P-ISSN : 2655-8661
E-ISSN : 2655-857X

UNIBI
Universitas Informatika dan Bisnis Indonesia

SisInfo
Vol. 8 No. 1 (2026)
JURNAL SISTEM INFORMASI DAN INFOMATIKA

Optimizing the New Student Admission Process Through the Implementation of a Laravel-Based CRM System at STABA
Annisa Nur Afinni, Tubagus Riko Rivanthio, Rita Komalasari

Analysis of the Usability of the Tinanggea Subdistrict Website Using the Usability Testing Method
Elda Indah Sanda Langi, Isnawaty Isnawaty, Frisilia Febiola, Suci Wulandari, Wa Ode Yurismawati, Wa Rahmiyanti, Awaliyah Fadhilatun Nisa, Nindy Asmawaty, Sheera Annisa, Hasmina Tari Mokoi, Bambang Pramono

Systematic Literature Review: Machine Learning Algorithm Performance Evaluation of Extract-Transform-Load
Muhammad Faisal Ashshidiq, Mohamad Nurkamal Fauzan

Design and Implementation of a Web-Based Marketing Information System for Oil Palm Midrib Products in the Kayu Raja Village Community
Riski Riski, Muh Rasyid Ridha, Usman Usman

Comparison of Chi-Square and Information Gain Feature Selection Methods for Support Vector Machine-Based Sentiment Analysis
Vitta Margaret Sinambela, Herlina Napitupulu, Nurul Gusriani

Activation Function Sensitivity in LSTM-Based Peak Stock Price Forecasting for High-Volatility Financial Time Series
Rizal Rafi Nugraha, Budiman Budiman, Imannudin Akbar

A Comparative Performance Analysis of Classification Algorithms for Hypertension Diagnosis
Imannudin Akbar, Titan Parama Yoga, Acep Hendra, Arnold Ropen Sinaga

A Gamified Learning Model for Business Intelligence in Technopreneurship Education: Case Study at FTI UNIBI
Acep Hendra, Imannudin Akbar, Titan Parama Yoga, Arnold Ropen Sinaga

Image-Based Cat (Felis Catus) Facial Expression Recognition Using YOLOv8n
Atanasius Manurip, Marwondo Marwondo, Venia Restreva Danestiara

Diterbitkan Oleh:
LPPM Universitas Informatika dan Bisnis Indonesia

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 Februari, 2026

DEWAN REDAKSI

Editor in Chief
Dr. Nur Alamsyah, S.T., M.Kom. (Universitas Informatika dan Bisnis Indonesia)

Managing Editor
Budiman, S.T., M.Kom. (Universitas Informatika dan Bisnis Indonesia)

Section Editor
- Titan Parama Yoga, M.Kom. (Universitas Informatika dan Bisnis Indonesia)
- Marwondo, S.T., M.Kom. (Universitas Informatika dan Bisnis Indonesia)
- Chairul Habibi, S.T., M.Kom. (Universitas Informatika dan Bisnis Indonesia)
- R. Yadi Rakhman Alamsyah, S.T., M.Kom. (Universitas Informatika dan Bisnis Indonesia)
- Imannudin Akbar, S.TP., M.Kom. (Universitas Informatika dan Bisnis Indonesia)
- Christian Tonyjanto, S.Kom., M.Pd. (Universitas Dhyana Pura)
- Anggraini Kusumaningrum, S.Kom., M.Cs. (Institut Teknologi Dirgantara Adisujtipto)
- Rio Andriyat Krisdiawan, M.Kom (Universitas Kuningan)
- Achmad Udin Zailani, S.Kom., M.Kom (Universitas Pamulang)

Reviewer
- Reni Nursyanti, M.Kom. (Universitas Informatika dan Bisnis Indonesia)
- Zatin Niqotaini, S.Tr.Kom, M.Kom. (Universitas Pembangunan Nasional Veteran Jakarta)
- Dr. Ati Suci Dian Martha (Universitas Telkom)
- Putri Taqwa Prasetyaningrum, S.T., M.T.MCE (Universitas Mercu Buana Yogyakarta)
- Dr. Erna Hikmawati, S. Kom., M. Kom (Universitas Nasional PASIM)
- Rita Komalasari, S.Si., M.Kom (Politeknik LP3I)
- Deden Pradeka, S.T., M.Kom. (Universitas Pendidikan Indonesia)
- Dr (Cand). Aji Gautama Putrada, S.T., M.T. (Telkom University)
- Dr (Cand). Mohamad Nurkamal Fauzan S.T., M.T. (Universitas Logistik dan Bisnis Internasional)
- Dr (Cand). Nur Ghaniaviyanto Ramadhan, S.Kom., M.Kom. (Telkom University Purwokerto)
- Erik Kurniadi, M.Kom. (Universitas Kuningan)
- Sumardiono, M.Kom. (Universitas Bina Insani)

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 Februari, 2026

DAFTAR ISI

Optimizing the New Student Admission Process Through the Implementation of a Laravel-Based CRM System at STABA 1-10
Annisa Nur Afinni, Tubagus Riko Rivanthio, Rita Komalasari

Analysis of the Usability of the Tinanggea Subdistrict Website Using the Usability Testing Method 11-19
Elda Indah Sanda Langi, Isnawaty Isnawaty, Frisilia Febiola, Suci Wulandari, Wa Ode Yurismawati, Wa Rahmiyanti, Awaliyah Fadhilatun Nisa, Nindy Asmawaty, Sheera Annisa, Hasmina Tari Mokoi, Bambang Pramono

Systematic Literature Review: Machine Learning Algorithm Performance Evaluation of Extract-Transform-Load 20-30
Muhammad Faisal Ashshidiq, Mohamad Nurkamal Fauzan

Design and Implementation of a Web-Based Marketing Information System for Oil Palm Midrib Products in the Kayu Raja Village Community 31-42
Riski Riski, Muh Rasyid Ridha, Usman Usman

Comparison of Chi-Square and Information Gain Feature Selection Methods for Support Vector Machine-Based Sentiment Analysis 43-51
Vitta Margaret Sinambela, Herlina Napitupulu, Nurul Gusriani

Activation Function Sensitivity in LSTM-Based Peak Stock Price Forecasting for High-Volatility Financial Time Series 52-59
Rizal Rafi Nugraha, Budiman Budiman, Imannudin Akbar

A Comparative Performance Analysis of Classification Algorithms for Hypertension Diagnosis 60-68
Imannudin Akbar, Titan Parama Yoga, Acep Hendra, Arnold Ropen Sinaga

A Gamified Learning Model for Business Intelligence in Technopreneurship Education: Case Study at FTI UNIBI 69-76
Acep Hendra, Imannudin Akbar, Titan Parama Yoga, Arnold Ropen Sinaga

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 Februari, 2026

Image-Based Cat (Felis Catus) Facial Expression Recognition Using YOLOv8n 77-85
Atanasius Manurip, Marwondo Marwondo, Venia Restreva Danestiara

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 February 2026

Optimizing the New Student Admission Process Through the Implementation of a Laravel-Based CRM System at STABA

Annisa Nur Afinni, Tubagus Riko Rivanthio, Rita Komalasari
Manajemen Informatika, Politeknik LP3I
Email: annisanurafinni.r23mi@plb.ac.id; tubagusrikorivanthio@plb.ac.id; ritakomalasari@plb.ac.id

Accepted:	Accepted After Revision:	Published:
14 January 2026	9 February 2026	27 February 2026

Abstract

The New Student Admission Process (PMB) is an important agenda in higher education institutions that requires structured data management and communication to run effectively and efficiently. The Bakti Asih College of Analysts (STABA) still manages prospective student data and promotional activities manually using Microsoft Excel and Google Drive. This approach risks data errors and makes communication monitoring less effective. This study aims to develop a Customer Relationship Management (CRM) system to assist in managing prospective student data, monitoring communication, and supporting targeted follow-up actions. The development method used is Waterfall, which includes needs analysis, design, implementation, verification, and maintenance. The CRM system provides features for managing prospects, WhatsApp integration, chat history, automatic status updates, and promotional team performance reports. Black Box testing shows that all functions work as needed. Overall, this system integrates data management, communication, and promotional performance monitoring into a single structured and controlled platform. The system is also designed to improve data accuracy, responsiveness, and efficiency in managing the overall PMB process.

Keywords: Customer Relationship Management, New Student Admissions, WhatsApp, Information System, Web-based application

1 INTRODUCTION

The development of educational institutions in Indonesia is increasing [1]. The high level of competition has resulted in universities having to manage their institutions professionally and focus on service quality [2]. In this competitive environment, the ability of institutions to provide fast, accurate, and targeted services has become one of the determining factors in attracting prospective students. The development of information and communication technology has become an important part of supporting this process, especially in data management and communication [3]. One concept that can be applied to improve service quality is Customer Relationship Management (CRM) [4]. CRM is a system developed to support companies in building close relationships with their customers [5], considering that customers are a very important aspect for the sustainability of an institution [6]. CRM focuses on managing customer data and interactions to improve service quality and customer satisfaction [7]. In higher education, CRM is a digital system that functions to manage data and communication with prospective new students in a centralized platform so that services can be carried out more optimally and purposefully [8].

The Bakti Asih College of Health Analysts (STABA) is a higher education institution that focuses on developing competencies in the field of health analysis. The new student admission process is an important stage in attracting prospective students, which is an annual agenda [9]. However, the management of applicant data and the reporting process at STABA are still done manually using Microsoft Excel and Google Drive. The lack of a system capable of recording and monitoring communication activities between the campus, especially the promotion team, and prospective students has resulted in uncontrolled interactions and follow-ups, leading to the loss of potential applicants. The delivery of information and campus promotion has also not been effective because there is no digital media that can directly connect the two [10].

In addition to technical problems in data management, another challenge faced in the New Student Admission (PMB) process is the suboptimal follow-up strategy for prospective students. Many prospective students who have made initial contact do

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

not receive continuous follow-up, resulting in lost enrollment potential. This is due to the limitations of the system in recording communication history and the absence of systematically measurable indicators of potential status.

The use of Customer Relationship Management (CRM) technology in education is a relevant solution to overcome these problems. CRM not only functions as a data management tool, but also as a strategic means to build long-term relationships between educational institutions and prospective students. With an integrated CRM system, the campus can monitor every interaction in real time, making the decision-making process faster and more accurate.

Based on this background, this study focuses on the application of a Laravel-based CRM system integrated with WhatsApp as the main communication medium. This system is expected to improve the effectiveness of the PMB process at STABA.

2 LITERATURE REVIEW

2.1 New Student Admissions (PMB)

New student admissions (PMB) are a routine agenda carried out by universities. The number of new student admissions reflects the public's perception and interest in the university. Private universities compete with each other to increase the number of new student admissions. Therefore, the number of new student admissions may increase or decrease each year. New student admissions (PMB) is the process of screening prospective students who are accepted into a university. Identifying patterns in PMB can provide useful information to universities in terms of study programs or prospective students who enroll in a study program [11]. In this process, prospective students usually come to the university they are interested in to register by bringing the required documents in accordance with the provisions. With the new student admission CRM system owned by universities, this process has changed. Prospective students no longer need to come in person but can simply do so through the system that has been provided. In other words, the new student admission information system is a very basic and necessary system for any institution or university, and is even the spearhead of marketing to obtain the expected number of students. Because the new student admission information system is an important system, it needs to receive more attention, namely by conducting continuous evaluations to determine the level of success in its implementation [12].

2.2 Customer Relationship Management (CRM)

Customer Relationship Management (CRM) is an approach to customer management through effective communication and relationships. It aims to build relationships and provide sustainable value [13]. Customer relationship management is also a combination of business processes and information technology used to understand various needs [14]. This concept is applied by educational institutions as subjects or service providers. This can attract the interest of the public, especially prospective students. Service is one of the keys to success for educational institutions because it involves a mutually beneficial relationship. With good management, educational institutions can provide more focused, personalized, and sustainable services that support the success of the New Student Admission (PMB) process [15].

2.3 Website

Websites are one of the most widely used sources of information [16]. A website is an application that contains multimedia documents (text, images, sound, animation, video) using the HTTP (hypertext transfer protocol) protocol and accessed using software called a browser [17]. A website is also an address (URL) that functions as a place to store data and information based on specific topics. The web is a hypertext system, consisting of millions of text pages linked by hyperlinks [18].

2.4 Visual Studio Code

Visual Studio Code is a very lightweight yet powerful source code editor that runs from the desktop. It comes with built-in support for JavaScript, scripts, and Node.js and has an array of extensions available for other languages, including C++, C#, Python, and PHP [19]. Visual Studio Code provides many features, including Intellisense, Git Integration, Debugging, and extension features that add to the text editor's capabilities. These features will continue to grow as new versions of Visual Studio Code are released.

2.5 PostgreSQL

PostgreSQL (pronounced Post-Gres-Q-L) or post-gres is one of a number of large databases that offer scalability, flexibility, and high performance. Its use is widespread across various platforms and supported by many programming languages [20]. PostgreSQL is an open source Object Relational Database Management System. PostgreSQL emphasizes extensibility, creativity, and compatibility. It competes with major relational database vendors such as Oracle, MySQL, SQL Server, and others [21].

2.6 Laravel Framework

A framework is a basic conceptual structure used

to solve or handle complex problems. In short, a framework is a container or framework for a website that will be built. By using this framework, the time spent on creating a website is reduced and repairs are made easier. Laravel is an open-source PHP-based framework that uses the model-view-controller concept. Laravel is licensed under the MIT License [22]. The Laravel framework has its own advantages that make it better than other frameworks. The following are the advantages of Laravel: faster performance, more stable data reloading, data security, the use of advanced features such as blade using the HMVC (Hierarchical Model View Controller) concept, the availability of ready-to-use libraries, and the migration management feature for creating table schemas in the database [23].

2.7 Node JS

Node.js is a freely accessible server environment that can operate on various platforms, including Linux, Windows, and Mac. The runtime environment provided by Node.js allows JavaScript to be executed outside of a browser environment. Node.js also has a rich ecosystem with various modules and tools that speed up the development process. With its speed and scalability, Node.js has become a popular choice for developers to build server-side applications, real-time web applications, and high-performance network applications [24]. Node.js complements the role of JavaScript so that it can also function as a server-side programming language, similar to PHP, Ruby, Perl, and so on. Node.js can run on Windows, Mac OS X, and Linux operating systems without any code changes. Node.js has its own HTTP server library, allowing it to run a web server without using web server programs such as Apache or Nginx [25].

3 RESEARCH METHODS

3.1 System Development

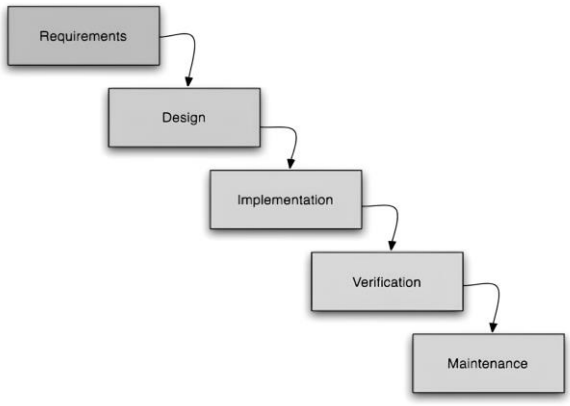


Figure 1. Waterfall Method [26]

The selection of the Waterfall method in the development of this CRM system is based on system requirements that have been clearly defined from the outset. The Waterfall method is a model developed for software development, creating software [27]. The model develops systematically from one stage to another in a waterfall-like mode as shown in Figure 1. This model proposes a systematic and sequential approach to software development, starting from the level of system progress in all analysis, design, coding, testing, and maintenance. This model covers the following activities: requirements analysis, design, implementation, verification, and maintenance. This development model is linear from the initial stage of system development, namely the planning stage, to the final stage of system development, namely the maintenance stage. The next stage will not be carried out until the previous stage has been completed and cannot be returned to or repeated [28].

At the analysis stage, we identify the requirements that must be met in order for the system to record and monitor the status of prospective students as well as all communication and follow-up activities carried out by the promotion team. The system provides user management features with different access rights, supports automatic report generation, and serves as a medium for more effective information delivery and campus promotion.

During the design phase, the structure and flow of the system are designed based on the results of the previously determined requirements analysis. The system design is visualized in the form of a Use Case Diagram to illustrate the interaction between actors and the system, an Activity Diagram to show the process flow and activities in the system, and a database structure design to define the tables, attributes, and relationships between the data used. The design stage aims to provide a comprehensive overview of the system to facilitate the implementation process and minimize errors in the development stage.

At the system implementation stage, the system was implemented with the first step being development using Visual Studio Code and XAMPP. The process began with database configuration and Laravel installation. Next, system features are developed, such as user management, prospective student data management, communication activity recording, potential status management, and report generation. This stage aims to produce a CRM system that can be used according to user needs.

At the verification stage, verification or testing of the system is carried out to ensure that it is in accordance with the design and that it functions smoothly. Testing is carried out by testing each feature of the system, starting from the login process, data management, recording of communication activities, to report generation. The

verification stage aims to find and fix errors or functional inconsistencies so that the system can run well, be stable, and be ready for use by users.

At the maintenance stage, researchers monitor the system's performance to ensure that it continues to run properly. Maintenance includes bug fixes, system performance improvements, and feature updates in accordance with user needs or developments in the new student admission process. This stage aims to maintain the quality of the system so that it remains optimal and can be used for a long period of time.

3.2 Data Collection

Observations were conducted for 3 weeks on the manual follow-up management process, starting from the collection of prospective student data, performance recap calculations, to registration. The purpose of this observation was to identify the weaknesses of the conventional system and define the system to be developed.

Interviews were conducted with the head of the promotion team. The questions asked focused on the difficulties of the conventional system, the features needed in the system, and expectations for the system to be implemented.

Documentation was carried out by collecting various supporting data from previous conventional management.

4 RESULTS AND DISCUSSION

The results of the Waterfall model make it easier for researchers to compile development stages in a systematic and structured manner. By following this process, researchers can ensure that system development runs smoothly and in a controlled manner. The following are the research stages that were applied:

4.1 System Design

a. Use Case Diagram

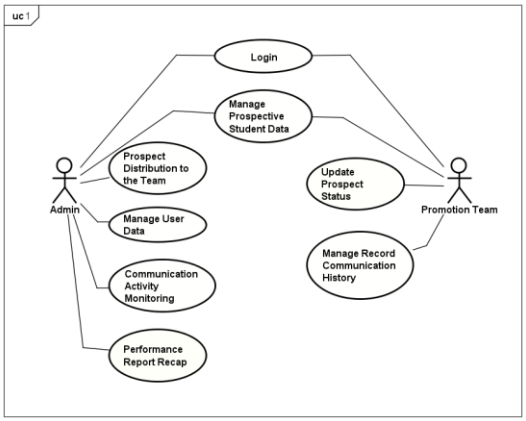


Figure 2. Use Case Diagram

In Figure 2, the Use Case Diagram involves two actors, namely Admin and Promotion Team. The Admin acts as the system manager who is responsible for managing user accounts, managing prospective student data, assigning prospects to the Promotion Team, monitoring communication activities, and performance reports recap. Meanwhile, the Promotion Team is responsible for handling prospective students by updating prospect status and recording communication history based on the interactions conducted. All activities performed by the Promotion Team are stored in the system database and can be monitored by the Admin to support performance evaluation.

b. Activity Diagram

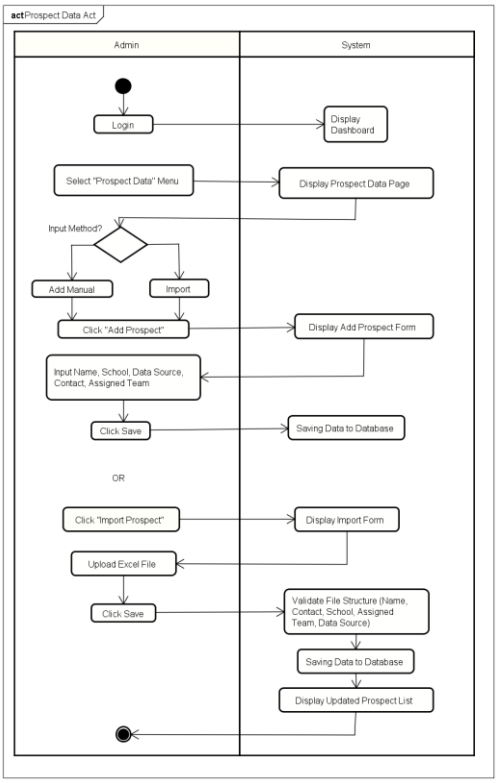


Figure 3. Prospective Student Data Activity Diagram

The Prospective Student Data Activity Diagram in Figure 3 illustrates the workflow of prospective student data management carried out by administrators through the CRM system, starting from the login process to the system displaying the dashboard page. The administrator then selects the prospective student data menu so that the system displays the prospective student data page. Administrators can add prospective student data

manually by clicking the "Add Prospective Student" button and filling in detailed information such as the prospective student's name, school of origin, data source, contact details, and assigned team, then saving the data to the database. In addition, administrators can also enter data in bulk through the data import feature by selecting an Excel file. The system will then validate the file structure to ensure that required data such as name, contact, school, assigned team, and data source are available before storing the data into the database. After that, the system will save all prospective student data to the database, display the updated prospective student data list, and the prospective student data management process is complete.

