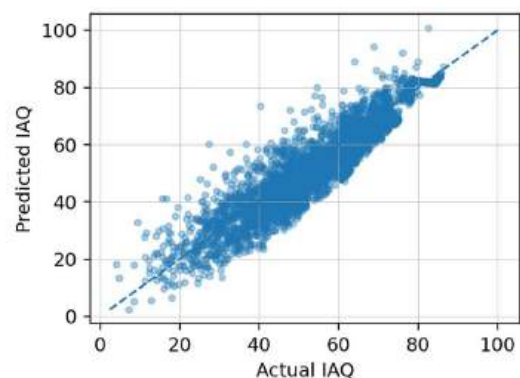


Figure. 2 Actual versus predicted IAQ scores for the compact MLP across validation folds.

This behavior is consistent with the design objective: the model is sufficiently expressive for initial IAQ estimation while remaining small enough for deployment on a microcontroller. Table 9 separates measured training-side metrics from implementation-oriented estimates. The parameter memory estimate is small, but actual flash usage and RAM usage will also depend on the TensorFlow Lite Micro runtime, operator resolver, tensor arena size, and firmware overhead. Therefore, the serial-monitor outputs are necessary in the next physical validation stage to quantify real latency, free heap, CPU load, and power implications on the ESP32 board.

Table 9. Additional implementation-oriented metrics

Metric	Value/status	Interpretation
Trainable parameters	121	Derived from the compact MLP architecture
Approximate float parameter memory	484 bytes	121 parameters multiplied by 4 bytes
Approximate int8 parameter memory	121 bytes	Analytical estimate after full quantization
Firmware serial outputs	IAQ, comfort, inference time, heap, CPU	Implemented as runtime feedback fields
Physical sensor validation	Not yet measured	Required before operational deployment

Figure 3 presents the serial monitor output generated by the ESP32 after the TensorFlow Lite Micro model was loaded and executed through the EloquentTinyML framework. The displayed environmental input consists of eight simulated sensor features, namely temperature, humidity, PM2.5, PM10, NO2, SO2, CO, and room type.

```

===== ENV INPUT =====
21.50
50.80
29.50
68.10
19.10
2.00
0.00
3.00

===== IAQ RESULT =====
Predicted IAQ: -94.83
Comfort Level: Sangat Nyaman (Excellent)
🕒 Inference Time (µs): 25
🗑️ Free Heap Memory (bytes): 327580
💻 Estimated CPU Usage (%): 0.00
=====

```

Figure. 3 The Output of Model in ESP 32.

The result demonstrates that the embedded firmware was able to pass the input vector to the deployed model, execute on-device inference, and return a predicted IAQ value together with a human-readable comfort label. In this execution cycle, the model produced an IAQ prediction of -94.83 and mapped the condition into “Sangat Nyaman (Excellent)”. The inference time was recorded at 25 microseconds, with 327,580 bytes of free heap memory remaining and an estimated CPU usage of 0.00%. These values indicate that the deployed model can run with very low computational overhead on the ESP32 platform.

However, the negative IAQ value also shows the importance of adding output post-processing, such as clipping the predicted score into the valid IAQ range, before interpreting the result as an environmental condition. Therefore, Figure 3 should be interpreted not as final field validation, but as evidence that the TinyML deployment pipeline, model loading, inference execution, memory monitoring, and serial-based feedback mechanism function correctly on the target embedded device.

IV. DISCUSSION

The central problem addressed in this study is the gap between IAQ prediction as an offline machine learning task and IAQ prediction as a deployable embedded intelligence function. In practical indoor environments, an IAQ prediction system is expected to operate locally, respond with low latency, remain inexpensive, and continue functioning even when network connectivity is limited or unstable. This requirement shifts the evaluation focus from predictive accuracy alone toward a broader embedded-AI perspective that includes model compactness, conversion feasibility, firmware integration, runtime observability, and the semantic validity of the output. Therefore, the proposed framework was not designed merely to optimize an IAQ regression score in Python, but to examine whether a trained model can be transformed into an executable TinyML function on a resource-constrained ESP32 device.

The experimental setup was intentionally organized as a proof-of-concept. Synthetic data were used to control the relationship among pollutant variables, IAQ scores, and comfort labels, while SMOTE was applied to reduce the dominance of majority comfort conditions during model training. Five-fold cross-validation was used to reduce split-specific bias and to obtain a more stable estimate of regression performance. The deployment pathway included TensorFlow Lite conversion, model transformation into C-array files, and Arduino firmware integration using EloquentTinyML with the TensorFlow Lite Micro backend. This workflow represents an important intermediate stage between offline IAQ modeling and physical deployment with calibrated sensors because it verifies whether the model can be embedded, called, and monitored within the target microcontroller environment.