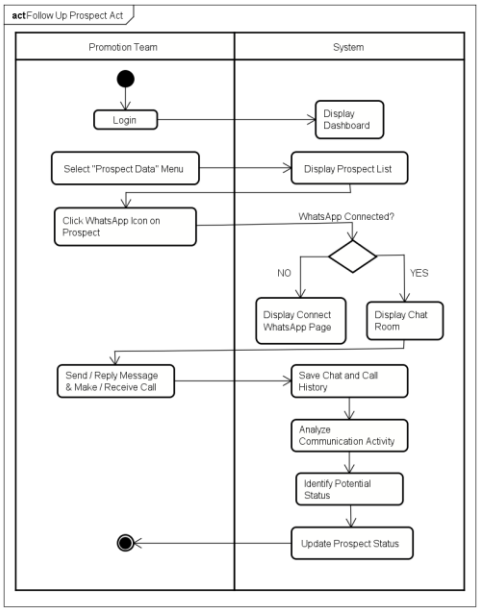


Figure 4. Follow-Up Activity Diagram

Figure 4, the Follow-up Activity Diagram, illustrates the follow-up process for prospective students involving users (the promotion team) and the CRM system. It begins when the user logs in and the system displays the dashboard page. The user then selects the prospective student data menu, and the system displays the prospective student data page. Next, the user connects to WhatsApp to start communicating with the prospective student. Once the connection is successful, the system displays a chat room as a medium for interaction. Based on the conversation content, the system identifies the prospective student's status as a result of the follow-up process, and once the potential status is recorded, the follow-up process is complete.

c. Database Structure

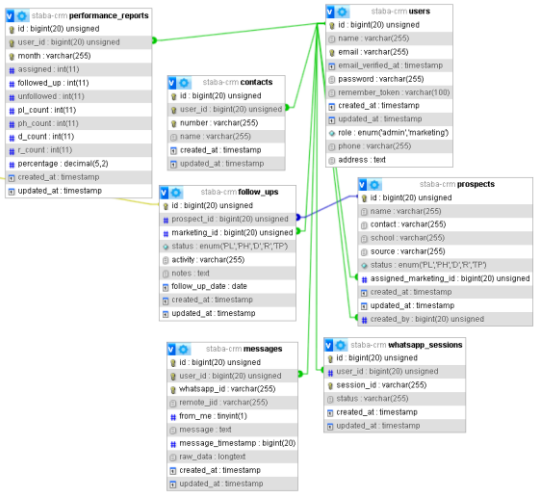


Figure 5. Database Structure

The database structure in Figure 5 of this CRM system consists of seven main integrated tables to support the new student registration process, namely the users, prospects, contacts, follow_ups, messages, whatsapp_sessions, and performance_reports tables. The users table serves to store system user data with roles as admin or promotion team and is the main table related to most of the other tables. The prospects table is used to store prospective student data along with their potential status and their relationship with assigned users and data recorders. The contacts table stores WhatsApp contact information linked to users, while the WhatsApp_sessions table stores WhatsApp connection session data for each user. The messages table is used to record all WhatsApp conversation activities between users and prospective students as part of the communication process. In addition, the follow_ups table is used to record follow-up activities carried out by the promotion team for prospective students, along with their status and follow-up schedule. Finally, the performance_reports table is used to store summary data on the marketing team's performance based on follow-up activities and communication results, all of which are connected through foreign key relationships to ensure data integrity and consistency in the CRM system.

4.2 System Implementation

The system implementation phase covers the results of applying the design based on the specified requirements. The implementation results show that the

system is capable of supporting the new student admission management process in a structured and integrated manner. All prospective student data is stored in a centralized database, minimizing the risk of data loss and duplication. The potential status monitoring feature provides a clear picture of each prospective student's position in the admission stages, allowing the promotion team to determine more appropriate follow-up strategies.

The integration of WhatsApp as the main communication medium makes it easier for the promotion team to interact directly with prospective students without having to switch platforms. All communication activities are automatically recorded, making it easier for administrators to monitor and evaluate the performance of the promotion team objectively. With performance recap reports that can be filtered based on specific periods, campus management can obtain accurate information as a basis for strategic decision-making in the new student admission process.

Based on the implementation results, the following is the main interface display of the CRM system used by administrators and users (promotional teams).

The Login page is accessed by administrators and users (promotional teams). Administrators create accounts for users so they can access the system. The system will verify the email and password entered and then direct users to the dashboard page according to their role, whether administrator or user (promotional team).

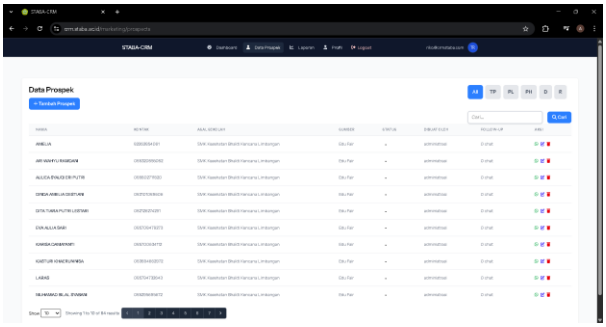


Figure 6. Prospect Data Page (User)

The Prospect Data page in Figure 6 is used by users to view a list of prospective students assigned by the admin. On this page, users can view basic information about prospective students, such as their name, school of origin, and potential status. This page serves as a basis for following up and communicating with prospective students through the CRM system.

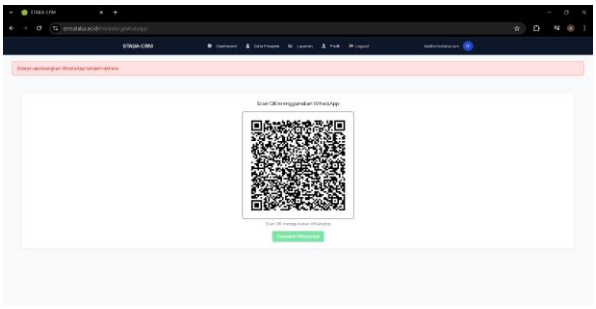


Figure 7. WhatsApp QR Scan Page (User)

The WhatsApp QR Scan page in Figure 7 is used by users to connect the CRM system with the WhatsApp application before the communication process is carried out. This page is accessed through the Prospect Data page, when the user clicks the WhatsApp icon on the prospective student's data. Once this page is open, the system will display a QR Code that must be scanned by the user using the WhatsApp application. If the scanning process is successful, the system will display an active WhatsApp connection status so that the user can continue the communication process with prospective students through the chat room page on the CRM system.

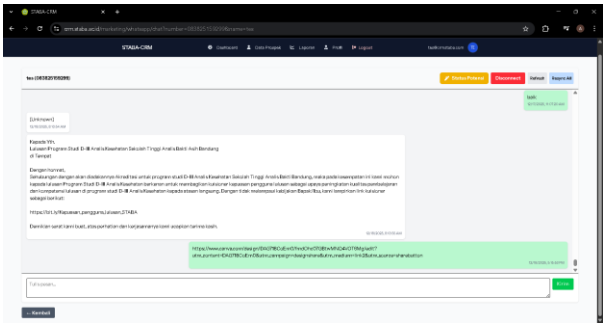


Figure 8. Chat Room Page (User)

The Chat Room page in Figure 8 is used by users to communicate directly with prospective students via WhatsApp. All incoming and outgoing messages are automatically recorded in the CRM system, can be monitored by the admin for monitoring purposes, and are used by the system to automatically update the status of prospective students based on the content of the conversation.

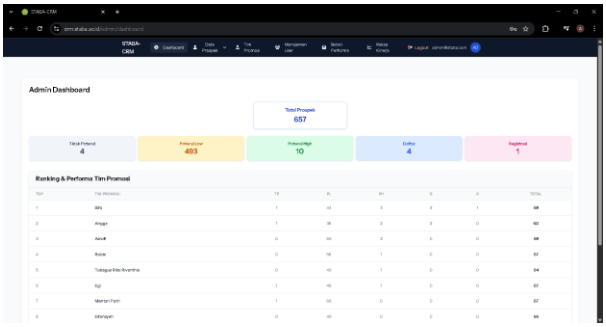


Figure 9. Dashboard Page (Admin)

The Admin Dashboard page in Figure 9 serves as the CRM system monitoring center. This page displays a summary of prospective student data based on potential status, as well as a table ranking the performance of the promotion team, sorted by the number of prospective students who have successfully registered. This dashboard helps administrators monitor and evaluate the performance of the promotion team and the overall new student admission process.

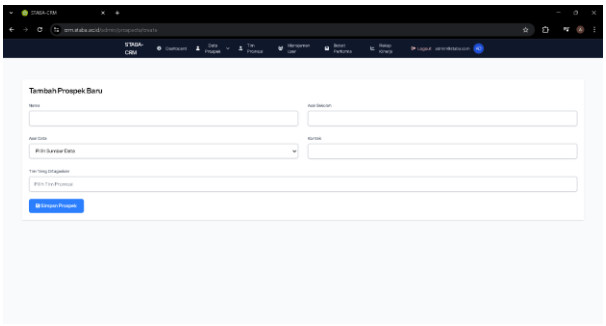


Figure 10. Add Prospective Student Data (Admin)

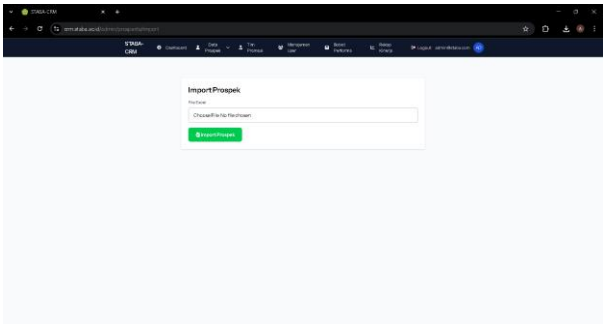


Figure 11. Import Prospect Data (Admin)

The Add and Import Prospect Data pages in Figures 10 and 11 are used by administrators to manage prospective new student data. On the add prospect data page, administrators can manually enter prospective student data one by one by filling in information such as name, school of origin, contact details, data source, and promotional team assignment. Meanwhile, the import prospective student data page is used to add prospective student data in bulk via an Excel file. After the saving process is complete, all prospective student data will be automatically stored in the CRM system database.

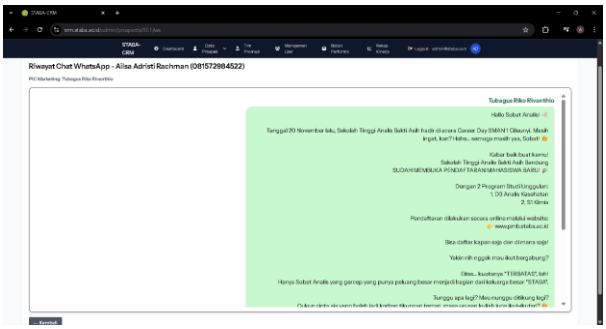


Figure 12. WhatsApp Message History Page (Admin)

The WhatsApp Message History page in Figure 12 is used to display the communication history between users and prospective students. This page is accessed through the Prospect Data page by clicking the WhatsApp icon on the prospective student's data. All incoming and outgoing messages that occur through the chat room will be stored and displayed on this page, so that it can be used by the admin as material for monitoring and evaluating the communication activities of the promotion team.

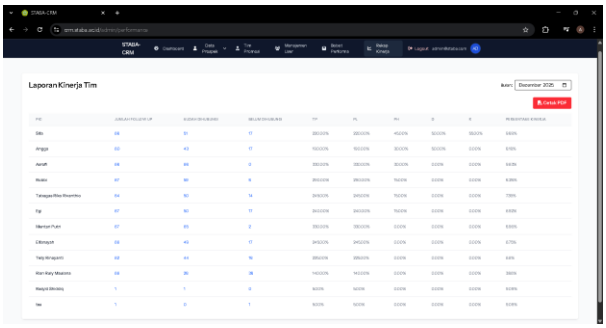


Figure 13. Performance Recap Report Page (Admin)

The Performance Recap Report page in Figure 13 is used by administrators to view and evaluate the performance of the promotion team based on their activities. On this page, administrators can filter reports by month, so that performance data is displayed according to the selected month. In addition, the system also provides a report printing feature that makes it easy for administrators to save or document the promotion team's performance recap results for evaluation and reporting purposes.

4.3 Testing

Testing was conducted using the black box method, as shown in the following table:

Table 1. System Testing

Feature Tested	Scenario	Result	Description
Login	Entering registered email and password	The system displays the dashboard according to the role	Succesful
Admin Dashboard	Accessing the admin dashboard page	The system displays a summary of the number of prospects based on potential status and promotion team performance ranking.	Successful
Prospect Data	Displaying prospect data	The system displays a list of prospective student data.	Successful
Add Prospect Data	Clicking the add prospect button and filling out the form	The system displays a form to add prospects and save data.	Successful
Import Prospect Data (Admin)	Clicking the import prospect button and uploading the file	The system displays a form to import prospects and save data.	Successful
Edit Prospect Data	Clicking the pencil icon and editing the data	The system successfully updates data.	Successful
Delete Prospect Data	Clicking the trash can icon and deleting the data	The system successfully deletes data.	Successful
WhatsApp Message History	Clicking the WhatsApp icon on the prospect data page	The system displays the user's WhatsApp message history with prospective students.	Successful
Team Performance Report (Admin)	Selecting the month filter and printing the report	The system displays data according to filters and prints promotional team performance reports.	Successful
Scan WhatsApp QR Code (User)	Scanning the WhatsApp QR Code	The system displays the WhatsApp connection status.	Successful
Chat Room (User)	Sending and receiving WhatsApp messages	Messages are displayed and stored in the CRM system.	Successful
Potential Status (User)	Conversing with prospective students	Potential status is automatically updated according to the content of the conversation.	Successful

5 CONCLUSION

Based on the results of the design, implementation, and testing of the Laravel-based Customer Relationship Management (CRM) system in the New Student Admission process at the Bakti Asih College of Analysts (STABA), it can be concluded that the system that was built is capable of improving the effectiveness and efficiency of prospective student data management. This system provides centralized prospect data recording facilities, potential status monitoring, and documentation of all communication activities and follow-ups carried out by the promotion team.

The application of the Waterfall method in system development provides a structured and systematic workflow, so that each stage of development can be carried out properly and in a controlled manner. The integration of WhatsApp as the main communication medium has proven to help improve the quality of interaction between the promotion team and prospective students, as well as facilitate the monitoring process by the admin.

6 SUGGESTION AND RECOMMENDATION

With this CRM system, the reporting process becomes faster and more accurate, the risk of data recording errors can be minimized, and the performance of the promotion team can be evaluated objectively based on available data. For further development, this system will be synchronized with the PMB website, so that data management and the new student admission process become more integrated, responsive, and sustainable.

REFERENCES

[1] A. L. Muhamad Muslihudin, "PERANCANGAN SISTEM APLIKASI PENERIMAAN MAHASISWA BARU DI STMIK PRINGSEWU MENGGUNAKAN PHP DAN MYSQL," *Jurnal TAM (Technology Acceptance Model) Vo,* pp. 32-39, 2014.

[2] Y. T. K. C. Maria Alfa Raniadita, "Analisis Efektivitas Marketing CommunicationProgram PMBpada Kampus yang Berubah Bentuk di wilayah Kab. Bandung Bar," *Jurnal Manajemen Bisnis dan Kesehatan (JMBK),* pp. 64-76, 2025.

[3] Z. I. Enggelbertus Gobai, "SISTEM INFORMASIPENERIMAAN MAHASISWA BARU SEKOLAH TINGGI ILMU HUKUM (STIH) MONOKWARI," *Jurnal Ilmiah Sistem Informasi dan Teknik Informatika "JISTI,* pp. 17-26, 2020.

[4] D. Hamidin, "MODEL CUSTOMER RELATIONSHIP MANAGEMENT (CRM)," *Seminar Nasional Aplikasi Teknologi Informasi,* pp. 31-34, 2020.

[5] G. T. M. Iqbal Aditya Pangestu, "SISTEM INFORMASI REKOMENDASI PRODUK

DENGAN PENDEKATAN CUSTOMER RELATIONSHIP MANAGEMENT DI CV. XYZ," *KOMPUTA : Jurnal Ilmiah Komputer dan Informatika,* pp. 78-86, 2021.

[6] D. H. Emma Rosinta, "Implementasi Customer Relationship Management (CRM) Pada AplikasiPenjualan Berbasis Web PT. Buana Telekomind," *JURNAL TECHNOLOGY INFORMATICS & COMPUTER SYSTEM,* pp. 8-14, 2018.

[7] L. K. Fauziah, "Implementasi CRM Customer Touching Application Memanfaatkan Model Cross Selling Dalam Promosi Produk Lain," *JTSI,* pp. 120-131, 2020.

[8] Y. S. Kartika Trisya Aprilla, "Rancangan Bangun CRM Operasional Penerimaan," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK),* pp. 1073-1085, 2024.

[9] L. S. R. A. S. H. M. Muhammad Nugraha, "RANCANG BANGUN SISTEM INFORMASI PENERIMAAN MAHASISWA BARU BERBASIS WEB DENGAN MENGGUNAKAN FRAMEWORK LARAVEL," *JITET (Jurnal Informatikadan Teknik Elektro Terapan),* pp. 1162-1169, 2024.

[10] D. H. P. N. Jajang Abdurrahman, "Implementasi Customer Relationship Management (CRM) Pada Penerimaan Siswa Baru (Studi Kasus di Pondok Pesantren Darussalam Kabupaten Garut)," *JURNAL NUANSA INFORMATIKA,* pp. 29-36, 2018.

[11] S. A. A. Z. Y. A. Aradea, "PENERAPAN DECISION TREE UNTUK PENENTUAN POLA DATA," *Penelitian Sitrotika,* pp. 1-7, 2011.

[12] E. U. N. Donny Yulianto, "Evaluasi Tingkat Kesuksesan Sistem Informasi Penerimaan Mahasiswa Baru Menggunakan Metode Hot Fit," *Journal of Computer, Information System, & Technology Management,* pp. 11-20, 2021.

[13] K. K. N. Kartika Imasari, "PENGARUH CUSTOMER RELATIONSHIP MANAGEMENT TERHADAP LOYALITAS PELANGGAN PADA PT BCA Tbk," *Fokus Ekonomi (FE),* pp. 183 - 192, 2011.

[14] H. I. Filania Rahayu, "PERANCANGAN SISTEM ELECTRONIC CUSTOMER RELATIONSHIP MANAGEMENT (E-CRM) GUNA MEMBANTU MENINGKATKAN PELAYANAN DAN LOYALITAS PELANGGAN STUDI KASUS: LEMBAGA PENDIDIKAN KUMON," *Jurnal IDEALIS,* pp. 205-212, 2019.