The core quantitative result shows that the compact MLP achieved RMSE 5.794, MAE 4.394, and R2 0.877 on the balanced synthetic dataset. These results indicate that a very small neural architecture is still capable of approximating the pollutant-driven IAQ score under the controlled synthetic formulation. Compared with larger or ensemble-based models, the compact MLP sacrifices part of its offline predictive capacity but offers a more favorable trade-off for embedded deployment. This trade-off is central to TinyML-based IAQ systems, where a model must be sufficiently accurate to support local interpretation but sufficiently small to satisfy memory, latency, and firmware integration constraints. The fold-level stability also suggests that the model does not depend excessively on a single train-test partition, which is important for later streaming inference scenarios.

The model comparison should, however, be interpreted carefully. Because the IAQ score in the synthetic dataset is generated using a deterministic pollutant-weighted formula, linear regression can act as an artificial upper-bound baseline rather than a realistic representation of IAQ prediction in real

buildings. In actual indoor environments, the relationship between pollutant concentration, thermal conditions, room context, ventilation, sensor position, occupant activity, and perceived air comfort is unlikely to remain purely linear. Thus, the strong performance of a linear baseline should not be generalized as evidence that linear modeling is sufficient for real IAQ intelligence. Instead, it highlights the limitation of the current synthetic generator and reinforces the need for physical sensor data, nonlinear environmental dynamics, and temporal validation in future work.

A deeper analytical insight is provided by the ablation study. SMOTE substantially improves the compact MLP, indicating that the distribution of comfort-related conditions affects the ability of a small model to represent less frequent states. Feature scaling also improves model behavior, confirming that preprocessing remains important even when the final target is a microcontroller-based deployment. The pollutant-only input configuration performs better than the full input configuration because the current IAQ score formulation is directly driven by PM2.5, PM10, NO2, and CO. However, this does not imply that temperature, humidity, and room type should be removed from a deployable IAQ-comfort framework. These contextual variables are important for interpreting comfort, supporting future physical validation, and extending the system from simple IAQ score prediction toward more realistic indoor exposure and comfort reasoning.

The ESP32 output shown in Figure 3 provides an important firmware-level finding. The serial monitor confirms that the deployed model can receive an eight-dimensional environmental input vector, execute inference on the ESP32, and return runtime information such as predicted IAQ, comfort label, inference time, free heap memory, and estimated CPU usage. The recorded inference time of 25 microseconds, remaining heap memory of 327580 bytes, and estimated CPU usage of 0.00% suggest that the deployed model imposes a very small computational burden on the device. This result supports the technical feasibility of executing the compact model within the ESP32 environment. Nevertheless, these metrics should be interpreted as initial runtime indicators rather than final embedded performance claims, because the current test uses simulated inputs and does not yet include sensor acquisition latency, communication overhead, power consumption, long-duration execution, or multi-cycle stability analysis.

More importantly, Figure 3 reveals a critical semantic-validity issue. The model produced a predicted IAQ value of -94.83, while the firmware mapped this output to “Sangat Nyaman (Excellent).” This finding shows that successful model execution on the microcontroller does not automatically guarantee meaningful environmental interpretation. The negative prediction is outside the intended IAQ range and should be treated as an invalid or out-of-range output. The incorrect assignment of an “Excellent” label occurs because the firmware-level label-mapping logic evaluates threshold conditions without first constraining or validating the regression output. This demonstrates the need for post-processing safeguards, including clipping the predicted IAQ score into a valid range, rejecting invalid values, checking input normalization consistency, and separating numerical inference from semantic decision rules. In embedded IAQ systems, such safeguards are not secondary

details; they are necessary to prevent misleading comfort feedback.

This finding also clarifies the distinction between deployment feasibility and deployment validity. From a deployment-feasibility perspective, the proposed pipeline successfully moves the trained model from Python into Arduino firmware and executes it on the ESP32 using EloquentTinyML. From a deployment-validity perspective, however, the observed negative IAQ output indicates that further refinement is required before the system can be interpreted as a reliable IAQ decision-support tool. The firmware should include bounded regression output, robust label mapping, sensor plausibility checks, and calibration-aware preprocessing. In addition, the same preprocessing used during model training, particularly feature scaling, must be consistently represented in the embedded inference stage. Without this consistency, a model that performs well during cross-validation may produce unrealistic values when deployed.

Practically, the proposed framework is valuable as an instructional and research prototype for embedded IAQ intelligence. It demonstrates how an IAQ model can move from dataset generation and training into firmware-level inference, how serial output can be used to inspect model behavior, and how runtime observability can be prepared through inference time, memory availability, and CPU usage fields. The framework also exposes the types of engineering issues that emerge only after deployment, such as invalid regression ranges and inappropriate threshold-based label mapping. These findings are useful because they show that embedded AI development requires not only model training but also output validation, runtime monitoring, and firmware-level control logic.