[15] R. H. Saeful Anam, "BRANDING IMAGE DAN CUSTOMER RELATIONSHIP MANAGEMENT (CRM) DALAM PENGUATAN MUTU PENDIDIKAN ISLAM DI SMA KHADIJAH SURABAYA," *Jurnal Akademik Ekonomi Dan ManajemeN,* pp. 168-181, 2025.

[16] S. M. Ekkal Prasetyo, "Rancang Bangun Sistem Informasi Sekolah Tinggi Ilmu Ekonomi Rahmanyah Kabupaten Musi Banyuasin Berbasis Website," *Jurnal Informanika,* pp. 19-30, 2015.

[17] M. Suhartanto, "Pembuatan Website Sekolah Menengah Pertama Negeri 3 Delanggu Dengan Menggunakan Php Dan MySQL," *Journal Speed – Sentra Penelitian Engineering dan Edukasi,* pp. 1-8, 2012.

[18] S. H. Ridwan Sanjaya, "RANCANG BANGUN WEBSITE PROFIL HOTEL AGUNG PRABUMULIH MENGGUNAKAN FRAMEWORK BOOTSTRAP," *Jurnal Teknologi dan Informasi (JATI) ,* pp. 57-64, 2017.

[19] S. Hartati, "PERANCANGAN SISTEM INFORMASI INVENTARIS BARANGPADA KANTOR NOTARIS DAN PPAT R.A LIA KHOLILA, S.H MENGGUNAKAN VISUAL STUDIO CODE," *Jurnal Siskomti,* pp. 37-48, 2020.

[20] S. J. Anik Sri Wahyuningsih, "PENERAPAN KNOWLEDGE MANAGEMENT SISTEM LAYANAN PERBAIKAN DEVICE DISTRIBUTION CENTER (DC) SUPPORT PT.INDOMARCO PRISMATAMA CABANG BEKASI," *Jurnal Informatika SIMANTIK,* pp. 47-51, 2020.

[21] T. S.Waruwu, "IMPLEMENTASI POSTGRESQL SEBAGAI SISTEM MANAJEMEN BASIS DATA PADA PENDAFTARAN MAHASISWA BARU BERBASIS WEB," *JURNAL MAHAJANA INFORMASI ,* pp. 57-61, 2019.

[22] A. I. N. Delia Mediana, "RANCANG BANGUN APLIKASI HELPDESK (A-DESK) BERBASIS WEB MENGGUNAKAN," *JURNAL MANAJEMEN INFORMATIKA,* pp. 75-77, 2018.

[23] A. P. I Gede Handika, "Pemanfaatan Framework Laravel Dalam Pembangunan Aplikasi E-Travel Berbasis Website," *Konferensi Nasional Sistem Informasi,* pp. 1329-1334, 2018.

[24] M. R. Y. Maria Gracela Sandra De Beny, "PERANCANGAN WEBSITELAPORAN DATA UMKM PROVINSI NUSA TENGGARA TIMUR MENGGUNAKAN JAVASCRIPTDAN NODE.JS," *JURNAL TEKNOLOGI INFORMAS,* pp. 43-50, 2024.

[25] H. F. R. Irfan Kurniawan, "REST API Menggunakan NodeJS pada Aplikasi TransaksiJasa Elektronik Berbasis Android," *JITSI,* pp. 127-132, 2020.

[26] A. S. B. A. S. R. D. R. Teguh Rijanandi, "Implementation of Encrypt National ID Card in Sinovi Application Use Waterfall Methodology," *Journal of Artificial Intelligence and Digital Business(RIGGS,* pp. 11-18, 2023.

[27] A. V. N. H. Dea Safryda Putri, "IMPLEMENTASI METODE WATERFALL DALAM PERANCANGAN SISTEM INFORMASI LAYANAN RW 41 KAMPUNG MARKAN BESI," *INFORMATION SYSTEM DEVELOPMENT,* pp. 7-14, 2023.

[28] H. P. Zendrato, "Perancangan Sistem Informasi Logistik Berbasis Web Pada PT.Unitama Huting Mandiri Menggunakan Metode Waterfall," *Riset dan E-Jurnal Manajemen Informatika Komputer,* pp. 19-35, 2020.

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 February 2026

Analysis of the Usability of the Tinanggea Subdistrict Website Using the Usability Testing Method

Elda Indah Sanda Langi[1)], Isnawaty[1)], Frisilia Febiola[1)], Suci Wulandari[1)], Wa Ode Yurismawati[1)], Wa Rahmiyanti[1)], Awaliyah Fadhilatun Nisa[1)], Nindy Asmawaty[1)], Sheera Annisa[1)], Hasmina Tari Mokoi[2)], Bambang Pramono[1)]

[1)] Department of Computer Science, Faculty of Engineering, Halu Oleo University
[2)] Department of Electrical Engineering, Faculty of Engineering, Halu Oleo University
Email: eldaindahsl1203@gmail.com; isnawaty@uho.ac.id, frisilia.febiolaa@gmail.com; pramitasuci632@gmail.com; waodeyuris@gmial.com; warahmiyanti@gmail.com; awaliyahfadhilatunis@gmail.com; nindyasmawaty0@gmail.com; sheeraannisa@gmail.com, hasmina.mokui@uho.ac.id, bambang.pramono@uho.ac.id

Accepted:	Accepted After Revision:	Published:
18 January 2026	10 February 2026	27 February 2026

Abstract

The Tinanggea Subdistrict website provides information and public services for the community at the subdistrict level and surrounding areas. However, the existence of a website must be supported by a good level of usability so that the information provided to users is conveyed properly. This study aimed to analyze the usability level of the Tinanggea Subdistrict website using a usability testing method. The research method used was descriptive with a task-based usability evaluation approach. Testing was conducted by involving eight people who were assigned as users of the website. The usability level was measured using the System Usability Scale (SUS) in the form of a questionnaire consisting of 10 questions. The SUS scores of each user were calculated and averaged to obtain the overall website usability score. The test results show that the Tinanggea Subdistrict website scored 51.1, indicating that it is usable but still not optimal. This study also shows that the website has implemented usability principles, especially in terms of ease of navigation and error prevention. However, the website still needs various improvements, particularly in terms of interaction efficiency, interface design consistency, and interactive service support, so that it can function optimally as a public service medium.

Keywords: Website, Tinanggea District, Usability Testing, ISO 9241-11, System Usability Scale (SUS)

1 INTRODUCTION

Information and communication technology is rapidly evolving. This ongoing progress has brought new innovations to all areas of human life. One such innovation is the e-government [1]. This aims to increase government transparency, effectiveness, and public access via official websites. Local government websites, including those for subdistricts and villages, are the main channels for public information, services, and communication with communities. However, a website's effectiveness depends heavily on its usability.

However, the existence of websites alone does not guarantee that the information and services provided can be optimally utilized by the community. Many government websites still face usability issues, such as confusing navigation, unclear information structure, and user-unfriendly interfaces. These problems can hinder the public from accessing important information and reduce their trust in government's digital services [2].

Usability is a measure that assesses how easy it is to use a website interface [3]. Usability refers to the extent to which a system can be used effectively, efficiently, and satisfactorily by users in a specific context. A website with good usability is expected to make it easier for users to find information, complete tasks, and understand the available functions without experiencing significant confusion [4].

Evaluations of the usability of local government websites have been conducted extensively in the context of research to measure users' real experiences with public digital services. Empirical research shows that aspects such as learnability, efficiency, memorability, error, and satisfaction have a significant effect on ease of use and public satisfaction; therefore, they need to be evaluated on government websites that are directly used by the public. Thus, usability serves as a technical

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

indicator and a measure of the government's success in providing digital services that are truly useful and responsive to the needs of the community [5].

In the context of local government, usability evaluation is increasingly important because website users come from diverse backgrounds, including age, education level, and digital literacy levels. Previous studies have shown that sub-district and village government websites often do not fully consider ease of use; therefore, a systematic evaluation is needed to identify existing usability issues [6]. Usability evaluation is expected to reveal the obstacles faced by users when accessing websites and provide a basis for recommendations for future system improvement.

Based on this background, this study aims to analyze the usability level of the Tinanggea Subdistrict Website using the usability testing method. Through this approach, the study is expected to provide a realistic picture of user experiences when interacting with websites and serve as an evaluative basis for efforts to improve the quality of web-based information services.

2 LITERATURE REVIEW

2.1 Usability in Government Websites

Research on the usability of government websites has been conducted extensively in Indonesia, particularly at the local government and public agency levels. Previous studies have found that many government websites still face problems with navigation structure, menu clarity, and display consistency [1]. Other studies have shown that good content quality does not necessarily guarantee that a website is easy to use if usability aspects are not comprehensively considered [7].

Based on these findings, it can be concluded that usability issues on government websites are common and can occur at various levels of agencies. Websites that are already available and actively used still have the potential to cause difficulties for users if they are not designed and evaluated with usability in mind. Therefore, usability is a fundamental aspect of public service website development.

2.2 Usability Testing Methods in Website Evaluation

Several studies have applied usability testing methods to evaluate government websites. A usability analysis using sub-district-level e-government websites found that users still experienced confusion in navigation and information search [6]. Furthermore, the usability testing approach has been shown to be effective in describing the real experience of users compared to evaluation methods based on developer assumptions [5].

In addition, usability evaluation using the System Usability Scale (SUS) approach is widely used to assess the usability of other government websites, such as in studies evaluating public service websites of government agencies. The results of the study show that SUS scores often reflect the need for improvements in terms of ease of use, navigation, and user satisfaction, thus providing a concrete image of the user experience when interacting with digital government services [8]. These findings confirm that the SUS-based usability testing method can complement traditional usability testing evaluations, especially in measuring quantitative user responses relevant to the development of sub-district government websites.

These findings show that usability testing functions not only as a measuring tool but also as a means of directly understanding user behavior. By observing how users complete specific tasks, researchers can identify usability issues that are often undetectable using other evaluation methods such as surveys. This makes usability testing relevant for applications on sub-district-level government websites with diverse user characteristics.

2.3 Usability of Local Government Websites

Other studies have highlighted the importance of usability on local government websites, including those for villages and subdistricts. A usability analysis of the Musi Village Website shows that information structure and menu clarity greatly affect users' success in accessing digital services [4]. Furthermore, usability evaluations of public service websites can provide improvement recommendations oriented towards the needs of end users [4], [9].

In addition to research highlighting aspects of information structure and user perception, several other empirical studies have shown that usability evaluations of local government websites provide new insights into the real experiences of citizens when they interact with digital services. For example, usability evaluations of e-government websites in the sub-district office environment show that the level of community involvement in utilizing online services is still low due to obstacles such as a lack of digital literacy and the community's preference for offline services [10]. This finding also emphasizes that usability evaluations must consider the diverse conditions of users at the local level so that the evaluation results can form the basis for inclusive and contextual recommendations for improving the development of sub-district and village government websites in the future.

Based on this study, it can be concluded that the usability evaluation of sub-district government websites still has room for improvement. Many studies focus on the context of cities and regencies or emphasize user perceptions, whereas task-based evaluations at the sub-district level are relatively limited. Therefore, this study

attempts to fill this gap by analyzing the usability of the Tinanggea Sub-district Website using the usability testing method to obtain a real picture of the user experience and relevant recommendations for improvement.

2.4 ISO 9241-11

ISO 9241-11 defines usability as the extent to which a system can be used by specific users to achieve specific goals in a specified context of use [11]. It considers effectiveness, efficiency, and satisfaction as key aspects of usability. Based on this standard, usability consists of three main components: effectiveness, efficiency, and user satisfaction.

Effectiveness refers to the error rate of a system; a good system is one that, when used, results in few errors made by users [12].

Efficiency refers to how quickly and accurately users can complete tasks using the system after learning its design. In this context, efficiency measures the resources expended in relation to the accuracy and completeness with which users achieve goals [12], [13].

Satisfaction is the level of satisfaction users feel when using the system provided to them for testing. Usually, this section will also present the possibility of a user using the system in the future.

These three components serve as a conceptual framework for interpreting the results of usability testing on the Tinanggea Subdistrict Website, enabling systematic analysis of the research results to be carried out systematically and in accordance with international usability standards.

2.5 System Usability Scale (SUS)

The System Usability Scale (SUS) is a usability evaluation instrument developed by John Brooke in 1996 as a simple questionnaire containing ten statements with a five-point Likert scale to measure users' subjective perceptions of the overall usefulness of the system [14]. The SUS is designed as a quick and practical evaluation tool, so it can be used efficiently in conditions of limited time and resources, and is neutral to the type of technology being evaluated. In its development, Brooke selected ten items with high correlations from a number of initial statements to produce alternating positive and negative response patterns to minimize response bias. The SUS primarily reflects the satisfaction aspect of usability, as defined in ISO 9241-11, and has been widely used in various interactive system evaluation contexts. Further research shows that although the SUS can provide a representative picture of usability with a relatively small number of respondents, increasing the sample size can improve the reliability of the findings and broaden the scope of usability issues identified [15].

3 RESEARCH METHODS

3.1 Research Type and Approach

This study used a descriptive method with a usability testing approach. The descriptive method was chosen because this research aims to describe and analyze the condition of the website as it is, without making significant developments or modifications to the system. The focus of the research is on how users interact with the website and the extent to which the website is easy to use in supporting users' information needs.

The usability testing approach was used because this method is one of the main methods in the field of human–computer interaction to evaluate interactive systems based on direct user experience. Through usability testing, researchers can observe user behavior, difficulties encountered, and interaction barriers that arise when users interact with a website. In addition to direct observation, this study also used a quantitative instrument in the form of the System Usability Scale (SUS) to obtain a numerical picture of the website's usability level and reinforce the evaluation results objectively.

3.2 Research Object

The research object in this study is the previously developed Tinanggea District website. The website functions as a medium for providing information containing various content, such as regional profiles, service information, news, and other supporting data for the project.

In the context of this study, the website is treated as a fully developed system that is ready for use by users. Therefore, this study does not discuss the design process or technical implementation of the website but focuses its analysis on the usability and ease of user interaction with the website interface based on the principles of human–computer interaction.

3.3 Research Subject

The research subjects were users of the website of the Tinanggea District, selected using purposive sampling. Eight respondents were involved in the study.

The number of respondents in this study refers to usability testing literature, which states that there is no absolute number of respondents; rather, it is adjusted to the objectives and context of the study. Previous studies have shown that using more than five respondents can increase the stability and coverage of usability findings, especially in systems used by individuals with diverse characteristics [16]. This is in line with a recent study which shows that in formative usability evaluations, a small number of respondents (6 to 9 people) is still effective and reliable for identifying the most critical usability issues on digital platforms [17].

In addition, another study explains that a range of five to ten respondents is commonly used in usability testing to obtain representative findings without producing excessively repetitive data [15]. Therefore, this study involved eight respondents, which was considered appropriate to achieve the objectives of descriptive evaluative usability evaluation.

The characteristics of the respondents in this study included the following:

1. Users who have accessed or are likely to access the Tinanggea District website.
2. Aged between 18 and 40 years old.
3. Basic skills in using digital devices and the internet.

3.4 Data Sources

The data sources used in this study consist of two types: primary and secondary data. Primary data were obtained from direct observation of user interactions when using the Tinanggea Subdistrict website. This data includes how users navigate the website, usage errors that occur, and difficulties experienced by users while interacting with the system.

Secondary data were obtained from the Tinanggea District website database. This database is used as a source of information displayed on the website. In this study, secondary data were not analyzed statistically but were used to create a realistic context for website usage so that usability evaluation could be carried out optimally.

3.5 Data Collection Techniques

The techniques used in data collection are divided into two categories, as follows.

3.5.1 Task-Based Usability Testing

At this stage, respondents will be asked to complete several scenario-based tasks. This is done to identify and record the general activities of website users. During this process, researchers observe the website, including the navigation flow, usage errors, and confusion experienced by users while using it.

3.5.2 System Usability Scale (SUS) Questionnaire

After completing all the given scenario tasks, respondents were asked to fill out a System Usability Scale (SUS) questionnaire consisting of 10 questions on a 1-5 Likert scale. This was done to obtain quantitative data on the website's usability to support the qualitative observation results. Details of the questions included in the questionnaire can be seen in Table 1 below.

Table 1. SUS Questionnaire

No	Statement
1	I think I will use this website often.
2	I find this website too complicated to use.
3	I find this website easy to use.
4	I need help from someone else or a technician to be able to use this website.
5	I feel that the various functions on this website are well integrated.
6	I feel that there are many inconsistencies on this website.
7	I feel that most people will quickly learn how to use this website.
8	I find this website very complicated to use.
9	I feel confident when using this website.
10	I need to learn a lot before I can use this website.

3.6 Task Scenario

A task scenario is a stage in the design of tasks related to a product or research being conducted [18]. The task scenarios presented are based on the general needs of users when accessing information on the Tinanggea Subdistrict website. This aims to make it easier for researchers to evaluate the ease of navigation on the website, the clarity of the information structure, and the ease of searching for information. The task scenarios in this study included four tasks that can be presented through Table 2 below.

Table 2. Task Scenario

Code	Task Scenario	Goal
T1	Go to the website and find detailed information about the profile or history of Tinanggea District.	Respondents successfully found and understood regional profile information.
T2	Search for information about the geographical location, boundaries, or map of Tinanggea District.	Respondents can accurately identify geographical information about the region.
T3	Find information about the potential of the region (such as natural resources, tourism, or MSMEs) in Tinanggea District.	Respondents can identify the potential strengths of the region.
T4	Find the official contact information, office address, or complete identity of the Tinanggea Subdistrict administrator.	Respondents successfully found contact details or agency identities for communication purposes

3.7 Data Analysis Techniques

The data analysis techniques used in this study were descriptive, qualitative, and quantitative. Quantitative data were obtained from the SUS questionnaire, specifically the final usability score of each respondent, which was then averaged to obtain an overview of the usability of the website. The quantitative analysis results supported the qualitative findings.

According to previous study, the usability test results are closely related to the concept of usability in ISO 9241-11, particularly the three concepts of effectiveness, efficiency, and satisfaction in the use of public service websites [19].

3.8 Research Stages

The research stages conducted in this study included the following:

1. Identification of Usability Problems on the Website
   This stage aimed to identify initial problems related to the usability of the Tinanggea District website. The identification was carried out through an initial review of the appearance, navigation structure, and presentation of the information on the website. The results of this stage were used as a basis for determining the focus of the usability evaluation to be conducted in the next stage.

2. Task Scenario Development
   In this stage, the researcher developed task scenarios that represented the general activities of users when accessing the website. The task scenarios were designed based on the needs of users in obtaining public information so that they could describe the actual conditions of website usage. The development of task scenarios aims to ensure that the usability evaluation is conducted in a focused manner and is relevant to the context of website usage.

3. Task-Based Usability Testing
   This stage is the core of the research, where respondents are asked to complete the task scenarios that were developed. During this process, researchers directly observe user interactions with the website, including the navigation flow, difficulties encountered, and errors that occur. The implementation of usability testing aims to obtain a real picture of the user experience when using a website.

4. Collection of Observation Data and SUS Questionnaires
   Data collection was carried out in two ways, namely direct observation and the System Usability Scale (SUS) questionnaire. Observational data were used to record user behavior, interaction barriers, and usability issues that arose during testing. The SUS questionnaire was used to obtain quantitative data in the form of usability scores that describe users' perceptions of the overall ease of use of the website.

5. Analysis of Usability Test Results
   The data obtained from the observations and SUS questionnaire were descriptively analyzed. The analysis was conducted by grouping the findings based on usability aspects, such as ease of use, navigation efficiency, and usage errors. The SUS scores were analyzed to provide a numerical overview of the website's usability level and were used to support the qualitative analysis results.