The novelty of this study is positioned at the system level by integrating IAQ regression, comfort-label balancing, compact MLP modeling, TensorFlow Lite conversion, C-array embedding, and EloquentTinyML-based ESP32 deployment into a single reproducible workflow. Rather than proposing a new neural network architecture, this study demonstrates how a lightweight IAQ predictor can be designed, trained, converted, embedded, executed, and evaluated on a resource-constrained microcontroller. This contribution is different from studies that mainly focus on sensor monitoring, cloud-based analytics, or offline prediction accuracy. However, several limitations remain. The dataset is synthetic and does not fully represent real indoor conditions, the data generator may produce unrealistic values, and the embedded test still uses simulated inputs instead of calibrated physical sensors. The negative IAQ value shown in Figure 3 also indicates the need for stronger output validation, such as clipping the predicted score before assigning a comfort label, while the reported CPU usage of 0.00% should be interpreted cautiously because it may be affected by rounding or measurement resolution. Future work should use calibrated indoor sensor data, apply bounded preprocessing, maintain consistent normalization between training and deployment, measure real ESP32 latency and energy consumption, and evaluate long-term stability across different room types.

V. CONCLUSION

This paper revised and extended the IAQ TinyML deployment study into an academic framework for implementing a machine learning based IAQ predictor on NodeMCU ESP32

using EloquentTinyML. The proposed workflow includes synthetic IAQ dataset generation, comfort-label construction, SMOTE balancing, feature scaling, compact MLP training, five-fold validation, TensorFlow Lite conversion, C-array model embedding, and Arduino firmware integration. The compact MLP achieved an average RMSE of 5.794, MAE of 4.394, and R2 of 0.877 while using only 121 trainable parameters.

The comparison and ablation results show that the model is not the most accurate offline alternative on a deterministic synthetic formula, but it provides a practical balance between regression performance, model compactness, nonlinear extensibility, and microcontroller deployment readiness. The framework is therefore suitable as a proof-of-concept and educational platform for embedded IAQ prediction. Future research should replace synthetic data with calibrated sensor measurements, constrain the data generator to physically realistic pollutant ranges, quantify real ESP32 latency and memory usage, and evaluate model robustness under long-term indoor environmental variation.

REFERENCES

- [1] World Health Organization. (2021). WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: World Health Organization.
- [2] Hayward, I., Hirst, E., Stewart, G., & Acton, W. J. F. (2024). Low-cost air quality sensors: Biases, corrections and challenges in their comparability. *Atmosphere*, 15(12), 1523. DOI: 10.3390/atmos15121523.
- [3] Chojer, H., Branco, P. T. B. S., Martins, F. G., Alvim-Ferraz, M. C. M., & Sousa, S. I. V. (2020). Development of low-cost indoor air quality monitoring devices: Recent advancements. *Science of the Total Environment*, 727, 138385. DOI: 10.1016/j.scitotenv.2020.138385.
- [4] Soro, S. (2021). TinyML for ubiquitous edge AI. arXiv preprint arXiv:2102.01255.
- [5] Abadade, Y., Temouden, A., Bamoumen, H., Benamar, N., Chtouki, Y., & Hafid, A. S. (2023). A comprehensive survey on TinyML. *IEEE Access*, 11, 96892-96922. DOI: 10.1109/ACCESS.2023.3294111.
- [6] David, R., Duke, J., Jain, A., Janapa Reddi, V., Jeffries, N., Li, J., Kreeger, N., Nappier, I., Natraj, M., Regev, S., Rhodes, R., Wang, T., & Warden, P. (2021). TensorFlow Lite Micro: Embedded machine learning for TinyML systems. *Proceedings of Machine Learning and Systems*, 3, 800-811.
- [7] EloquentArduino. (2024). EloquentTinyML: An eloquent interface to TensorFlow Lite for Microcontrollers. Arduino Libraries Documentation.
- [8] Espressif Systems. (2025). ESP32 Series Datasheet. Shanghai: Espressif Systems.
- [9] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- [10] Warden, P., & Situnayake, D. (2019). TinyML: Machine learning with TensorFlow Lite on Arduino and ultra-low-power microcontrollers. Sebastopol, CA: O'Reilly Media.
- [11] Banbury, C., Reddi, V. J., Torelli, P., Holleman, J., Jeffries, N., Kiraly, C., Montino, P., Kanter, D., Ahmed, S., Pau, D., Thakker, U., Torrini, A., Warden, P., Cordaro, J., Di Guglielmo, G., Duarte, J., Gibellini, S., Parekh, V., Tran, H., Tran, N., Wenxu, N., & Xuesong, X. (2021). MLPerf Tiny benchmark. arXiv preprint arXiv:2106.07597.
- [12] Hymel, S., Banbury, C., Situnayake, D., Elum, A., Ward, C., Kelcey, M., Baaijens, M., Majchrzycki, M., Plunkett, J., Tischler, D., Grande, A., Moreau, L., Maslov, D., Beavis, A., Jongboom, J., & Reddi, V. J. (2023). Edge Impulse: An MLOps platform for Tiny Machine Learning. *Proceedings of Machine Learning and Systems*, 5.
- [13] Guo, Z., Wang, X., & Ge, L. (2023). Classification prediction model of indoor PM_{2.5} concentration using CatBoost algorithm. *Frontiers in Built Environment*, 9, 1207193. DOI: 10.3389/fbuil.2023.1207193.
- [14] Yu, W., Nakisa, B., Loke, S. W., Stevanovic, S., Guo, Y., & Rastgo, M. N. (2024). Indoor PM_{2.5} forecasting and the association with outdoor air pollution: A modelling study based on sensor data in Australia. arXiv preprint arXiv:2405.07404.
- [15] Desouza, P., Kahn, R., Stockman, T., Obermann, W., Crawford, B., Wang, A., Crooks, J., Li, J., & Kinney, P. (2022). Calibrating networks of low-cost air quality sensors. *Atmospheric Measurement Techniques*, 15, 6309-6328. DOI: 10.5194/amt-15-6309-2022.
- [16] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. DOI: 10.1145/2939672.2939785.
- [17] Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. DOI: 10.1023/A:1010933404324.

Development of a Marker-Based Augmented Reality Application for Computer Hardware Learning Using the Multimedia Development Life Cycle

Muhamad Adila Syahputra^{1*}, Ade Syahputra², Silvester Dian Handy Permana³

^{1,2,3}Department of Informatics, Faculty of Science, Engineering, and Design, Universitas Trilogi
Email: ¹adila.syahputra@trilogi.ac.id, ²adesyahputra@trilogi.ac.id, ³handy@trilogi.ac.id

Abstract – Understanding computer hardware is a fundamental competency for vocational students. However, limited access to physical hardware components often restricts practical learning activities in vocational schools. This study aims to develop and evaluate VirtuComp, a marker-based Augmented Reality (AR) application for computer hardware learning using the Multimedia Development Life Cycle (MDLC) methodology. The application visualizes seven essential computer hardware components as interactive three-dimensional (3D) objects to support students in recognizing and understanding hardware structures through an immersive learning experience. The development process followed the six phases of the MDLC framework. The application was evaluated through functional performance testing and usability assessment involving 68 tenth-grade students from SMK Muhammadiyah 02 Cileungsi using the User Experience Questionnaire (UEQ). The functional evaluation demonstrated that the marker-based tracking system operated reliably under adequate lighting conditions with an optimal detection distance of 4–12 cm. The usability evaluation produced positive results across all six UEQ dimensions. Efficiency achieved a score of 1.89, which was classified as Excellent. Attractiveness (1.74), Perspicuity (1.78), and Stimulation (1.38) were rated as Good, while Dependability (1.40) and Novelty (0.96) were categorized as Above Average. These findings indicate that the proposed application provides a reliable and user-friendly learning medium that supports computer hardware education, particularly in vocational schools with limited laboratory resources.

Keywords – *Augmented Reality, Multimedia Development Life Cycle, User Experience Questionnaire, Computer Hardware, Central Processing Unit.*

I. INTRODUCTION

Mastery of computer hardware is a fundamental competency for students enrolled in the Computer and Network Engineering (TKJ) vocational program. Understanding the structure, functions, and interrelationships of hardware components is essential for developing the practical skills required in computer maintenance and system assembly. However, many vocational schools face challenges in providing sufficient and up-to-date laboratory equipment due to budget limitations. Consequently, students often have limited opportunities to interact directly with physical hardware, which may reduce the effectiveness of practical learning activities [1][2].

Recent advances in educational technology have introduced Augmented Reality (AR) as an effective medium for enhancing interactive learning experiences. Previous studies have reported that AR improves students' motivation, engagement, and conceptual understanding by integrating virtual three-dimensional objects into real-world environments. In parallel, the Multimedia Development Life Cycle (MDLC) has been widely adopted as a systematic framework for developing multimedia-based educational applications because it provides structured stages for design, implementation, and evaluation [3][4]. Furthermore, usability evaluation using standardized instruments such as the User Experience Questionnaire (UEQ) has become increasingly important for assessing the quality of educational applications from the users' perspective [13].

Despite these developments, existing studies have primarily emphasized the implementation of AR technology or the development of multimedia learning applications without comprehensively integrating a systematic development framework and standardized usability evaluation in the context of vocational computer hardware education. In addition, studies focusing specifically on the visualization of internal computer hardware components for vocational learning remain limited. This gap indicates the need for an educational application that not only provides realistic three-dimensional visualization but also demonstrates both technical reliability and positive user experience through empirical evaluation [6][7][14][15].

To address this gap, this study develops VirtuComp, a marker-based Augmented Reality application for learning computer hardware using the Multimedia Development Life Cycle (MDLC) methodology. The application visualizes seven essential computer hardware components in an interactive three-dimensional environment and is evaluated through technical performance testing and the User Experience Questionnaire (UEQ) [5]. The main contribution of this study is the integration of a structured multimedia development framework with comprehensive usability evaluation to provide an effective interactive learning medium for vocational computer hardware education. The findings are expected to support technology-enhanced learning by offering an accessible alternative for schools with limited laboratory resources while improving students' learning experience through immersive visualization.