6. Preparation of Conclusions and Recommendations
   The data obtained from the observations and SUS questionnaires were descriptively analyzed. The analysis was conducted by grouping the findings based on usability aspects, such as ease of use, navigation efficiency, and usage errors. The SUS scores were analyzed to provide a numerical overview of the website's usability level and were used to support the qualitative analysis results.

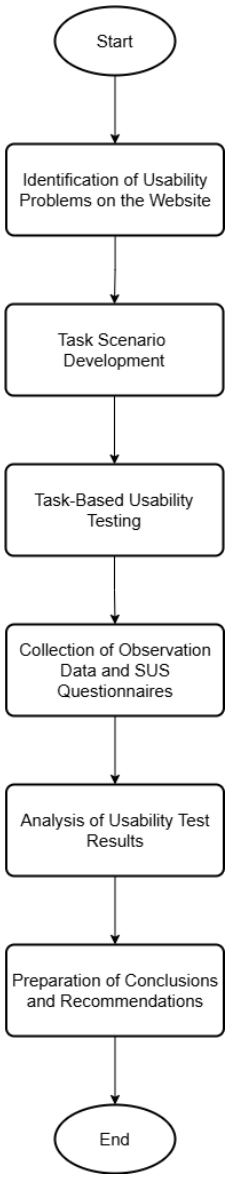


Figure 1. Research Stages

The research stages were arranged systematically and consistently with the characteristics of descriptive evaluative research, without covering the process of system design or redevelopment of the system.

4 RESULTS AND DISCUSSION

The results of the usability testing of the Tinanggea District website were obtained using the System Usability Scale (SUS) instrument. Based on the calculations from eight respondents, the average SUS score was 51.2, indicating that the website's usability level was in the low to marginal category. This score indicates that the website is still functionally usable as a medium for providing public information but does not yet provide an optimal user experience. The distribution of the System Usability Scale (SUS) scores from all respondents as a basis for assessing the level of website usability is presented in Table 3.

Table 3. System Usability Scale (SUS) Calculation Results

Respondents	SUS Scores
R1	52.5
R2	45.0
R3	47.5
R4	57.5
R5	55.0
R6	45.0
R7	52.5
R8	55.0
Averange	**51.2**

The SUS results were further analyzed with reference to the concept of usability in ISO 9241-11, which covers the aspects of effectiveness, efficiency, and satisfaction in the context of using public service websites. In terms of effectiveness, most respondents were able to complete task scenarios related to basic information searches, such as sub-district profiles and regional potential. However, some users still needed time to adapt during the initial stages of interaction to understand the information presentation. This condition shows that the system's effectiveness has been achieved, but there is still room for improvement to accelerate the achievement of the user goals.

The homepage of the Tinanggea Subdistrict website, which contains regional identity and the main navigation menu, is shown in Figure 2. The basic information presented directly on the homepage helps users understand the purpose of the website without the need for lengthy exploration. The consistency of the navigation menu on each page also supports ease of use when moving between the pages. In terms of usability, this condition shows that the website has fulfilled the aspect of learnability because the usage pattern can be learned by users in a relatively short time.

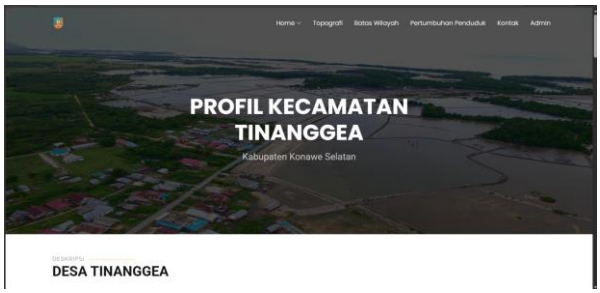


Figure 2. Tinanggea Subdistrict Website Home Page Display

In presenting information on the potential of leading villages, the website utilizes visual elements in the form of images and village tab navigation, as shown in Figure 3. This visual approach helps users obtain a quick and intuitive overview of the region's condition. However, the limited textual explanation means that the information presented tends to be visually descriptive and does not fully support the need for more in-depth information on the topic. This shows that it has achieved effectiveness in conveying information, but the completeness and depth of the content still need to be improved to optimize the quality of public information services.



Figure 3. Page on the Potential of Leading Villages with Village Tab Navigation

In addition to functioning as a medium for providing public information, the Tinanggea Subdistrict website is equipped with an Admin Dashboard feature for user data management, as shown in Figure 4. The existence of this feature shows that the website is not only static but also has operational functions that support system management and maintenance. Based on the observation results, the system implements an error prevention mechanism through input data validation and clear error messages when errors occur during data entry. This mechanism helps users understand the errors that occur and prevents repeated errors during data management.

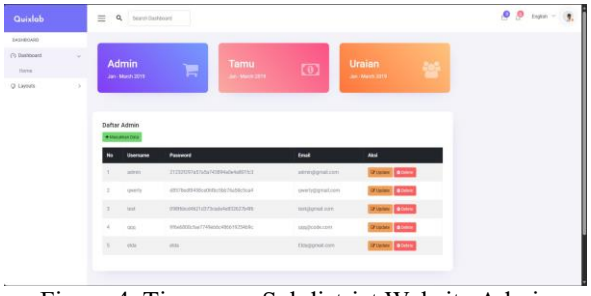


Figure 4. Tinanggea Subdistrict Website Admin Dashboard Display

An example of the application of an error prevention mechanism in the form of input data validation is shown in Figure 5, which shows that the system provides immediate feedback when mandatory data are not filled in by the user. The existence of this mechanism contributes to increasing the effectiveness and efficiency of system usage because errors can be prevented before the data storage process is carried out. With clear validation and feedback, user interactions with the system become more focused and controlled.

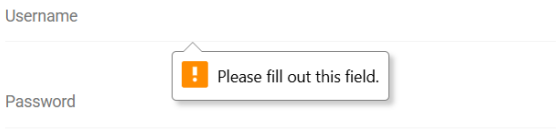


Figure 5. Example of Input Data Validation on the Admin Dashboard

Although error prevention aspects have been well implemented, the evaluation results show that the Admin Dashboard still has limitations in terms of interaction. The system does not yet provide features or space that allow administrators to respond to messages, questions, or feedback from website users. The absence of this two-way communication feature has the potential to limit the role of the website as a responsive and interactive public service medium. In addition, several inconsistencies were found in the presentation of the interface, such as differences in menu terms and component layouts on several pages, which can affect the consistency of the user's experience.

Overall, the results and discussion show that the Tinanggea Subdistrict website has fulfilled its basic function as a medium for providing public information and has implemented several usability principles, particularly in terms of ease of navigation and error prevention. However, based on a usability analysis referring to ISO 9241-11, there are still aspects that need to be improved, especially in terms of interaction efficiency, interface consistency, and communication support between administrators and users. Improvements in these aspects not only affect ease of use but also have direct implications for the quality of public services and user satisfaction. Therefore, the results of this study can be used as a basis for developing a more responsive, consistent, and user-oriented website for the Tinanggea Subdistrict

5 CONCLUSION

Based on the results of research conducted using descriptive methods with a usability testing approach, the conclusion that can be drawn is that the Tinnangea Subdistrict website has fulfilled its basic function of providing public information. This is evidenced by the ability of most respondents to complete task scenarios related to searching for basic information, such as subdistrict profiles, regional information, and regional potential.

Usability testing using the System Usability Scale (SUS) resulted in an average score of 51.2. This score falls into the low-to-marginal category, indicating that the website is still functionally usable but does not yet provide an optimal user experience. This finding is reinforced by observations showing that users need to adapt to understand the navigation structure and information presentation.

When reviewed based on the concept of usability in ISO 9241-11, the effectiveness aspect of the Tinanggea Subdistrict website has been achieved because users are able to complete their main usage objectives successfully. However, the efficiency and satisfaction aspects still need to be improved, especially in terms of interface consistency, navigation efficiency, and the completeness and depth of the information presented. In addition, the limited two-way communication features between users and administrators are one of the factors that have the potential to reduce the quality of interaction and user satisfaction.

6 SUGGESTION AND RECOMMENDATION

Overall, the results of this study show that the Tinanggea Subdistrict website has implemented several usability principles, particularly in terms of ease of navigation and error prevention. However, improvements are still needed in terms of interaction efficiency, interface design consistency, and more interactive service support. This improvement to the website can function more optimally as a responsive public information service medium that is oriented towards user needs.

REFERENCES

[1] Guido A. Ama Duli, Hendrik Toda, Delila A. Nahak Seran, and Belandina Liliana Long, "Efektivitas Penerapan E-Government dalam Mendukung Transparansi dan Keterbukaan Informasi Publik," *Jurnal Ilmiah Teknik Informatika dan Komunikasi*, vol. 5, no. 2, pp. 601–617, Jun. 2025, doi: 10.55606/juitik.v5i2.1230.

[2] R. H. Adirasyid, H. Muslimah Az-Zahra, and N. Y. Setiawan, "Evaluasi Usability Situs Web Resmi Pemerintah Kabupaten Sidoarjo Menggunakan Metode Usability Testing dan Evaluasi Heuristic," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 9, pp. 8804–8813, 2019, Accessed: Jan. 13, 2026. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6268

[3] A. F. Aqil, "Evaluasi User Interface pada Website Pemerintah Kabupaten Asahan," 2024. Accessed: Jan. 13, 2026. [Online]. Available: http://eprints.ipdn.ac.id/id/eprint/19481

[4] A. H. Muhammad, I. K. A. Mogi, and G. A. V. M. Giri, "Analisis Usability Testing Pada Website Desa Musi," *Jurnal Pengabdian Informatika (JUPITA)*, vol. 1, no. 2, pp. 447–454, 2023, Accessed: Feb. 09, 2026. [Online]. Available: https://ejournal1.unud.ac.id/index.php/jupita/article/view/187

[5] K. C. Pangemanan and M. R. Fahlevvi, "Pengaruh Evaluasi Kebergunaan Website Resmi Pemerintah Kota Tomohon (tomohon.go.id): Pendekatan Usability Testing untuk Meningkatkan Pelayanan Publik," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, pp. 2011–2020, 2025, doi: 10.23960/jitet.v13i3.6658.

[6] S. Hijriah and D. Irawan, "Analisis Website E-Government Kecamatan Menggunakan Metode Usability Testing," *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 3, pp. 419–430, Sep. 2023, doi: 10.31849/zn.v5i3.15798.

[7] S. Sajida, L. T. Husna, K. W. Rahayu, and I. A. Suryawati, "Evaluating Public Website Performance: Content Analyses on Malang City Government Website," *Journal of Governance*, vol. 8, no. 2, pp. 263–278, Jun. 2023, doi: 10.31506/jog.v8i2.19248.

[8] I. G. S. Aryandana, L. A. Susanti, P. E. Suardana, and P. V. Nugraha, "Pengujian Kualitas Website Dinas Kependudukan dan Pencatatan Sipil (DUKCAPIL) Kota Denpasar Bali Menggunakan Metode System Usability Scale (SUS)," *Jurnal Teknologi Informasi dan Komputer*, vol. 11, no. 1, pp. 14–25, Apr. 2025, doi: 10.36002/jutik.v11i1.3747.

[9] M. L. Nuriman and N. Mayesti, "Evaluasi Ketergunaan Website Perpustakaan Universitas Indonesia Menggunakan System Usability Scale," *BACA: Jurnal Dokumentasi Dan Informasi*, vol. 42, no. 2, pp. 253–269, 2020, Accessed: Jan. 12, 2026. [Online]. Available: https://ejournal.brin.go.id/baca/article/view/7540

[10] S. Wahyuni, D. Febriyanti, and N. Kencana, "Evaluasi Pengunaan (Usability) Website E-Pemerintahan di Kantor Kelurahan Demang Lebar Daun Kota Palembang," *WEDANA: Jurnal Kajian Pemerintahan, Politik dan Birokrasi*, vol. 9, no. 1, pp. 31–41, Sep. 2023, doi: 10.25299/wedana.v9i1.12388.

[11] F. S. Putra, H. Muslimah Az-Zahra, and L. Fanani, "Evaluasi Usability Aplikasi Perangkat Bergerak AlgoritmaKopi Menggunakan Metode Usability Testing," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3,

no. 8, pp. 8130–8139, 2019, Accessed: Feb. 04, 2026. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6036

[12] M. S. Tuloli, R. Patalangi, and R. Takdir, "Pengukuran Tingkat Usability Sistem Aplikasi e-Rapor Menggunakan Metode Usability Testing dan SUS," *Jambura Journal of Informatics*, vol. 4, no. 1, pp. 13–26, Apr. 2022, doi: 10.37905/jji.v4i1.13411.

[13] J. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds., London, United Kingdom: Taylor & Francis, 1996, pp. 189–194. Accessed: Feb. 03, 2026. [Online]. Available: https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale

[14] International Organization for Standardization, "ISO 9241-11:2018 Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts," Geneva, 2018. Accessed: Feb. 03, 2026. [Online]. Available: https://www.iso.org/standard/63500.html

[15] A. Cazañas, A. de San Miguel, and E. Parra, "Estimating Sample Size for Usability Testing," *Enfoque UTE*, vol. 8, no. 1, pp. 172–185, Feb. 2017, doi: 10.29019/enfoqueute.v8n1.126.

[16] L. Faulkner, "Beyond the Five-User Assumption: Benefits of Increased Sample Sizes in Usability Testing," *Behavior Research Methods, Instruments, & Computers*, vol. 35, no. 3, pp. 379–383, Aug. 2003, doi: 10.3758/BF03195514.

[17] P. Vlachogianni and N. Tselios, "Investigating the impact of personality traits on perceived usability evaluation of e-learning platforms," *Interactive Technology and Smart Education*, vol. 19, no. 2, pp. 202–221, May 2022, doi: 10.1108/ITSE-02-2021-0024.

[18] H. Farid, D. Yusup, and Carudin, "Analisis Usability pada Aplikasi Momby Spa Menggunakan Metode Usability Testing," *Jurnal Ilmiah Wahana Pendidikan*, vol. 2022, no. 14, pp. 155–163, 2022, Accessed: Feb. 04, 2026. [Online]. Available: https://jurnal.peneliti.net/index.php/JIWP/article/view/2082

[19] M. L. Nurhakim, B. T. Hanggara, I. Sartika, and E. Maghfiroh, "Analisis Usability dan Pengalaman Pengguna Website Sman 5 Bogor Versi Mobile Menggunakan Mecue Questionnaire dan System Usability Scale (SUS)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 2, pp. 773–780, 2023, Accessed: Feb. 04, 2026. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12310

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 February 2026

Systematic Literature Review: Machine Learning Algorithm Performance Evaluation of Extract-Transform-Load

Muhammad Faisal Ashshidiq, Mohammad Nurkamal Fauzan
Informatics Engineering, Vocational School, Universitas Logistik dan Bisnis Internasional
Email: sidiqfaisal30@gmail.com; m.nurkamal.f@ulbi.ac.id

Accepted:	Accepted After Revision:	Published:
16 July 2025	5 February 2026	27 February 2026

Abstract

The exponential growth of data in the digital era poses significant challenges for effective data utilization. The Extract, Transform, Load (ETL) process is the foundation for preparing large-scale, unstructured data from various sources (NoSQL databases, log files) for analysis in a data warehouse. However, handling complex data structures such as nested arrays in MongoDB is a major obstacle during the transformation phase. In addition, the purpose of the transformation process is to maintain data quality and integrity. This crucial need requires a robust mechanism for anomaly detection to identify unusual patterns or events that indicate data corruption or system errors. The process of handling system errors requires analyzing nested array data structures using relevant machine learning algorithms for anomaly detection. This literature study is expected to provide valuable insights and identify relevant algorithms in data anomaly detection after the ETL process.

Keywords: Machine Learning, Anomaly Detection, Nested Array, SLR

1 INTRODUCTION

The digital age, marked by rapid data growth known as Big Data, poses significant challenges for organizations in managing and using data effectively to support evidence-based decisions. Various industries currently collect large amounts of data from diverse sources, including relational databases and NoSQL systems such as MongoDB. However, raw data is typically unstructured, inconsistent, and contains redundancies and missing values, making it difficult for advanced analysis.

To resolve these issues, organizations implement the Extract–Transform–Load (ETL) process as a core part of modern data architecture. The ETL process involves three main steps: (1) Extract, which retrieves data from sources such as MongoDB in JSON format; (2) Transform, which includes schema modifications, normalization, nested array splitting, and integrating data relationships; and (3) Load, which involves loading the transformed data into a data warehouse for analytics and reporting.

MongoDB, as a document-oriented database, can store complex data structures, including nested arrays and hierarchical documents. Its flexibility offers storage advantages, but the complexity of these structures increases computational load during the ETL transformation stage. Processing errors can affect data quality after ETL, so evaluating data quality using metrics such as null value percentage, attribute inconsistency, duplication, and outliers is necessary. Indicators of poor data quality suggest that the ETL process has failed to meet business analytics needs.

Research and analysis clearly indicate that a mature data management strategy is essential. Unsupervised machine learning (ML) methods are used to automate ETL performance evaluation and detect anomalies in data quality metrics. Techniques such as Isolation Forest, Local Outlier Factor (LOF), and One-Class Support Vector Machine (OC-SVM) are used to identify abnormal patterns in large-scale unlabeled data. Using ML improves anomaly detection accuracy while maintaining ETL system scalability and reducing processing delays in production.

However, existing research shows that the most popular algorithms are not always the most effective in every scenario. Some studies choose certain methods for ease or popularity, while others highlight the superior performance of alternative algorithms under specific conditions, such as high dimensionality or concept drift. Therefore, a systematic approach is needed to differentiate between how often algorithms are used and

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

how effectively they are based on comprehensive evaluation criteria.

Performance evaluation of anomaly detection during ETL processing cannot rely on a single metric, such as accuracy. This study adopts an integrated approach, considering three key dimensions: (1) detection accuracy, including precision, recall, and F1-score; (2) scalability, involving computation time, ability to handle large data, and compatibility with real-time ETL; and (3) interpretability, which assesses how well data practitioners can understand detection results to support operational decisions.LO

High anomaly-detection accuracy can be misleading, especially with highly imbalanced data where anomalies are rare compared to normal data. Results show models can achieve high accuracy simply by predicting the majority of class but fail to detect critical issues that indicate data quality problems. Hence, the literature increasingly emphasizes recall-based metrics, false positive rates, and the business impact of detection errors.

This journal presents a Systematic Literature Review (SLR) following PRISMA guidelines to examine research published from 2019 to 2025 on applying machine learning to ETL evaluation and optimization, particularly for semi-structured data and nested arrays. The article selection process aligns with the PRISMA flow diagram, covering identification, screening, eligibility, and final inclusion. As part of the review, several selected journal articles are summarized in a table, including authors, publication year, dataset types, algorithms used, and main contributions:

Table 1. Summary table of selected articles

Author	Year	Dataset	Algorithm	Contribution
Marcelli E, Barbariol T, Susto G	2022	Sensor data (Node-RED),	Isolation Forest	High AUC (0.997), F1 (95.29) active learning for anomaly detection [1].
Usman N, Utami E, Hartanto A	2023	Credit transactions.	LOF	Accuracy 97.6%, F1 (97). Density-based outlier scoring [2].
Usman N, Utami E, Hartanto A	2023	High Dimensional Sensor data	OC-SVM	AUC 0.9, TPR 0.9. efficient for high-dim anomalies [2].

The SLR also includes a dedicated subsection on quality assessment of the studies, considering clarity of methodology, reproducibility, dataset relevance, and transparency in reporting evaluation metrics. The approach ensures that the synthesis is both quantitative and methodologically solid.

Based on the literature review, the study identifies open research gaps such as limited exploration of hybrid ML-rule-based ETL approaches that combine machine learning flexibility with deterministic rule reliability, the scarcity of studies on online anomaly detection in streaming ETL environments especially for real-time settings with concept drift and the lack of integration of Explainable Artificial Intelligence (XAI) methods to clarify the causes of anomalies in ETL processes, helping practitioners to follow up effectively.