II. RESEARCH METHODOLOGY

2.1. Research Design and Object

This study employed a research and development (R&D) approach to design, develop, and evaluate a marker-based Augmented Reality application for computer hardware learning. The development process followed the Multimedia Development Life Cycle (MDLC) methodology, which provides a systematic framework for multimedia application development through six sequential phases: Concept, Design, Material Collecting, Assembly, Testing, and Distribution. The proposed application was subsequently evaluated to assess its technical performance and user experience.

The research was conducted at SMK Muhammadiyah 02 Cileungsi, with tenth-grade students from the Computer and Network Engineering (TKJ) program serving as the research participants. These participants were selected because the developed application was specifically designed to support the learning of computer hardware components within the vocational school curriculum.

Figure 1 illustrates the overall research methodology adopted in this study. The research process began with problem identification and requirement analysis through field observation and literature review, followed by application development using the Multimedia Development Life Cycle (MDLC) method [3]. The final stage involved technical performance evaluation through marker detection testing under different environmental conditions, as well as usability evaluation using the User Experience Questionnaire (UEQ) to assess users' perceptions of the proposed application. This systematic research framework ensured that both the technical functionality and the educational usability of the developed application were comprehensively evaluated.

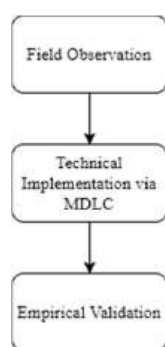


Figure. 1 Research Methodology.

2.2. Development Framework (MDLC)

The proposed application was developed using the Multimedia Development Life Cycle (MDLC) methodology, which provides a structured framework for developing multimedia-based applications. The MDLC approach consists of six sequential phases that ensure the systematic design, implementation, and evaluation of the proposed Augmented Reality application [3].

1. **Concept:** This phase identified the instructional objectives, target users, and functional requirements of the application, as well as the

computer hardware components to be visualized in the learning environment.

2. **Design:** The system architecture, user interface, application workflow, and marker-based interaction mechanism were designed to support intuitive navigation and effective visualization of three-dimensional objects.
3. **Material Collecting:** Learning materials, three-dimensional models, textures, hardware specifications, and multimedia assets were collected and prepared for integration into the application.
4. **Assembly:** All multimedia assets were integrated into the development environment using the Unity game engine and the Vuforia Software Development Kit (SDK). Interactive features, including marker recognition, object visualization, and user interaction, were implemented during this phase.
5. **Testing:** The application underwent functional and technical evaluations to verify marker detection accuracy, object visualization stability, and application performance under different operating conditions. Usability was subsequently assessed using the User Experience Questionnaire (UEQ).
6. **Distribution:** The completed application was packaged as an Android application package (APK) and deployed for evaluation by the target users in the learning environment.

Figure 2 illustrates the six phases of the Multimedia Development Life Cycle (MDLC) implemented in this study. The framework provides a systematic development process, beginning with concept formulation and application design, followed by multimedia asset integration, system implementation, testing, and final deployment. The adoption of the MDLC methodology ensured that the proposed Augmented Reality application was developed in a structured manner while meeting both functional requirements and educational objectives.

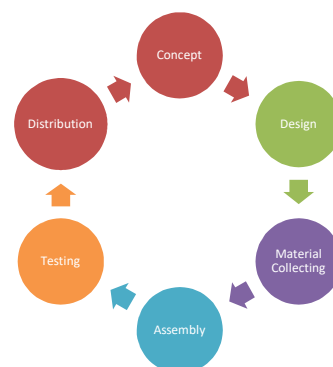


Figure. 2 Stages of the Multimedia Development Life Cycle (MDLC) Method.

2.3. Usability Evaluation

The usability of the proposed Augmented Reality application was evaluated using the User Experience

Questionnaire (UEQ), a standardized instrument widely employed to assess users' perceptions of interactive systems. The UEQ measures six dimensions of user experience: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty, providing a comprehensive evaluation of both pragmatic and hedonic aspects of system usability.

The evaluation was conducted after participants completed the learning activities using the developed application. A total of 68 tenth-grade students from the Computer and Network Engineering (TKJ) program at SMK Muhammadiyah 02 Cileungsi participated in the evaluation. The sample size was determined using the Slovin formula to ensure that the selected respondents adequately represented the target population.

The collected UEQ responses were analyzed by calculating the mean score for each evaluation dimension and comparing the results with the established UEQ benchmark categories. This benchmarking process enabled the usability and user experience of the proposed application to be interpreted according to internationally recognized evaluation standards, thereby providing empirical evidence of the application's effectiveness and user acceptance.

III. RESULTS AND DISCUSSION

3.1. Application Implementation and 3D Visualization

The implementation phase resulted in the development of VirtuComp, an Android-based Augmented Reality (AR) application designed to facilitate the learning of computer hardware components through interactive three-dimensional visualization [8]. The application was developed following the Multimedia Development Life Cycle (MDLC) framework and integrates marker-based AR technology to present virtual hardware models in a real-world environment. This implementation enables users to explore computer hardware interactively, providing a more engaging and immersive learning experience than conventional two-dimensional instructional media.