Therefore, this systematic review aims to provide an overview of research trends, algorithm effectiveness, evaluation criteria, and future directions for applying machine learning to improve data quality and Extract-Transform-Load (ETL) system performance.

1.1 Research Questions

There are several research questions in this study, among others:

1. RQ1: What is the most effective Machine Learning method for anomaly detection in nested array data after the Extract-Transform-Load process?
2. RQ2: How is the implementation and performance evaluation of machine learning algorithms in optimizing the Extract-Transform-Load process on large-scale data?

2 LITERATURE REVIEW

2.1 ETL Process (Extract, Transform, Load)

The ETL process is a fundamental methodology in data warehousing that consists of three stages: extracting data from MongoDB in JSON format (Extract), transforming the data schema, and modifying some related data for analysis purposes (Transform), and the transformed data is loaded into the data warehouse (Load) for further analysis as reporting [3]. Data with large volumes is known as “Big Data”. To handle such

large amounts of data, an ETL process is needed to make data processing easier, faster, and automated, using Apache Airflow [4].

2.2 Machine Learning

Machine Learning (ML) is a technology to address various challenges, especially in the final output of the ETL process. ML algorithms can automate several aspects of data transformation, proactively identify anomaly patterns, and predict potential issues [5]. Machine learning has two types of approaches: supervised and unsupervised. Supervised machine learning requires labeled data to train the model, whereas unsupervised machine learning does not, as its approach focuses on discovering patterns within the data independently [6].

2.3 MongoDB

MongoDB is a leading NoSQL database and a choice for storing structured and semi-structured data. MongoDB is one of the sources used for analysis to obtain business information within an organization [7] MongoDB's unique ability to store nested arrays provides flexibility but also presents challenges in the ETL process.

Research conducted by [8] shows that the schema-less MongoDB data structure requires an adaptive ETL approach to handle the variations in dynamic data structures [9] A comparative study comparing ETL performance between relational databases. The main challenges in the ETL process using data from MongoDB include denormalization, handling nested objects, and query optimization for unstructured data [10] Proposing that MongoDB's complex array improves ETL performance by up to 40% [11], [12].

2.4 Anomaly Detection

Anomaly detection is the process of identifying patterns that do not conform to normal expectations. Data anomalies can include unreasonable values, unusual patterns, or data inconsistencies that can affect the quality of analysis [12]. Anomaly detection after the ETL process becomes crucial to maintain the integrity of the processed data [13]

3 RESEARCH METHODS

3.1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

The Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) method is a systematic and structured framework for conducting literature reviews [14]. The process includes the identification, evaluation, and synthesis of results from several previous studies in a systematic and structured manner [14]. The approach is chosen as a review of Machine Learning algorithms in optimizing the Extract-Transform-Load (ETL) process, especially in handling complex data such as nested arrays and anomaly detection [3]. In this study, we followed the PRISMA guidelines to review various machine learning algorithms, reducing potential bias and ensuring the objectivity of the literature review process through transparent and systematic reporting [14].

3.2 Stages of Systematic Literature Review

This research involves several stages that must be carried out when studying and summarizing previous research to provide an overview and guide the process from planning to final reporting of results.

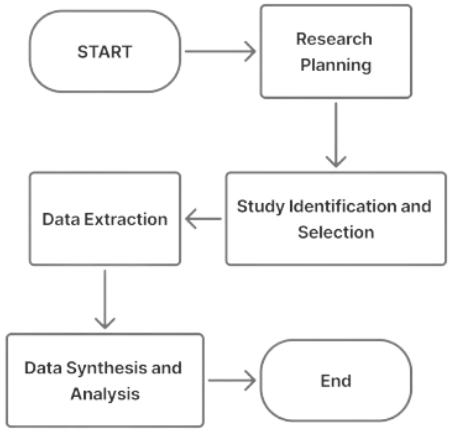


Figure 1. Flowchart Systematic Literature Review

Figure 1 shows the planning diagram for the systematic literature review conducted to examine and summarize several journal articles from 2019 to 2025. There are some discussions in the systematic literature review chart to understand the process carried out. Here is the discussion of the flowchart:

1. Formulation of research questions
   This research is guided by two Research Questions (RQs) designed to explore in depth the role of machine learning algorithms in evaluating the performance of the Extract-Transform-Load (ETL) process. Specifically related to anomaly detection in complex data structures.

2. Inclusion and Exclusion Criteria
   This process ensures that the results are relevant and that the quality of the studies to be analyzed is high. A set of inclusion and exclusion criteria is established. The inclusion and exclusion criteria serve as filters in the journal article selection

process. Additionally, the researcher will focus on the main objective, which is literature relevant to the predetermined main question. The following are the objectives of the research, among others:

a. Inclusion Criteria
Relevant journal articles on the topic are published by reputable publishers such as IEEE, ACM, Springer, Elsevier, MDPI, and others. The journal articles specifically discuss the application of machine learning algorithms to evaluate Extract-Transform-Load performance, anomaly detection, data processing for storage in data warehouses, data schema transformation, and several processes to address issues with nested array data types, all with a modular structure.

b. Exclusion Criteria
Identified duplicate studies, deleting irrelevant journal articles based on a review of titles and abstracts, and studies that do not directly discuss the application of machine learning for anomaly detection or the performance of Extract-Transform-Load on nested array data [15].

3. Data Sources and Search Strategies
Literature research is conducted through various methods, using a variety of sources, including reputable academic databases, to ensure coverage of relevant studies [16]. The literature will be processed using well-known academic databases, such as Scopus.

This search strategy uses keywords, then compiles several relevant journal articles. The results obtained are approximately 40 journal articles. These results are based on a systematic literature review process.

3.3 Study Identification and Selection

This stage involves identifying journal article data and screening studies according to the established planning criteria. Based on the process carried out, the PRISMA guidelines need to be followed to ensure transparency and reproducibility. Here are some of the steps taken, among others:

1. The specified database search is for journal articles indexed academically, one of the well-known examples being Scopus.
2. Data deletion of irrelevant journal articles.
3. Filtering titles and abstracts.
4. Full text screening for journal articles that pass the relevance verification process and comply with inclusion and exclusion criteria.

3.4 Data Extractions

This stage follows the analysis of relevant journal articles, in which input data are selected to classify the articles that will serve as references for this research. This process will produce information on several important aspects, including the publication, the research methodology used, the characteristics of the dataset employed, the performance metrics used for evaluation, and the solutions proposed by previous studies to assist future research on similar topics.

3.5 Data Synthesis and Analysis

In the final stage, the extraction and processing stage, the final visualization synthesis results are produced based on the publication years of the journal articles, which are set from 2019 to 2025. From the approximately 7 years of selected publication years, the data will be analyzed to answer the research questions.

4 RESULTS AND DISCUSSION

The results of the process carried out through a systematic literature review present the final outcomes of the identification and selection of journal articles, based on well-known academic studies. Based on the processing conducted using the PRISMA method, the following will be discussed:

4.1 Keyword Search Results

The stages of identifying and selecting studies are the initial steps in collecting relevant literature. Processing begins with keyword searches related to the journal article topics referenced. There are several well-known journal publishers, and then the next step is screening, aimed at ensuring that journal articles meeting the inclusion criteria are selected as the final results. The selection of criteria is based on the results of the journal article screening process, summarized in Table 1, which outlines a systematic procedure for sorting and selecting relevant studies.

Table 2. Stages of a systematic literature review

Stages	Explanation
Research	The process of removing duplicate literature is limited by search processors, resulting in the database in Watase or the selection of literature that does not match the topic.

Selection Based on Title and Abstract	Filtering journal articles based on titles and abstracts. Selecting relevant and irrelevant studies.
Selection Based on the method used for research topic adjustment	Filtering relevant topics based on the method or model used.
Selection based on full text	A comprehensive analysis for the extraction process of relevant journal article selection results will be automatically saved in the full text menu.
Documentation of the Selection Process	The entire selection process is documented using PRISMA to show the journal articles reviewed at each stage.

The discussion based on the table visualization above has resulted in various keyword searches. Here are the final results of the keyword search.

No	Keyword	Raw
1	performance evaluation, big data	47
2	Machine Learning Algorithms, big data	64
3	anomaly detection, isolation forest	56
4	algorithm data warehouse	30
5	extract transform load	16

Figure 2. Identify and select keywords

Note that Figure 2 shows the final result of keyword identification and selection processing. The process aims to identify relevant journal articles for the most general topic. The search was conducted using general keywords and more specific research keywords. Below are some of the final results obtained from the search for relevant journal articles.

1. "Performance evaluation, big data." This keyword retrieves journal data. A total of 47 data points were filtered down to about 14.
2. "Machine Learning Algorithm, big data" This keyword retrieves 64 journal data entries, which are then filtered down from 64 to approximately 24 entries.
3. "Anomaly detection, isolation forest." This keyword retrieved a total of 56 journal data entries, which were then filtered down to 32.
4. "Algorithm data warehouse." This keyword has 30 journal data entries, which are then filtered down to 7.
5. "Extract, transform, load." For this keyword, there are 16 journal data entries, which were then filtered down to about 5.

The keyword search above filters journal article data in the Watase database. The publication year follows a 7-year cycle, from 2019 to 2025, and is indexed by reputable publishers such as Scopus. After successful processing, the next step is to produce a PRISMA diagram.

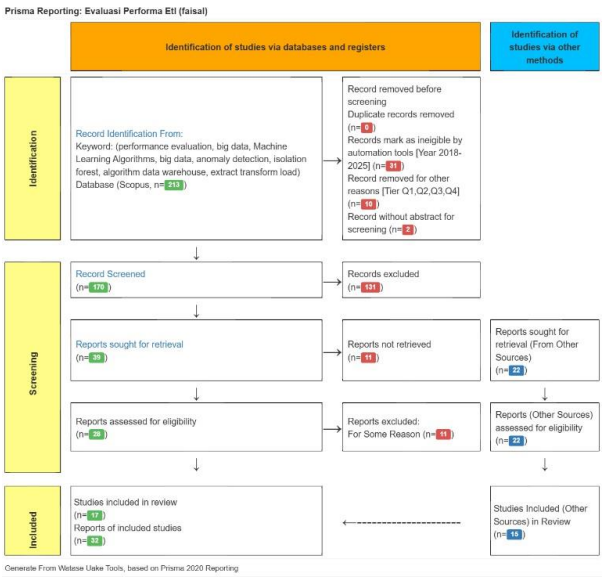


Figure 3 PRISMA Result

Please refer to Figure 3, which shows the final results of the keyword search process, also known as the identification and selection of studies. The search process was conducted using databases from various reputable publishers. A total of 213 journal article data entries were found. Of these, 31 irrelevant and duplicate journal articles were removed. Thus, 182 data entries remained. Next, approximately 10 journal articles that did not meet the criteria were deleted, leaving approximately 172. The remaining data were reduced by 2 journal articles that lacked abstracts, resulting in a final PRISMA output of 170 journal articles needed for the study.

The processing of these 170 journal articles will be repeated, with some excluded, resulting in 39 articles suitable for the research. Next, a retrieval process was carried out to select relevant articles. The selection was done manually, resulting in 11 reports that could not be obtained, while the remaining 28 reports were successfully retrieved and deemed suitable as research references. Additionally, the final results of the keyword search for journal articles produced a visualization of a synthesis graph by year, showing trends over time, in accordance with external data from publishers.

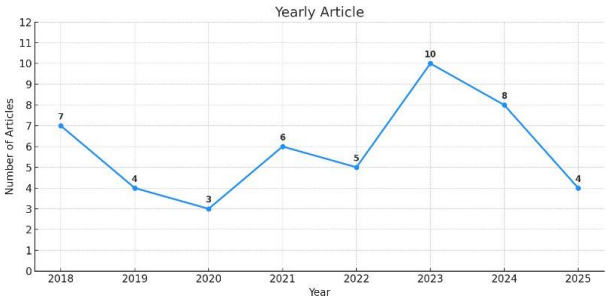


Figure 4. Exclusive results based on the year of journal article publication

Note that Figure 4 shows the graph of the number of journal articles selected based on the publication year, from 2018 to 2025. The publication year 2018 successfully obtained 7 journal articles, the year 2021 obtained 6 journal articles, the year 2022 obtained 5 journal articles, and the year 2024 obtained 8 journal articles. The most journal articles were obtained from the publication year 2023, with 10 articles, while the fewest were in 2020, with 3 articles. Based on the line graph, the average is 4 journal articles for 2019 and 2025.

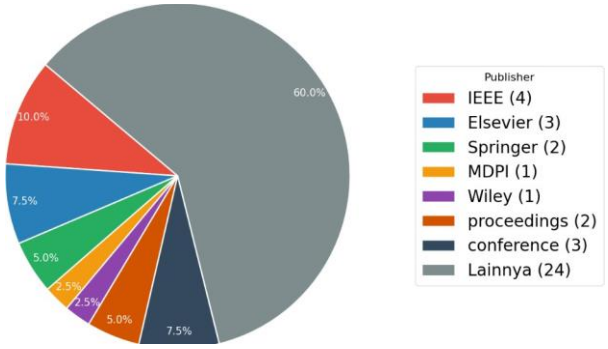


Figure 5. Results of the illusion based on a reputable publisher

Note that Figure 5 is a visualization of the final inclusion results based on various reputable publishers. The visualization was conducted to analyze the distribution of relevant journal articles successfully found by different publishers. Based on the visualization, the source with the most articles is IEEE, with around 4 journal articles, while the source with the fewest is MDPI, and Wiley managed 1 journal article. The final results, displayed as a pie chart, will proceed to the next stage: processing to answer the research questions.

RQ1: What is the most effective Machine Learning method for anomaly detection in nested array data after the Extract-Transform-Load process?

Anomaly detection is the process of identifying unusual patterns or events, such as results from data transformation processes that experience discrepancies or are considered null [17]. Anomaly is also known as outlier, deviation, or exception [18]. Anomalies are data points or entire patterns that deviate from normal behavior. Therefore, anomaly detection is performed to tidy up data with nested array types by converting them into a modular format so the system can understand them. Managing nested data types using the Extract-Transform-Load strategy presents complex challenges that require machine learning algorithms to handle hierarchical and unstructured data structures. Thus, the systematic literature review conducted reveals various machine learning methods that demonstrate the effectiveness of anomaly detection for nested array data patterns [19].



Figure 6. Types of machine learning algorithms

Figure 6 is a visualization of the mind mapping, based on the analysis process that identified two types of machine learning algorithms. From among these algorithms, the researcher will select the most efficient for anomaly detection. The selection is based on the results of the analysis, through visualization and understanding, that have been conducted. The most commonly used algorithm for anomaly detection is unsupervised, and the metric most commonly used to assess whether the data is anomalous or normal is accuracy.



Figure 7. Types of Algorithms for Anomaly Detection

Based on scientific analysis, machine learning algorithms are divided into two types: supervised and unsupervised. The researcher further analyzed and found that anomaly detection processing is often performed using unsupervised algorithms [6], [20]. Therefore, it has been concluded that the results of unsupervised algorithm analysis are frequently used, including by other researchers. Unsupervised algorithms branch into many methods, so researchers choose efficient methods for anomaly detection, data management, and error checking in data. The following are:

1. Isolation Forest
   This algorithm is designed to detect anomalies. Its approach is to isolate data observations that are considered anomalies from those considered normal. [21], [22] The Isolation Forest method demonstrates strong performance, with an AUC of 0.997, an F1-Score of 95.29%, a Precision of 95.73%, and a recall of 94.85%.

2. Local Outlier Factor (LOF)
   This algorithm is designed to measure the 'anomaly' value, such as data points based on data density compared to neighbor looseness. Data points with a significantly higher Local Outlier Factor than their neighbors are considered outliers [23].

3. One-Class Support Vector Machine (OC-SVM)
   This algorithm is designed to classify normal data, identifying data points that are 'extraordinarily' normal as anomalies [24] stated that the OC-SVM algorithm shows performance with an AUC value of 0.9, FPR of 0.39, and TPR of 0.9.

Table 3. Total Data

Algorithm	Citation	Low Metrics	High Metrics	Data Transformation
Isolation Forest	[1]	AUC: 0.935 F1-Score: 93.7 Precision: 93.86 Recall: 93.08	AUC: 0.997 F1-Score: 95.29 Precision: 95.73 Recall: 94.85	Historical machine sensor data from Node-RED to tabular.
Local Outlier Factor (LOF)	[2]	AUC: 0 Accuracy: 77.8 F1-Score: 50 F2-Score: 14 Precision: 50.3 Recall: 50.4 AUROC: 0.5465	AUC: 0.956 Accuracy: 97.6 F1-Score: 97 F2-Score: 95.7 Precision: 97.7 Recall: 97.6 AUROC: 0.4859	Credit card transaction data into tabular.
One-Class Support Vector Machine (OC-SVM)	[2]	AUC: 0.64 FPR: 0.0 TPR: 0.4	AUC: 0.9 FPR: 0.39 TPR: 0.9	Sensor data dimensions are high for feature representation.

Based on the analysis of the table visualization above, the most frequently used algorithm is Isolation Forest to detect anomalies in data that has been processed using the Extract-Transform-Load strategy. [17], [19].

After managing data using the Extract-Transform-Load (ETL) strategy with nested array data from a company in Indonesia, the processing needs to consider several methods as key to the success of implementing machine learning algorithms. Here are some effective data preprocessing steps, among others:

a. Flattening/De-nesting is the process of transforming a nested array structure into a tabular representation.
b. Feature Engineering is a process of extracting various relevant features from nested data array patterns that can be described through the data characteristics comprehensively, such as the number of elements in the array, the data type of elements, or specific patterns.
c. Missing Value Imputation is the process of removing or eliminating unnecessary or potentially appearing values during the Extract-Transform-Load process.
d. Encoding categorical data is the process of converting categorical data into a numerical format.

Based on the analysis, it was found that the most frequently used algorithm is Isolation Forest for anomaly detection, and that the processing takes a long time. Therefore, it is necessary to find out information so that the processing time is kept under about 10 minutes to avoid taking too long in data management.

RQ2: How is the implementation and performance evaluation of machine learning algorithms in optimizing the Extract-Transform-Load process on large-scale data?

The application of machine learning (ML) in the Extract-Transform-Load process offers great potential to improve efficiency, accuracy, and automation. However, the implementation faced challenges, such as difficulties in making accurate and careful decisions. The implementation involved a transformation process by adjusting the data schema or data warehouse structure (BigQuery). Then, an implementation issue occurred, so an evaluation of data performance was necessary by

applying machine learning to identify the root cause and determine which data needs to be followed up on.

Here are some metrics that can be used as considerations [25]:

1. Anomaly detection accuracy
   This metric is used to measure precision, recall, and F1-Score.

2. Efficiency of time and computational resources.
   This metric is applied to process data quickly, within approximately 1 minute, without burdening the device's resources.

3. Quality of transformed data
   This metric is used to measure the consistency and completeness of data from the Extract-Transform-Load process.

Among the three materials, the researcher took a wise step by conducting further analysis to understand the meaning of the information obtained. As a result, several machine learning application methods were identified that can assist in many aspects, including:

Table 4. Name of the machine learning method for evaluating ETL performance

Machine Learning Method	**Explanation**	**Citation**
Improvement of Data Quality	This method is used to identify and mark inconsistent, duplicate, or functional data to ensure the quality of data that has been uploaded to the data warehouse.	[26], [27]
Data Transformation Automation	This method is used to handle nested arrays. Additionally, the application of machine learning can learn transformation patterns efficiently and reduce the need for complex manual rules.	[5], [28]
Proactive Anomaly Detection	This Machine Learning method is capable of performing real-time or near-real-time anomaly pattern detection during the ETL process.	[13], [23], [24], [29]
Performance Optimization of Extract-Transform-Load	This Machine Learning method can analyze logs and performance metrics of Extract-Transform-Load processes to identify bottlenecks or areas that need optimization. Examples include job scheduling or resource allocation.	[19], [25], [30]

Through the Extract-Transform-Load (ETL) workflow, the design is processed to prepare data for effective use [31]. Data is collected and then processed according to the needs of analysis, reporting, and ultimately used to make wise business decisions. [32] With the three stages completed, the results are accurate, consistent, and relevant data [3].

The processing is carried out, and the analysis has found that the understanding that can be drawn is a transformation process by changing raw data into a format ready for analysis. An example is managing data using transaction data. [33] The initial data was in a nested array pattern, then its structure was transformed into a tabular format that can be understood by analytics, so it can be analyzed into detailed explanations. Processing was carried out using an unsupervised method so that the design could explain activities, scope, objects, materials, tools, locations, and data collection techniques, and operational definitions of research variables as an effective analysis technique.

5 CONCLUSION

The processing conducted using a systematic literature review strategy can be summarized that there are two types of machine learning algorithms, namely supervised and unsupervised, which are then further narrowed down according to the research needs. Based on the results of narrowing down various types of unsupervised algorithms, three were identified: isolation

forest, local outlier factor, and one-class support vector, all of which have good effectiveness in detecting anomalies in nested array data patterns after processing data with the Extract-Transform-Load strategy. The algorithms obtained were then selected based on their frequent use because they demonstrated some of the best performance, with isolation forest being the most prominent. The results of the isolation forest algorithm include data evaluation metrics with the implementation of optimized Extract-Transform-Load processing to improve efficiency, accuracy, and automation. However, there are challenges in processing and managing data, so careful consideration of the evaluation metrics results is necessary, such as recalculating anomaly detection accuracy, time efficiency, and data quality from the data transformation results.

6 SUGGESTION AND RECOMMENDATION

This research provides a solid foundation for developing data management using a more effective Extract-Transform-Load (ETL) strategy, integrating machine learning algorithms to detect anomalies in nested array data structures.

REFERENCES

[1] E. Marcelli, T. Barbariol, and G. A. Susto, "Active Learning-based Isolation Forest (ALIF): Enhancing Anomaly Detection in Decision Support Systems," Jul. 2022.

[2] N. Usman, E. Utami, and A. D. Hartanto, "Comparative Analysis of Elliptic Envelope, Isolation Forest, One-Class SVM, and Local Outlier Factor in Detecting Earthquakes with Status Anomaly using Outlier," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, IEEE, Feb. 2023, pp. 673–678. doi: 10.1109/ICCoSITE57641.2023.10127748.

[3] J. C. Nwokeji and R. Matovu, "A Systematic Literature Review on Big Data Extraction, Transformation and Loading (ETL)," 2021, pp. 308–324. doi: 10.1007/978-3-030-80126-7_24.

[4] J. Nwokeji, F. Aqlan, A. Anugu, and A. Olagunju, "Big Data ETL Implementation Approaches: A Systematic Literature Review (P)," Jul. 2018, pp. 714–721. doi: 10.18293/SEKE2018-152.

[5] F. Raymand, B. Najafi, A. Haghighat Mamaghani, A. Moazami, and F. Rinaldi, "Machine learning-based estimation of buildings' characteristics employing electrical and chilled water consumption data: Pipeline optimization," *Energy Build.*, vol. 295, p. 113327, Sep. 2023, doi: 10.1016/j.enbuild.2023.113327.

[6] Y. Gong, F. Gu, K. Chen, and F. Wang, "The Architecture of Micro-services and the Separation of Frond-end and Back-end Applied in a Campus Information System," in *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications(AEECA)*, IEEE, Aug. 2020, pp. 321–324. doi: 10.1109/AEECA49918.2020.9213662.

[7] A. A. Yulianto, "Extract Transform Load (ETL) Process in Distributed Database Academic Data Warehouse," *APTIKOM Journal on Computer Science and Information Technologies*, vol. 4, no. 2, pp. 61–68, Jul. 2019, doi: 10.11591/APTIKOM.J.CSIT.36.

[8] M. Gorawski, K. Pasterak, A. Gorawska, and M. Gorawski, "The stream data warehouse: Page replacement algorithms and quality of service metrics," *Future Generation Computer Systems*, vol. 142, pp. 212–227, May 2023, doi: 10.1016/j.future.2023.01.003.

[9] S. R. Cheruku, S. Jain, and A. Aggarwal, "Managing Data Warehouses in Cloud Environments: Challenges and Solutions," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 6, no. 8, Sep. 2024, doi: 10.56726/IRJMETS61249.

[10] F. F. Hasan and M. S. A. Bakar, "Data Transformation from SQL to NoSQL MongoDB Based on R Programming Language," in *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2021, pp. 399–403. doi: 10.1109/ISMSIT52890.2021.9604548.

[11] B. R. Chang, H.-F. Tsai, and Y.-D. Lee, "Integrated High-Performance Platform for Fast Query Response in Big Data with Hive, Impala, and SparkSQL: A Performance Evaluation," *Applied Sciences*, vol. 8, no. 9, p. 1514, Sep. 2018, doi: 10.3390/app8091514.

[12] A. Herreros-Martínez, R. Magdalena-Benedicto, J. Vila-Francés, A. J. Serrano-López, S. Pérez-Díaz, and J. J. Martínez-Herráiz, "Applied Machine Learning to Anomaly Detection in Enterprise Purchase Processes: A Hybrid Approach Using Clustering and Isolation Forest," *Information*, vol. 16, no. 3, p. 177, Feb. 2025, doi: 10.3390/info16030177.

[13] M. Nalini, B. Yamini, C. Ambhika, and R. Siva Subramanian, "Enhancing early attack detection: novel hybrid density-based isolation forest for improved anomaly detection," *International Journal of Machine Learning and Cybernetics*, vol. 16, no. 5–6, pp. 3429–3447, Jun. 2025, doi: 10.1007/s13042-024-02460-5.

[14] A. R. Fadillah and M. N. Fauzan, "Systematic Literature Review: Identifying Key Variables and Measuring Maximum Loan Limits," *Jurnal ELTIKOM : Jurnal Teknik Elektro, Teknologi Informasi dan Komputer*, vol. 8, no. 2, pp. 100–110, Dec. 2024, doi: 10.31961/eltikom.v8i2.1156.

[15] S. Mishra, S. Konidala, and J. Manda, "Improving the ETL process through declarative transformation languages ," *Distributed Learning and Broad Applications in Scientific Research*, vol. 5, Jun. 2019.

[16] B. Oliveira, Ó. Oliveira, T. Matos, V. Santos, and O. Belo, "AN ETL PATTERN FOR LOG CONFIGURATION AND ANALYSIS," in *Proceedings of the International Conferences Big Data Analytics, Data Mining and Computational Intelligence 2019; and Theory and Practice in Modern Computing 2019*, IADIS Press, Jul. 2019, pp. 39–46. doi: 10.33965/bigdaci2019_201907L005.

[17] A. Herreros-Martínez, R. Magdalena-Benedicto, J. Vila-Francés, A. J. Serrano-López, S. Pérez-Díaz, and J. J. Martínez-Herráiz, "Applied Machine Learning to Anomaly Detection in Enterprise Purchase Processes: A Hybrid Approach Using Clustering and Isolation Forest," *Information*, vol. 16, no. 3, p. 177, Feb. 2025, doi: 10.3390/info16030177.

[18] B. J. Wheeler and H. A. Karimi, "Enhancing Hyperspectral Anomaly Detection Algorithm Comparisons: Leveraging Dataset and Algorithm Characteristics," *Remote Sens. (Basel).*, vol. 16, no. 20, p. 3879, Oct. 2024, doi: 10.3390/rs16203879.

[19] A. Gautama Putrada, I. Dian Oktaviani, M. Nurkamal Fauzan, and N. Alamsyah, "CNN Pruning for Edge Computing-Based Corn Disease Detection with a Novel NG-Mean Accuracy Loss Optimization," *Telematika*, vol. 17, no. 2, pp. 68–83, Aug. 2024, doi: 10.35671/telematika.v17i2.2899.

[20] E. F. Agyemang, "Anomaly detection using unsupervised machine learning algorithms: A simulation study," *Sci. Afr.*, vol. 26, p. e02386, Dec. 2024, doi: 10.1016/j.sciaf.2024.e02386.

[21] D. Sartor, T. Barbariol, and G. A. Susto, "Bayesian active learning isolation forest (B-ALIF): A weakly supervised strategy for anomaly detection," *Eng. Appl. Artif. Intell.*, vol. 130, p. 107671, Apr. 2024, doi: 10.1016/j.engappai.2023.107671.

[22] G. Hannák, G. Horváth, A. Kádár, and M. D. Szalai, "<scp>Bilateral-Weighted</scp> Online Adaptive Isolation Forest for anomaly detection in streaming data," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 16, no. 3, pp. 215–223, Jun. 2023, doi: 10.1002/sam.11612.

[23] M. S. Hossain and H. Mahmood, "Short-Term Load Forecasting Using an LSTM Neural Network," in *2020 IEEE Power and Energy Conference at Illinois (PECI)*, IEEE, Feb. 2020, pp. 1–6. doi: 10.1109/PECI48348.2020.9064654.

[24] Y. Qiao, K. Wu, and P. Jin, "Efficient Anomaly Detection for High-Dimensional Sensing Data With One-Class Support Vector Machine," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 404–417, Jan. 2023, doi: 10.1109/TKDE.2021.3077046.

[25] J. C. Quiroz, T. Chard, Z. Sa, A. Ritchie, L. Jorm, and B. Gallego, "Extract, transform, load framework for the conversion of health databases to OMOP," *PLoS One*, vol. 17, no. 4, p. e0266911, Apr. 2022, doi: 10.1371/journal.pone.0266911.

[26] E. Widad, E. Saida, and Y. Gahi, "Quality Anomaly Detection Using Predictive Techniques: An Extensive Big Data Quality Framework for Reliable Data Analysis," *IEEE Access*, vol. 11, pp. 103306–103318, 2023, doi: 10.1109/ACCESS.2023.3317354.

[27] T. Nguyen, H.-T. Nguyen, and T.-A. Nguyen-Hoang, "Data quality management in big data: Strategies, tools, and educational implications," *J. Parallel Distrib. Comput.*, vol. 200, p. 105067, Jun. 2025, doi: 10.1016/j.jpdc.2025.105067.

[28] F. F. Hasan and M. S. A. Bakar, "Data Transformation from SQL to NoSQL MongoDB Based on R Programming Language," in *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, Oct. 2021, pp.

399–403. doi: 10.1109/ISMSIT52890.2021.9604548.

[29] S. Alam, S. K. Sonbhadra, S. Agarwal, and P. Nagabhushan, "One-class support vector classifiers: A survey," *Knowl. Based. Syst.*, vol. 196, p. 105754, May 2020, doi: 10.1016/j.knosys.2020.105754.

[30] Q. Yang and Y. Tang, "Big Data-based Human Resource Performance Evaluation Model Using Bayesian Network of Deep Learning," *Applied Artificial Intelligence*, vol. 37, no. 1, Dec. 2023, doi: 10.1080/08839514.2023.2198897.

[31] J. Awiti, "Algorithms and Architecture for Managing Evolving ETL Workflows," 2019, pp. 539–545. doi: 10.1007/978-3-030-30278-8_51.

[32] D. Andriansyah, "Implementasi Extract-Transform-Load (ETL) Data Warehouse Laporan Harian Pool," *Jurnal Teknik Informatika*, vol. 8, no. 2, pp. 45–49, Aug. 2022, doi: 10.51998/jti.v8i2.486.

[33] M. Hendayun, E. Yulianto, J. F. Rusdi, A. Setiawan, and B. Ilman, "Extract transform load process in banking reporting system," *MethodsX*, vol. 8, p. 101260, 2021, doi: 10.1016/j.mex.2021.101260.

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 February 2026

Design and Implementation of a Web-Based Marketing Information System for Oil Palm Midrib Products in the Kayu Raja Village Community

Riski, Muh Rasyid Ridha, Usman
Information Systems, Engineering and Computer Science, Universitas Islam Indragiri
Email: daenkrhisky11@gmail.com; rasyid4sky@gmail.com; usmanovsky13411@gmail.com

Accepted:	Accepted After Revision:	Published:
23 February 2026	27 February 2026	27 February 2026

Abstract

The marketing of oil palm midrib products in Kayu Raja Village remains conventional, limiting market reach and sales efficiency. While numerous studies have examined web-based marketing systems for general MSMEs, limited research has addressed user-centered system design and evaluation for rural communities marketing agricultural by-products such as oil palm midribs. This study aims to design, implement, and evaluate a web-based marketing information system tailored to the characteristics of the Kayu Raja Village community. The research adopts the Software Development Life Cycle (SDLC) using the Waterfall model, encompassing requirements analysis, UML-based system design, implementation, and testing. System evaluation was conducted through black-box testing to assess functional performance and user acceptance testing (UAT) to measure usability and user satisfaction. The findings indicate that the developed system effectively provides product catalog management, online ordering, and structured transaction data processing. User evaluation results demonstrate that the system is functional, user-friendly, and aligned with rural user capabilities. The implementation contributes to expanding market access, improving marketing efficiency, and strengthening the economic competitiveness of the village community.

Keywords: Marketing Information System, Web-Based Marketing, Rural MSMEs, Oil Palm Midrib, Marketing Digitalization.

1 INTRODUCTION

The digitalization of marketing systems among rural micro, small, and medium enterprises (MSMEs) has become an important issue in Indonesia, particularly in improving market access and data management efficiency. In Indonesia, digitalization has become a key catalyst in supporting the growth of the micro, small, and medium enterprises (MSME) sector. According to data from Statistics Indonesia (BPS) in 2024, approximately 64.2% of MSMEs have adopted digital technologies; however, the majority remain concentrated in urban areas. Meanwhile, rural MSME actors continue to lag behind in technology adoption due to limitations in infrastructure and digital literacy [1].

Kayu Raja Village is one of the rural areas with economic potential derived from oil palm midrib products, a by-product of oil palm leaves utilized as raw material for broomsticks and other household items. Based on field observations and interviews conducted in this study, oil palm midrib processing represents a recurring productive activity among local households and contributes to the village's small-scale economic activities. These products have commercial value and relatively stable market demand. However, the marketing of oil palm midrib products in Kayu Raja Village is still conducted through traditional practices, namely direct sales to collectors or local markets. Such marketing patterns result in a very limited market reach, and product prices tend to be determined by intermediaries, thereby preventing the community from achieving optimal income levels.

In addition to the limited market reach, the community of Kayu Raja Village also lacks an integrated marketing information system to manage product data, promotion, and sales transactions. Sales recording processes are still carried out manually, making it difficult for business actors to monitor business development and make strategic decisions. This condition is consistent with previous findings indicating that rural entrepreneurs generally face constraints in utilizing digital technology due to limited knowledge and inadequate supporting facilities [2].

These findings indicate that rural MSMEs face multiple barriers in adopting digital marketing, including limited digital literacy and infrastructure constraints [2]. While previous studies have widely examined digital marketing adoption and its impact on MSME

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

performance, limited attention has been given to the development and evaluation of structured web-based marketing information systems specifically tailored to rural communities and agricultural by-product commodities. In particular, research focusing on user-oriented system design for small-scale rural enterprises remains insufficient. Therefore, there is a need to design and implement a marketing information system that aligns with the operational characteristics and technological capabilities of rural business actors. The implementation of web-based marketing information systems has been shown to enhance the effectiveness of MSME product marketing. Research demonstrates that the adoption of digital marketing significantly influences the sustainable growth of MSMEs in Indonesia [3]. Web-based systems enable business actors to display product catalogs, update pricing information, and receive online orders at relatively lower costs compared to conventional marketing methods [2].

The limited understanding of the strategic value of digitalization has led many micro-enterprises to underutilize digital technologies in their marketing activities. Previous studies indicate that digital marketing plays a crucial role in promoting MSME growth and sustainability. Further findings confirm that the adoption of digital marketing significantly enhances the sustainable growth of MSMEs in Indonesia [3]. Factors such as a positive attitude toward technology and the intention to use digital marketing have been shown to exert a direct impact on business growth.

The challenges faced by the Kayu Raja Village community are not limited to the lack of digitalization facilities but also include the absence of an integrated marketing information system. The manual processes involved in the sale of oil palm midrib products make it difficult for the community to manage sales data, promotions, and customer relationships. In fact, web-based information systems can assist business actors in expanding market networks and improving marketing efficiency. Research indicates that the key success factors of digital marketing among Indonesian MSMEs lie in continuous training and mentoring [4]. With the availability of a simple and easily accessible web-based information system, rural communities can gain practical understanding of effective digital marketing strategies without relying on external parties.

Meanwhile, research and implementation of web-based marketing systems have predominantly focused on the food, beverage, and retail sectors, whereas non-food agricultural by-product–based commodities such as oil palm midribs have received limited scholarly attention [5]. Moreover, many digital systems are designed without adequately considering the capabilities of rural users, making their sustained implementation challenging. Therefore, there is a need for a web-based marketing information system that is simple, user-friendly, and aligned with the characteristics of the Kayu Raja Village community.

In addition to technological factors, government support and the development of digital infrastructure also play a crucial role. A study emphasizes that government digital transformation programs must be accompanied by efforts to enhance community digital literacy in order to optimize infrastructure utilization [6]. Without local innovations such as information systems tailored to community needs, digitalization will not be implemented effectively.

Accordingly, the author is motivated to conduct research on the design and implementation of a web-based marketing information system, given the high urgency of this study. The development and implementation of such a system are intended as a solution to overcome the limitations of traditional marketing practices currently employed in Kayu Raja Village. From an academic perspective, this research contributes to the advancement of the literature on the utilization of information technology for rural community economic empowerment. From a practical standpoint, the resulting system is expected to assist the Kayu Raja Village community in expanding market reach, improving the efficiency of marketing data management, and supporting increased income as well as the economic independence of the rural community.

2 LITERATURE REVIEW

Previous studies have demonstrated that web-based marketing information systems improve product promotion, operational efficiency, and structured data management among MSMEs [7]. These systems enable real-time access to product and transaction information, thereby enhancing business performance and expanding promotional reach beyond geographical constraints. Moreover, prior research emphasizes the importance of applying structured system development methodologies, particularly the Software Development Life Cycle (SDLC), to ensure alignment between system design and user requirements.

Despite these contributions, most studies concentrate on general retail sectors or urban-based enterprises, with limited attention to rural community contexts. In addition, although digital marketing adoption has been widely discussed, fewer studies focus on the systematic design and evaluation of web-based marketing information systems tailored to small-scale rural enterprises, especially those operating in agricultural by-product commodities.

From a marketing strategy perspective, the integration of web-based systems with Customer Relationship Management (CRM) approaches has been shown to enhance customer engagement and promotional effectiveness [8]. Similarly, digitalization

in the agricultural sector contributes to improved distribution efficiency and transaction processes[9]. However, these studies primarily emphasize marketing outcomes rather than examining the suitability of system design for users with limited technological capabilities in rural areas. In terms of system development approaches, various methodologies such as Waterfall, Rapid Application Development (RAD), Object-Oriented Design, and Software as a Service (SaaS) have been applied in previous research [10]. While SaaS-based models offer accessibility advantages for MSMEs, comparative evaluations of development approaches in rural village-level implementations remain limited. Furthermore, empirical studies focusing specifically on oil palm midrib product marketing within rural communities are still underrepresented in the literature. Therefore, this study addresses these gaps by designing, implementing, and evaluating a web-based marketing information system using the SDLC approach, specifically tailored to the operational characteristics and technological capabilities of the Kayu Raja Village community.

3 RESEARCH METHODS

This study employs the Software Development Life Cycle (SDLC) approach using the Waterfall model in the design and implementation of a web-based marketing information system for oil palm midrib products for the Kayu Raja Village community. The Waterfall model was selected because it provides systematic and sequential development stages, in which each phase must be completed before proceeding to the next, thereby facilitating control and evaluation of the system development process [11].

The Waterfall model is considered appropriate for this research because system requirements were clearly identified from the initial stage through observation and interviews, and the system developed has low to moderate complexity [12].