Figure. 3 Application UI.

Figure 3 presents the implemented Android application user interface. The interface consists of the Instruction and Main Menu screens, which serve as the primary interaction points for users. The Instruction screen provides guidance on operating the application, while the Main Menu offers access to the application's core functions, including AR visualization, learning materials, and additional features. The interface was designed with an intuitive navigation

structure to enhance usability and facilitate efficient interaction.

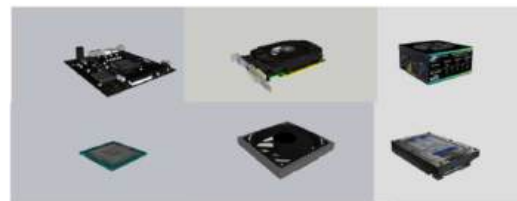


Figure. 4 3D Visualization.

Figure 4 illustrates the three-dimensional visualization of computer hardware components implemented within the application. The developed models, including the motherboard and other CPU components, were designed with realistic geometric structures and textures to closely resemble actual hardware. The integration of interactive 3D visualization allows users to observe each component from multiple perspectives, thereby improving conceptual understanding of hardware structures and supporting a more immersive learning process [9][11].



Figure. 5 AR Markers.

Figure 5 presents a representative AR marker used in the proposed marker-based Augmented Reality application. The marker functions as the primary recognition target, enabling the system to detect and track physical images before rendering the corresponding three-dimensional computer hardware model together with its instructional information. This marker-based recognition mechanism ensures stable object visualization, accurate tracking, and seamless interaction, thereby enhancing both the reliability of the application and the overall learning experience [12].

3.2. Experimental Results of Performance Evaluation

The performance evaluation was conducted to assess the effectiveness of the marker-based tracking mechanism under different environmental conditions, including variations in detection distance and lighting intensity. The evaluation aimed to determine the operating conditions that provide stable object recognition and visualization, thereby ensuring the reliability of the proposed Augmented Reality application during instructional activities.

Table 1 summarizes the experimental results obtained from the marker detection tests. The results indicate that the

application achieved its most stable performance when the marker was positioned at a distance ranging from 4 cm to 12 cm from the device camera. Within this range, the system consistently recognized the marker and rendered the corresponding three-dimensional object with accurate alignment and minimal tracking instability. At shorter distances, the marker occupied a larger portion of the camera frame, while distances within the optimal range provided sufficient image detail for reliable feature extraction and tracking.

Table 1. Testing Parameters

Testing		Good	Fair	Poor
Distance (cm)	4	✓		
	8	✓		
	12	✓		
	27		✓	
	50		✓	
Tilt Angle	0°	✓		
	45°	✓		
	70°	✓		
Lighting Condition	Bright	✓		
	Dark		✓	

As the detection distance increased beyond approximately 27 cm, the tracking performance gradually deteriorated. This reduction in accuracy can be attributed to the decreasing image resolution of the marker captured by the camera, which limits the number of distinguishable feature points available for the computer vision algorithm. Consequently, the system experienced delayed recognition or failed to render the virtual object consistently.

Lighting conditions also had a significant influence on tracking performance. Stable visualization was achieved under adequate illumination, whereas darker environments reduced image contrast and degraded the visibility of marker features. Since marker-based Augmented Reality relies on feature detection and image matching, insufficient lighting negatively affected the system's ability to distinguish the marker from the surrounding background, resulting in unstable tracking and reduced visualization accuracy.

Overall, the experimental findings demonstrate that the proposed application provides reliable marker recognition and stable three-dimensional visualization when operated within the recommended distance and lighting conditions. These results confirm the technical feasibility of employing marker-based Augmented Reality as an effective instructional medium for computer hardware learning.

3.3. Usability Evaluation and Empirical Findings

The usability of the proposed application was evaluated using the User Experience Questionnaire (UEQ), which measures user perceptions across six dimensions: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. The evaluation aimed to assess both the usability and overall user experience of the marker-based Augmented Reality application after users completed the learning activities [10].

The UEQ Benchmark graph and the corresponding mean scores are presented in Figures 6 and 7, providing a comprehensive comparison between the obtained results and the established UEQ benchmark categories. These

findings offer quantitative evidence of the application's usability and interaction quality, allowing its performance to be evaluated against a large collection of benchmarked interactive systems.

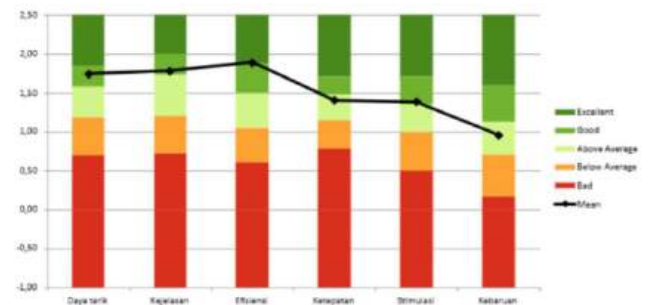


Figure. 6 UEQ Benchmark Graph.