The research stages were conducted sequentially in accordance with the Waterfall model, as follows:

1. Problem Identification and Data Collection
   The initial stage aimed to understand the condition of oil palm midrib product marketing, which is still carried out traditionally. Data collection was conducted through:
   a. Observation, namely direct observation of marketing processes and community sales activities to obtain factual field data.
   b. Interviews, conducted in a semi-structured manner with business actors and village officials to explore information regarding marketing constraints and expected system requirements [13].
   c. Literature review, by examining relevant books and scientific journals related to information systems, digital marketing, and web-based system development as the theoretical foundation of the study [14].

2. System Analysis
   System analysis was carried out to transform collected data into system requirements to be developed. This analysis aimed to identify problems in the existing system and formulate more effective solutions for a new system [19]. The system analysis includes:
   a. Functional requirements, namely the main system functions such as providing a product catalog, enabling online product ordering, and managing product and transaction data by the admin [15].
   b. Non-functional requirements, including aspects of usability, data security, system performance, interface design, and accessibility across various devices [16].

3. System Design
   The system design stage aimed to translate system requirements into technical and visual designs. The design was conducted using Unified Modeling Language (UML) to clearly model system structure and workflow [17]. The UML diagrams used include Use Case Diagrams, Activity Diagrams, Sequence Diagrams, and Class Diagrams to ensure that all functional system requirements were well defined prior to implementation.

4. System Implementation
   Implementation was carried out by applying the design results into a web-based marketing information system. The website was used as a promotional medium by displaying product information, prices, and ordering contacts. This stage aimed to realize the system design into a functional system that could be directly used by the community [18].

5. System Testing
   System testing was conducted using functional testing to ensure that all system features operated according to user requirements. The testing involved the Kayu Raja Village community as users to assess usability and clarity of information provided by the system [19].

6. Evaluation and Maintenance
   The evaluation stage was carried out to assess system effectiveness based on testing results and user feedback. System maintenance includes

content updates, information corrections, and ensuring continuous system accessibility for sustainable use [20].

4 RESULTS AND DISCUSSION

A. System Design

Overview The development of the web-based marketing information system employed the Software Development Life Cycle (SDLC) using the Waterfall model to ensure a structured and systematic development process. Recent studies emphasize that SDLC remains relevant for small-scale information systems requiring clear requirement definition and sequential development stages [21]. Unified Modeling Language (UML) was utilized to model system requirements through Use Case, Activity, and Sequence Diagrams. Contemporary research highlights that UML modeling enhances requirement clarity and reduces design ambiguity, particularly in community-based system development projects [22].

B. System Implementation

The system was implemented as a web-based application with two primary actors: administrator and customer. Core functionalities include product catalog management, online ordering, transaction recording, and reporting. Recent studies confirm that web-based marketing systems significantly improve accessibility and real-time data management for rural MSMEs, particularly in areas with limited physical market access [23]. The use of structured databases supports data consistency and improves operational efficiency in small enterprises transitioning from manual processes to digital platforms [24]. The interface design followed usability-oriented principles to accommodate users with limited digital literacy. Current research indicates that simplicity, clarity, and minimal navigation complexity are essential for technology adoption in rural communities [25].

C. Functional Testing

Results Functional testing was conducted using the black-box testing approach to verify system compliance with predefined functional requirements. Black-box testing continues to be widely used in validating web-based systems due to its effectiveness in assessing user-level functionality [26]. Testing results demonstrated that all primary features registration, authentication, product management, ordering, and reporting operated according to specifications. No major functional errors were identified during testing. These findings align with recent research indicating that structured validation processes improve system reliability and user trust in newly implemented digital platforms [22],[26].

D. Usability Evaluation

User Acceptance Testing (UAT) was conducted through direct trials involving members of the Kayu Raja Village community. Usability is defined as the extent to which a system can be used effectively, efficiently, and satisfactorily within a specific context [27]. Based on user feedback, the system was perceived as easy to understand and operate. The digital catalog improved information accessibility, and structured transaction records reduced manual recording errors. Similar findings have been reported in recent studies on rural digitalization, where usability directly influences technology adoption and sustained usage [28].

E. System Performance and Adoption Impact

From a performance perspective, the system demonstrated stable functionality under small-scale operational conditions. Data processing and retrieval were executed without observable delays. Research in recent years shows that even basic web-based systems significantly enhance transparency and transaction efficiency in rural enterprises [24] [28]. Before implementation, marketing activities relied on intermediaries and manual recording. After implementation, product promotion became digitally accessible, and transaction data were systematically managed. Digital transformation studies published in the past five years indicate that structured marketing information systems contribute to improved competitiveness, wider market reach, and increased economic resilience among rural MSMEs [24] [29].

Although long-term sustainability depends on continuous digital literacy development, initial adoption results in this study indicate positive acceptance among local users. Based on the results of the research on the implementation and testing of the web-based marketing information system for oil palm midrib products developed in this study, the discussion is focused on the system's functional performance, its ease of use for the community, and its contribution to supporting the digital expansion of product marketing in Kayu Raja Village.

4.1 System Design

In this study, Unified Modeling Language (UML) was employed to design the processes of the web-based marketing information system for oil palm midrib products, utilizing three types of diagrams, namely:

1. Use Case Diagram
2. Activity Diagram
3. Sequence Diagram

I. Use Case Diagram

The Use Case Diagram in this study illustrates the interaction between two primary actors administrator and customer and the core functionalities of the web-based marketing information system. The administrator manages product data, monitors transactions, and generates sales reports, while the customer browses products and places orders.

The structured separation of roles reflects the transformation from informal and manual marketing practices to a digitally managed workflow. Clear role definition in system modeling has been shown to improve requirement clarity and reduce functional ambiguity in small-scale information system development [30].

By explicitly defining system boundaries and actor responsibilities, the use case model enhances transaction transparency and strengthens data accountability. Recent studies indicate that structured digital workflows in rural MSMEs contribute to improved operational control and reduced reliance on intermediaries [31] [32].

Furthermore, aligning system functions with actual business processes is critical in community-based system implementation, particularly where users possess limited technological familiarity [25]. Therefore, the use case design in this research not only model's system interaction but also serves as a strategic foundation for supporting autonomous digital marketing practices in Kayu Raja Village. In this study, the system comprises two main actors, namely:

a) Admin / Seller
   1. Register
   2. Log in to the system
   3. Manage product data (add, edit, delete)
   4. Manage orders and transactions
   5. View sales reports
   6. Log out

b) User / Buyer (General Public)
   1. Register
   2. Log in to the system
   3. View the product catalog and prices
   4. Add desired products to the shopping cart
   5. Purchase selected products
   6. Log out.

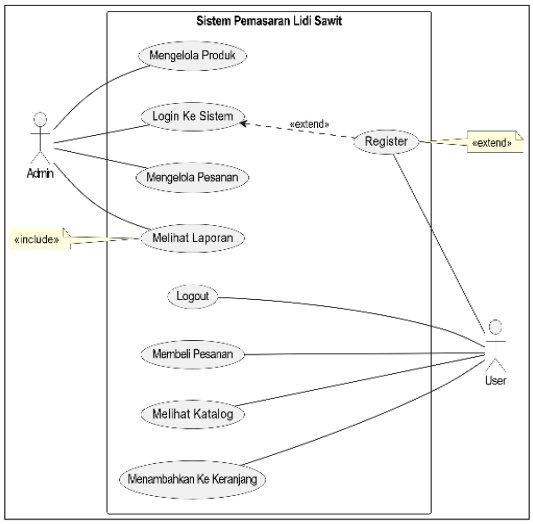


Figure 1. Use Case Diagram

II. Activity Diagram

The Activity Diagram models the operational workflow of the web-based marketing information system, particularly in product data management, promotional activities, and user interaction in accessing and ordering products. The modeled workflow begins from user authentication, proceeds to product browsing and order submission, and ends with transaction recording by the administrator.

Rather than merely representing procedural steps, the activity model formalizes previously informal and unstructured marketing practices observed in Kayu Raja Village. By translating manual communication-based transactions into sequential and logically controlled digital processes, the system reduces operational ambiguity and improves traceability. Structured workflow modeling has been shown to improve system clarity, minimize implementation errors, and enhance process control in small-scale information systems [33]. In the context of rural MSMEs, clear digital process mapping contributes to operational efficiency and supports gradual digital adoption [23]. Therefore, the Activity Diagram in this study does not only describe system flow but also represents the transformation of traditional marketing mechanisms into a controlled and accountable digital workflow. The processes occurring within this system include the following:

A. Admin / Seller Activity Diagram Process
   a. The admin accesses the web-based marketing information system page.
   b. The admin logs in by entering a username and password.

c. The system validates the login credentials.
   a) If the data are valid, the system displays the admin dashboard.
   b) If the data are invalid, the system displays an error message and returns to the login page.
d. The admin manages oil palm midrib products, including:
   a) Add product data
   b) Edit product data
   c) Delete product data
e. The system stores every data modification in the database.
f. The admin manages marketing information and system content.
g. The admin completes the activities and logs out.
h. The system terminates the admin session.

B. User / Buyer Activity Diagram Process
   a. The buyer accesses the marketing information system website.
   b. The system displays the main page.
   c. The buyer views the list of oil palm midrib products.
   d. The buyer selects a product to view detailed information.
   e. The system displays the product details and marketing information.
   f. The buyer obtains the required information.
   g. The buyer exits or closes the website page.
   h. The process ends.

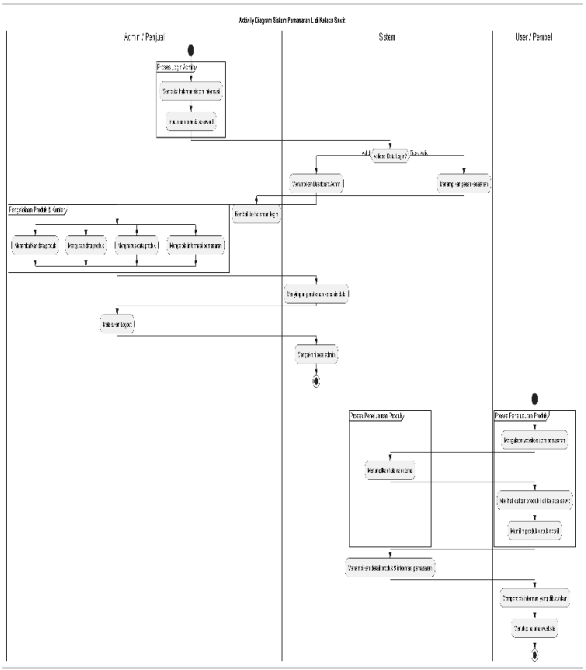


Figure 2. Activity Diagram

III. Sequence Diagram

The Sequence Diagram in this study models the transaction flow within the web-based marketing information system, particularly during product ordering and product data management processes. The interaction begins when a user submits a request through the interface, such as browsing products or placing an order. The system processes the request by validating the input data before forwarding it to the application logic layer.

After validation, the system communicates with the database to retrieve product information or to store transaction records. The database integration ensures that product data and transaction histories are stored consistently and can be accessed in real time. Once the database operation is completed, the system generates a response that is returned to the user interface, confirming the success or failure of the transaction.

In the product management process, the administrator submits product updates, which are processed and synchronized directly with the database. This structured interaction between the user interface, application layer, and database improves data integrity and reduces the risk of logical inconsistencies. Recent studies emphasize that interaction-based modeling enhances system reliability by clearly defining message flow and transactional dependencies in web-based systems [34]. Furthermore, structured transaction sequencing supports traceability and minimizes implementation errors in small-scale information

systems [23]. By formalizing transaction flow and database integration, the Sequence Diagram in this research contributes to improving operational transparency and ensuring accountable digital marketing processes within the rural enterprise context.

The sequence diagram for product ordering includes the following:

A. Objects Involved:
   1. User
   2. Web Server
   3. Database
   4. Admin

B. Process Flow:
   1. The user opens the product catalog page and sends a request to the server.
   2. The server receives the request and retrieves product data from the database.
   3. The database returns the product data to the server.
   4. The server displays the product data to the user.
   5. The user selects a product and completes the order form.
   6. The server receives the order data from the user and stores it in the database.
   7. The admin logs into the system and accesses the order data.
   8. The system displays new order data to the admin.
   9. The admin updates the order status to "processed."
   10. The server saves the updated order status to the database.
   11. The system sends a notification to the user that the order has been received and is being processed.
   12. The sequence diagram process ends.

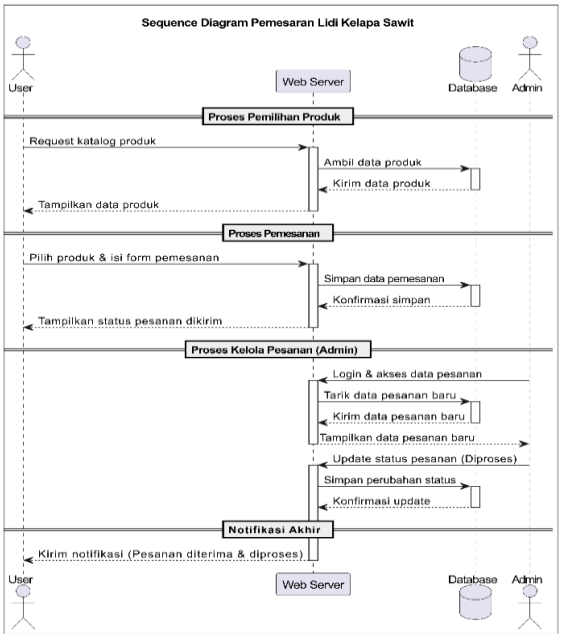


Figure 3. Sequence Diagram

This process indicates that the system operates in real time, allowing administrators and users to be interconnected through the web server.

With the application of UML based process design:

1. Researchers can ensure that every functional requirement is represented in the system design.
2. System documentation becomes clearer and easier to understand for the development team and other stakeholders.
3. The implementation phase becomes more efficient because a structured workflow reference is already available.
4. System testing is facilitated, as each scenario has been modeled in UML.

4.2 System Design and Implementation

The web-based marketing information system was implemented to operationalize the structured models developed in the design phase. The system integrates a web application layer with a relational database to support product management, transaction processing, and digital marketing activities.

Rather than merely providing an online interface, the implemented system formalizes previously informal marketing practices into a controlled digital environment. The product catalog feature enables structured presentation of product information, reducing dependency on oral communication and intermediaries. The ordering module supports systematic transaction recording, ensuring that each purchase request is validated, processed, and stored within the database.

The integration between the application layer and the database ensures data consistency and real-time synchronization between product updates and user access. This integration minimizes manual recording errors and enhances traceability of transactions, which was identified as a major limitation in traditional practices. From an operational perspective, the implementation demonstrates how structured web-based systems can transform rural marketing mechanisms into accountable digital processes. Similar findings have been reported in recent studies on digital transformation in rural MSMEs, where system integration improves efficiency and market accessibility[35] [36].

Therefore, the implementation results do not merely represent technical deployment but illustrate the practical realization of digital process restructuring within the Kayu Raja Village enterprise context.

A. User (Buyer) Implementation

The user component is designed to facilitate prospective buyers in accessing product information and placing orders online.

The following presents the implementation of features on the user side:

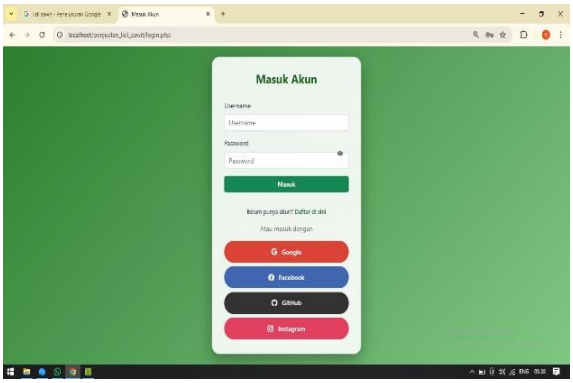


Figure 4. Login Page

Figure 4 illustrates the login page (“Account Sign In”) of the oil palm midrib sales web application. This page provides fields for username and password, a Sign In button, and an option to register a new account. It also offers login via third-party accounts such as Google, Facebook, GitHub, and Instagram. This page functions as the authentication gateway for users (buyers and sellers) before accessing the system.

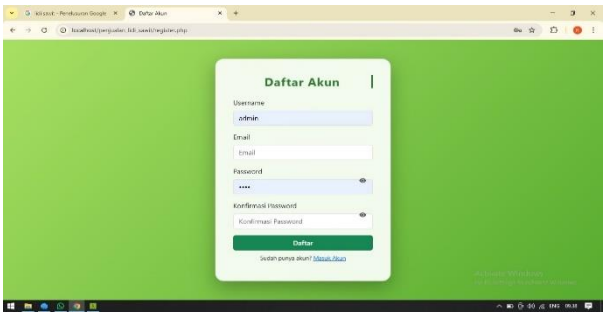


Figure 5. Account Registration Page (Register)

Figure 5 illustrates the account registration page (“Create Account”) of the oil palm midrib sales web application. This page functions to register new users by providing input fields for username, email, password, and password confirmation. A “Register” button is available to save user data, along with a Sign In link for users who already have an account.

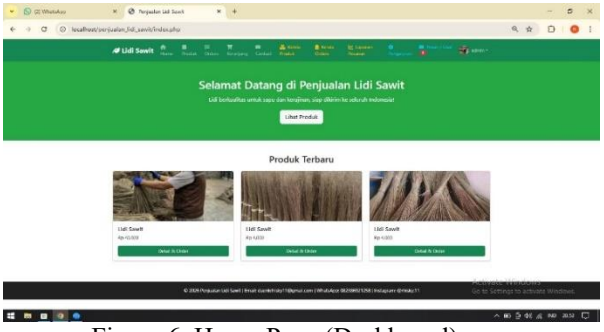


Figure 6. Home Page (Dashboard)

Figure 6 illustrates the home page (dashboard) of the oil palm midrib sales system, which is equipped with a comprehensive navigation menu, including product management, orders, reports, settings, and messaging features. This page functions as the central hub for control and information by displaying a system welcome message and a list of the latest products to support management and transaction activities.

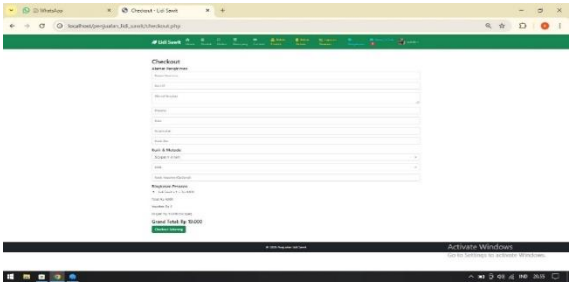


Figure 7. Product Catalog Page

Figure 7 presents the product catalog page of the oil palm midrib sales system. Figure 4 illustrates the implementation of the product catalog module, which operationalizes structured product data management and supports digital accessibility for customers.

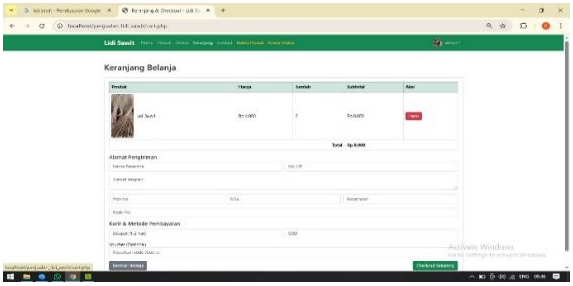


Figure 8. Shopping Cart Page

Figure 8 illustrates the shopping cart page of the oil palm midrib sales web application. This page functions to display and manage items selected for purchase, including product information, price, quantity, subtotal, and total payment. It also provides a shipping address form, courier and payment method options, and a checkout button. The page represents the transition from product selection to the transaction process.