The benchmark analysis indicates that the application achieved a mean score of 1.89 for the Efficiency dimension, placing it in the Excellent category. This result demonstrates that users were able to complete learning tasks efficiently, navigate the interface with minimal effort, and interact with the application's features without unnecessary complexity. The high efficiency score reflects the effectiveness of the interface design and interaction flow in supporting computer hardware learning.

The remaining dimensions also produced favorable results. Attractiveness (1.74), Perspicuity (1.78), and Stimulation (1.38) were classified as Good, indicating that users perceived the application as visually appealing, easy to understand, and engaging throughout the learning process. These findings suggest that the combination of intuitive navigation, clear instructional content, and interactive three-dimensional visualization contributed positively to the overall learning experience.

Meanwhile, Dependability (1.40) and Novelty (0.96) were categorized as Above Average, demonstrating that users considered the application reliable while also recognizing its innovative characteristics. Although the Novelty score was lower than the other dimensions, it still exceeded the average benchmark, indicating that the application offers a satisfactory level of innovation while leaving opportunities for future enhancements, such as richer interactive features, adaptive learning scenarios, or more immersive Augmented Reality interactions.

Dimensi	Mean	Comparison to benchmark	Interpretation
Daya tarik	1.74	Good	10% of results better, 75% of results worse
Kejelasan	1.78	Good	10% of results better, 75% of results worse
Efisiensi	1.89	Excellent	In the range of the 10% best results
Ketersapan	1.40	Above Average	25% of results better, 50% of results worse
Stimulasi	1.38	Good	10% of results better, 75% of results worse
Kebaruan	0.96	Above Average	25% of results better, 50% of results worse

Figure. 7 UEQ Mean Scores.

Overall, the empirical findings demonstrate that the proposed application delivers a positive user experience across all six UEQ dimensions. The consistently favorable benchmark ratings indicate that the application not only functions reliably from a technical perspective but also provides an effective, engaging, and user-friendly learning environment. These results support the feasibility of employing marker-based Augmented Reality as an instructional medium for computer hardware education and

confirm that the developed application satisfies key usability requirements for educational technology.

3.4. Scientific Discussion

The findings of this study demonstrate that the integration of marker-based Augmented Reality with the Multimedia Development Life Cycle (MDLC) methodology provides an effective approach for developing interactive learning media for computer hardware education. The technical evaluation confirmed that the proposed application achieved stable marker recognition under appropriate operating conditions, while the usability evaluation indicated positive user perceptions across all six User Experience Questionnaire (UEQ) dimensions. These results suggest that the application successfully fulfills both its functional requirements and usability objectives.

Compared with previous studies, such as Ramadhan (2021), which primarily emphasized the overall quality of Augmented Reality systems, the present study specifically focuses on the visualization of computer hardware components within a vocational education context [9]. This domain-specific implementation enables students to interact directly with three-dimensional representations of hardware components, thereby supporting a more engaging and intuitive learning process. The adoption of the MDLC framework also provides a structured development process that facilitates systematic multimedia application design and evaluation.

The Efficiency score of 1.89, categorized as Excellent according to the UEQ benchmark, indicates that users were able to operate the application with minimal effort and navigate its features effectively. In addition, the positive scores obtained for Attractiveness, Perspicuity, Dependability, Stimulation, and Novelty demonstrate that the application offers a balanced user experience by combining functional usability with engaging visual interaction. These findings indicate that the developed application is suitable for supporting practical learning activities in vocational education.

From an educational perspective, the proposed application offers a practical alternative for introducing computer hardware concepts in situations where access to physical laboratory equipment is limited. Rather than replacing conventional laboratory practice, the application complements existing instructional methods by providing interactive three-dimensional visualization that enhances students' conceptual understanding and learning engagement. Consequently, the proposed Augmented Reality application has the potential to support technology-enhanced learning environments while reducing the dependence on extensive physical hardware resources.

IV. CONCLUSION

This study developed VirtuComp, a marker-based Augmented Reality (AR) application for computer hardware learning using the Multimedia Development Life Cycle (MDLC) method. The application enables students to explore three-dimensional models of computer hardware components interactively, providing a more engaging learning experience than conventional instructional media.

The experimental evaluation demonstrated that the marker-based tracking mechanism operated reliably under adequate lighting conditions and at an optimal detection distance of 4–12 cm, ensuring stable visualization of virtual objects. Furthermore, the User Experience Questionnaire (UEQ) results indicated positive user perceptions across all six evaluation dimensions, with the Efficiency scale achieving an Excellent benchmark rating, reflecting the application's ease of use and effective interaction design. These findings suggest that the proposed application is a practical and effective learning medium for supporting computer hardware education while reducing the need for extensive physical laboratory equipment. Future research may focus on integrating virtual hardware assembly simulations, incorporating interactive assessment features, and improving marker tracking performance under challenging environmental conditions to further enhance learning effectiveness and user experience.

REFERENCES

- [1] Anakotta, R., Sugiarto, B. A., & Tulenan, V. (2023). Augmented Reality Identifikasi Perangkat Keras Komputer Untuk Anak Kelas 7. *Jurnal Teknik Elektro Dan Komputer*, 12(1), 25–34. Retrieved from <https://ejournal.unsrat.ac.id/v3/index.php/elekdankom/article/view/43850/42456>
- [2] Andreas, N., & Suryantara, I. G. N. (2023). Pengenalan Komponen Komputer Berbasis Augmented Reality Pada Android Dengan Metode Single Marker. *Jurnal Algoritma, Logika Dan Komputasi*, 6(1). <https://doi.org/10.30813/j-alu.v6i1.3610>
- [3] Atmanto, E. A., Pudjoatmodjo, B., & Sularasa, A. (2021). Implementasi Augmented Reality Pada Aplikasi Mutiara Furniture Berbasis Android Menggunakan Metode MDLC. *E-Proceeding of Applied Science*, 7(6), 3412–3419.
- [4] Dalimunthe, H. F., & Simanjuntak, P. (2023). APLIKASI PENGENALAN PERANGKAT KERAS KOMPUTER BERBASIS ANDROID MENGGUNAKAN AUGMENTED REALITY. *Computer and Science Industrial Engineering (COMASIE)*, 9(2). <https://doi.org/10.33884/comasiejournal.v9i2.7624>
- [5] Fadillah, M. A., Yunus, A., & Budianto, A. E. (2022). ANALISIS USER EXPERIENCE PADA AUGMENTED REALITY ORGANOLOGY



- MENGGUNAKAN USER EXPERIENCE QUESTIONNAIRE (UEQ). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 512–518. <https://doi.org/10.36040/jati.v6i2.4888>
- [6] Khoiron, M., Huda, W. S., & Sholiha, H. (2023). Media Pengenalan Hardware Komputer Berbasis Game Menggunakan AR Markerless. *Larisa Teknik Informatika*, 1(01), 24–33.
- [7] Putra, J. A., & MZ, Y. (2023). Implementasi Serta Edukasi Aplikasi Augmented Reality Pengenalan Perangkat Keras Jaringan Menggunakan Metode Marker-Based Tracing Pada SMKN 3 Yogyakarta. *Jurnal ADARMA*, 49–55. Retrieved from <https://e-journal.janabadra.ac.id/index.php/adarma/article/view/2485/1657>
- [8] Putri, A. E., & Hendriyani, Y. (2023). Pengembangan E-Modul Berbasis Augmented Reality Untuk Mata Pelajaran Komputer dan Jaringan Dasar Siswa Kelas X TKJ di SMK Negeri 3 Seluma. *JAVIT: Jurnal Vokasi Informatika*, 3(1), 56–63. <https://doi.org/10.24036/javit.v3i1.70>
- [9] Ramadhan, A. F., Putra, A. D., & Surahman, A. (2021). Aplikasi Pengenalan Perangkat Keras Komputer Berbasis Android Menggunakan Augmented Reality (AR). *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(2), 1–8. Retrieved from <http://jim.teknokrat.ac.id/index.php/JTISI>
- [10] Rasio, H. S., & Perdana, S. R. (2020). Evaluasi User Experience Sistem Informasi Akademik Mahasiswa pada Perguruan Tinggi Menggunakan User Experience Questionnaire. *Jurnal Komputer Terapan*, 6(1), 69–78. Retrieved from <https://jurnal.pcr.ac.id/index.php/jkt/>
- [11] Rinov, M. (2023). Pelatihan Pembelajaran Berbasis Augmented Reality (AR) di Era Teknologi 5.0 pada Materi Instalasi Komputer. *JPM: Jurnal Pengabdian Masyarakat*, 2(2), 209. <https://doi.org/10.52434/jpm.v2i2.2744>
- [12] Saifuddeen, M., Fahim, M., Sahil, M., Sahil, M., & Afeel, A. (2023). Markerless Augmented Reality Implementation Using Computer Vision. *Journal of Advance Research in Mobile Computing*, 5(1).
- [13] Schankin, A., Budde, M., Riedel, T., & Beigl, M. (2022). Psychometric Properties of the User Experience Questionnaire (UEQ). *CHI Conference on Human Factors in Computing Systems*, 1–11. New York, NY, USA: ACM. <https://doi.org/10.1145/3491102.3502098>
- [14] Tasrif, E., Mubai, A., Huda, A., & Rukun, K. (2020). Pemanfaatan media pembelajaran berbasis augmented reality menggunakan aplikasi Ar_Jarkom pada mata kuliah instalasi jaringan komputer. *Jurnal Konseling Dan Pendidikan*, 8(3), 217. <https://doi.org/10.29210/153400>
- [15] Taufik, A., Alamsyah, N., & Saputra, A. (2023). Development of Augmented Reality-Based Mobile App Learning Media for Computer Network Hardware Material. *EduLine: Journal of Education and Learning Innovation*, 3(3), 320–327. <https://doi.org/10.35877/454RI.eduline2004>