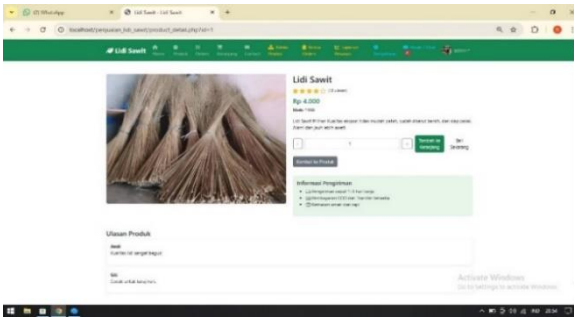


Figure 9. Checkout Page

Figure 9 illustrates the checkout page of the oil palm midrib sales system. This page is used to complete the transaction process by providing a shipping address form, courier and payment method options, and an order summary displaying the total cost and grand total.

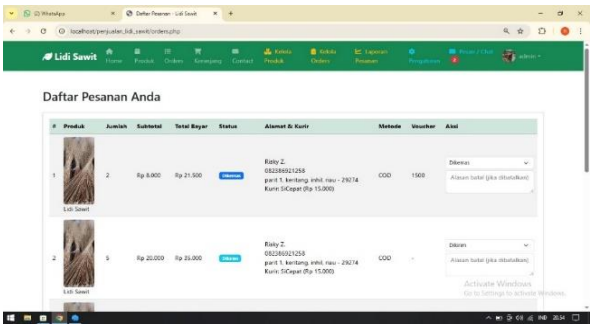


Figure 10. Order List Page

Figure 10 presents the order list page of the oil palm midrib sales system. This page displays transaction information in tabular form, including product, quantity, subtotal, total payment, order status, shipping address, payment method, and action options. The interface enables systematic monitoring and management of order progress.

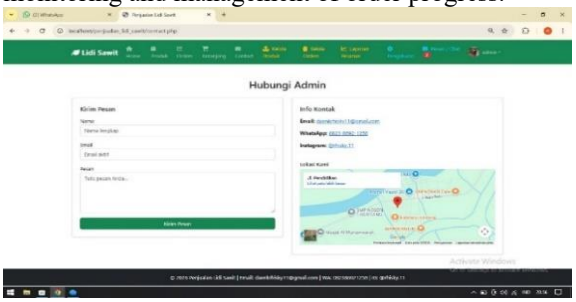


Figure 11. Contact Page (Contact Admin)

Figure 11 presents the contact page (“Contact Admin”) of the oil palm midrib sales system. This page provides a message submission form for users as well as admin contact information, including email, WhatsApp, and Instagram, accompanied by a location map. This feature facilitates direct communication and customer service within the system.

B. Admin (Seller) Implementation

The admin component is used by business actors or system administrators to manage product data and process incoming orders. The following presents the implementation results on the admin side:

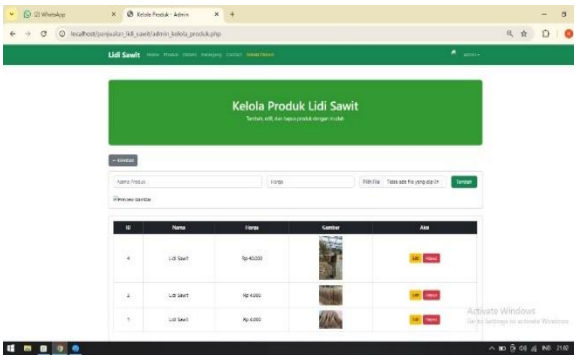


Figure 12. Dashboard Page (Admin)

Figure 12 presents the admin dashboard page for order management in the oil palm midrib sales system. This page displays detailed transaction information, including product data, order quantity, total payment, shipping address, payment method, and order status that can be updated. This feature functions as a systematic control mechanism for managing the ordering process from the seller's perspective.

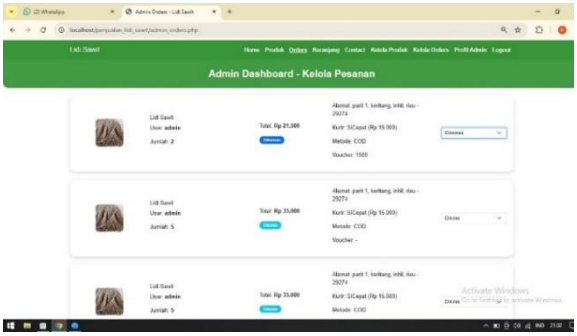


Figure 13. Product Management Page

Figure 13 presents the product management page for the seller (admin) in the oil palm midrib sales system. This page is used to add, edit, and delete product data, which are displayed in a table containing product name, price, image, and action buttons.

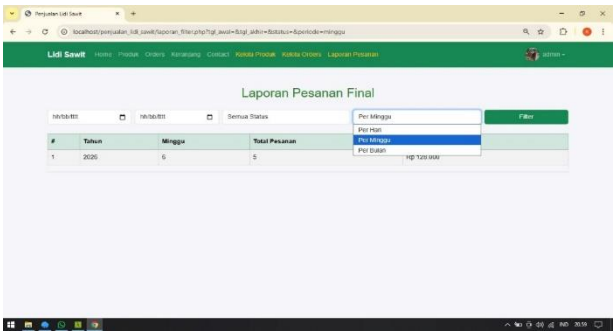


Figure 14. Order Report Page

Figure 14 presents the order report page of the oil palm midrib sales system on the admin side. This page provides data filtering features based on date range, order status, and reporting period (daily, weekly, monthly), and displays a summary of order quantities in tabular form. This feature supports systematic analysis of transaction performance.

C. Database Implementation

The database is used to store all information related to the system, including product data, admin data, buyer data, and order transaction data. The database system is designed to ensure data security, consistency, and integrity. The database is implemented in MySQL and created through the MySQL terminal in the Laragon environment.

5 CONCLUSION

This study developed and implemented a web-based marketing information system for oil palm midrib products to address the limitations of traditional marketing practices in Kayu Raja Village. The system was designed using the Software Development Life Cycle (SDLC) with the Waterfall model, ensuring a structured development process aligned with identified user requirements. Based on functional testing using the black-box method, all core features including product management, online ordering, and transaction recording operated according to system specifications. Furthermore, User Acceptance Testing (UAT) conducted with community members indicated that the system was perceived as easy to use and understandable, particularly in accessing product information and recording transactions.

The implementation results demonstrate that the system successfully formalizes previously manual marketing processes into a structured digital workflow supported by database integration. This transition improves transaction traceability and data organization under small-scale operational conditions. Although long-term impact measurement requires further observation, initial adoption results indicate positive

user acceptance within the rural community context. Therefore, this research contributes not only practically through system deployment but also academically by demonstrating the applicability of structured web-based marketing systems in supporting digital transformation initiatives for rural micro-enterprises.

REFERENCES

[1] L. Anatan dan Nur, "Micro, Small, and Medium Enterprises' Readiness for Digital Transformation in Indonesia," *Economies*, vol. 11, no. 6, hlm. 156, 2023, doi: 10.3390/economies11060156.

[2] J. Wiratama, S. F. Wijaya, H. Budiyanto, R. Setiadi, dan A. L. Raphael, "Studi Awal Pemanfaatan Teknologi Digital oleh UMKM UP2K Kenanga di Desa Curug Sangereng: Analisis Kesiapterapan Digitalisasi Pemasaran," *I-Com*, vol. 5, no. 3, hlm. 1504–1514, Sep 2025, doi: 10.70609/i-com.v5i3.8005.

[3] P. Adi *dkk.*, "Pemberdayaan Masyarakat melalui Pelatihan Digital Marketing pada UMKM Produk Pertanian di Desa Bodag, Madiun, Jawa Timur," *prima j comm. empw. serv.*, vol. 6, no. 2, hlm. 126, Agu 2023, doi: 10.20961/prima.v6i2.65249.

[4] Oki Dermawan, Amara Silvi Cardona, Elfa Diana, dan Suci Rahmatu Wahidah, "Digital Strategy for MSMEs: Marketing Optimization of Woven Bags," *communautaire*, vol. 3, no. 1, hlm. 40–51, Agu 2024, doi: 10.61987/communautaire.v3i1.385.

[5] S. Ramadhani, V. Sihombing, dan G. Juni Yanris, "Web-based Palm Oil Seedling Sales Information System (Case Study: CV. XYZ)," *International Journal of Science, Technology & Management*, vol. 6, no. 1, hlm. 257–262, Jan 2025, doi: 10.46729/ijstm.v6i1.1213.

[6] E. N. Dewi dan S. Sumarno, "Web-based Product Marketing Information System for MSMEs Tanggulangin Wallet and Bag Manufacturers," *PELS*, vol. 4, Jul 2023, doi: 10.21070/pels.v4i0.1417.

[7] A. Mulyaningsih dan M. A. Gustalika, "Design and Development of a Website for Micro, Small, and Medium Enterprises (MSMEs) of Pekunden Pottery Craft Using the Extreme Programming Method," *IJOEM Indones. J. E-learning Multimed.*, vol. 4, no. 3, hlm. 248–264, Okt 2025, doi: 10.58723/ijoem.v4i3.497.

[8] Wizra Aulia, Stefani Hardiyanti Putri, dan Imelda Juniarta Emin, "Penerapan Sistem Informasi Pemasaran Toko Oleh-Oleh Makanan Khas Danau Maninjau Berbasis WEB," *Neptunus*, vol. 3, no. 3, hlm. 289–300, Agu 2025, doi: 10.61132/neptunus.v3i3.1035.

[9] H. Subhi Malallah dan M. Bahjat Abdulrazzaq, "Web-Based Agricultural Management Products for Marketing System: Survey," *ACAD J NAWROZ UNIV*, vol. 12, no. 2, hlm. 49–62, Apr 2023, doi: 10.25007/ajnu.v12n2a1532.

[10] P. Setiaji dan A. Setiawan, "Software As A Service (Saas) As A Handicraft Portal Bag," *J. Phys.: Conf. Ser.*, vol. 1430, no. 1, hlm. 012047, Jan 2020, doi: 10.1088/1742-6596/1430/1/012047.

[11] M. Ridwan, I. Fitri, dan B. Benrahman, "Rancang Bangun Marketplace Berbasis Website menggunakan Metodologi Systems Development Life Cycle (SDLC) dengan Model Waterfall," *jtik*, vol. 5, no. 2, hlm. 173–184, Jun 2021, doi: 10.35870/jtik.v5i2.209.

[12] M. Al Fajar, M. H. Dar, dan R. Rohani, "Application of Waterfall model in development of family planning participants information system," *SinkrOn*, vol. 7, no. 2, hlm. 679–686, Mei 2022, doi: 10.33395/sinkron.v7i2.11387.

[13] D. F. Arum, J. M. Peranginangin, dan I. Purwidyaningrum, "PIECES Method for Evaluating the Management Information System in the Pharmacy Installation of Hospital X," *FJMR*, vol. 4, no. 4, hlm. 1811–1824, Apr 2025, doi: 10.55927/fjmr.v4i4.161.

[14] S. Ravi dan S. R. C. Rajasekaran, "A Perspective of Digital Marketing in Rural Areas: a Literature Review," *J. Professional Business Review*, vol. 8, no. 4, hlm. e01388, Mar 2023, doi: 10.26668/businessreview/2023.v8i4.1388.

[15] Ms. S. Prabakar, "E-COMMERCE WEBSITE USING NEXT.JS," *IJSREM*, vol. 09, no. 04, hlm. 1–9, Apr 2025, doi: 10.55041/IJSREM46353.

[16] F. M. Ghabban dan W. Ghaban, "Analyzing Non-Functional Requirements of Mobile Applications: Usability, Reliability, Performance, and Supportability," *JCC*, vol. 13, no. 08, hlm. 260–279, 2025, doi: 10.4236/jcc.2025.138013.

[17] M. M. I. Molla, J. Ahmad, dan W. M. N. Wan Kadir, "A Comparison of Transforming the User Stories and Functional Requirements into UML Use Case Diagram," *Int J Innov Comp*, vol. 14, no. 1, hlm. 29–36, Mei 2024, doi: 10.11113/ijic.v14n1.463.

[18] C. N. Febriyanti, D. Puspita, M. R. Badruzzaman, M. F. A. Fatar, dan M. G. Ruba', "Pembuatan Website Katalog Produk Kuliner Sebagai Media Promosi dan Pemasaran UMKM Sate Ayam Ponorogo Pak Ici," *Welfare*, vol. 1, no. 3, hlm. 582–588, Sep 2023, doi: 10.30762/welfare.v1i3.784.

[19] A. D. Samala, M. Ricci, C. J. Angel Rueda, L. Bojic, F. Ranuharja, dan W. Agustiarmi,

"Exploring Campus through Web-Based Immersive Adventures Using Virtual Reality Photography: A Low-Cost Virtual Tour Experience," *Int. J. Onl. Eng.*, vol. 20, no. 01, hlm. 104–127, Jan 2024, doi: 10.3991/ijoe.v20i01.44339.

[20] F. Asrin dan G. V. Utami, "Implementing Website-Based School Information Systems in Public Elementary Schools Using Waterfall Model," *J. Inf. Syst. Informatics*, vol. 5, no. 2, hlm. 590–614, Mei 2023, doi: 10.51519/journalisi.v5i2.495.

[21] S. Pargaonkar, "A Comprehensive Research Analysis of Software Development Life Cycle (SDLC) Agile & Waterfall Model Advantages, Disadvantages, and Application Suitability in Software Quality Engineering," *IJSRP*, vol. 13, no. 8, hlm. 120–124, Agu 2023, doi: 10.29322/IJSRP.13.08.2023.p14015.

[22] A. Ferrari, S. Abualhaija, dan C. Arora, "Model Generation with LLMs: From Requirements to UML Sequence Diagrams," dalam *2024 IEEE 32nd International Requirements Engineering Conference Workshops (REW)*, Jun 2024, hlm. 291–300. doi: 10.1109/REW61692.2024.00044.

[23] N.-J. Shih dan T.-Y. Chen, "Assessment of Lantern Festivals by Government Procurements," *Sustainability*, vol. 14, no. 9, hlm. 5147, Apr 2022, doi: 10.3390/su14095147.

[24] P. Tasatanattakool, K. Pradit, P. Nilsook, dan P. Wannapiroon, "Digital transformation of organizations: Intelligence financial management system," *ijirss*, vol. 8, no. 1, hlm. 773–783, Jan 2025, doi: 10.53894/ijirss.v8i1.4422.

[25] X. Wan dan Y. Ren, "Exercise identity, exercise behavior and mobile phone addiction: A cross-sectional mediation study with a sample of rural left-behind children in China," *Heliyon*, vol. 9, no. 4, hlm. e14953, Apr 2023, doi: 10.1016/j.heliyon.2023.e14953.

[26] T. Li *dkk.*, "A Survey on Web Application Testing: A Decade of Evolution," 2024, *arXiv*. doi: 10.48550/ARXIV.2412.10476.

[27] W. Wulandari, N. Nofiyani, dan H. Hasugian, "USER ACCEPTANCE TESTING (UAT) PADA ELECTRONIC DATA PREPROCESSING GUNA MENGETAHUI KUALITAS SISTEM," *IlmuKomputer*, vol. 4, no. 1, hlm. 20–27, Mar 2023, doi: 10.24127/ilmukomputer.v4i1.3383.

[28] C. Liang, M. Umar, F. Ma, dan T. L. D. Huynh, "Climate policy uncertainty and world renewable energy index volatility forecasting," *Technological Forecasting and Social Change*, vol. 182, hlm. 121810, Sep 2022, doi: 10.1016/j.techfore.2022.121810.

[29] J. Nako, C. Psychalinos, A. S. Elwakil, dan D. Jurisic, "Design of Higher-Order Fractional Filters With Fully Controllable Frequency Characteristics," *IEEE Access*, vol. 11, hlm. 43205–43215, 2023, doi: 10.1109/ACCESS.2023.3271863.

[30] M. Waseem, P. Liang, M. Shahin, A. Di Salle, dan G. Márquez, "Design, monitoring, and testing of microservices systems: The practitioners' perspective," *Journal of Systems and Software*, vol. 182, hlm. 111061, Des 2021, doi: 10.1016/j.jss.2021.111061.

[31] N.-J. Shih dan T.-Y. Chen, "Assessment of Lantern Festivals by Government Procurements," *Sustainability*, vol. 14, no. 9, hlm. 5147, Apr 2022, doi: 10.3390/su14095147.

[32] C. Liang, M. Umar, F. Ma, dan T. L. D. Huynh, "Climate policy uncertainty and world renewable energy index volatility forecasting," *Technological Forecasting and Social Change*, vol. 182, hlm. 121810, Sep 2022, doi: 10.1016/j.techfore.2022.121810.

[33] M. Morales-Sandoval, R. De-La-Parra-Aguirre, H. Galeana-Zapien, dan A. Galaviz-Mosqueda, "A Three-Tier Approach for Lightweight Data Security of Body Area Networks in E-Health Applications," *IEEE Access*, vol. 9, hlm. 146350–146365, 2021, doi: 10.1109/ACCESS.2021.3123456.

[34] M. Morales-Sandoval, R. De-La-Parra-Aguirre, H. Galeana-Zapien, dan A. Galaviz-Mosqueda, "A Three-Tier Approach for Lightweight Data Security of Body Area Networks in E-Health Applications," *IEEE Access*, vol. 9, hlm. 146350–146365, 2021, doi: 10.1109/ACCESS.2021.3123456.

[35] S. Ullah, R. Attah-Boakye, K. Adams, dan G. Zaefarian, "Assessing the influence of celebrity and government endorsements on bitcoin's price volatility," *Journal of Business Research*, vol. 145, hlm. 228–239, Jun 2022, doi: 10.1016/j.jbusres.2022.01.055.

[36] F. A. Handoko, Tundo, F. I. Sarky, dan R. Wijaya, "Transformasi Digital: Pengembangan Sistem Informasi Penjualan Berbasis Web pada UMKM Barokah Jaya Cell di Bekasi dengan Pendekatan UI/UX," *jtik*, vol. 9, no. 3, hlm. 1150–1157, Apr 2025, doi: 10.35870/jtik.v9i3.3820.

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 February 2026

Comparison of Chi-Square and Information Gain Feature Selection Methods for Support Vector Machine-Based Sentiment Analysis
(Case Study: Vidio Application Reviews on Google Play Store)

Vitta Margaret Sinambela, Herlina Napitupulu, Nurul Gusriani
Mathematics, Faculty of Mathematics and Natural Sciences, Padjadjaran University
Email: vitta21001@mail.unpad.ac.id; herlina@unpad.ac.id; nurul.gusriani@unpad.ac.id.

Accepted:	Accepted After Revision:	Published:
4 August 2025	24 February 2026	27 February 2026

Abstract

Vidio is a local streaming platform that dominates the Indonesian market, but still faces challenges in improving user satisfaction as reflected by its 3.5 rating. To enhance the application, user experience insights are needed, which can be identified through sentiment analysis. This study aims to analyze the sentiment of Vidio application user reviews and compare the performance of the Support Vector Machine model using Chi-Square and Information Gain feature selection. The dataset comprises 4,670 reviews collected from July 01 to November 30, 2024. Model evaluation utilizes Balanced Accuracy metrics optimized through hyperparameter tuning to ensure fair assessment on imbalanced data. The experimental results demonstrate that Chi-Square feature selection yields the optimal performance, achieving a peak Balanced Accuracy of 94.78%. Significantly, this result was attained using a computationally efficient Linear Kernel ($\mathrm{C} = 1$). In contrast, the Information Gain method yielded a lower Balanced Accuracy of 94.20% despite utilizing a complex Polynomial Kernel ($\mathrm{C} = 1, \mathrm{r} = -1, \mathrm{p} = 3, \gamma = 0.1$). These findings conclude that Chi-Square provides a superior trade-off between classification accuracy and model complexity, offering a more robust solution for sentiment analysis.

Keywords: Sentiment Analysis, Support Vector Machine, Chi-Square, Information Gain, Vidio.

1 INTRODUCTION

According to [1], the number of video streaming application user in Indonesia continues to increase, reflecting the growing role of streaming platforms in digital life. The local platform that dominates this market in Indonesia is Vidio [2]. The Vidio application has a rating of 3.5 from approximately 670 thousand users, indicating the need for further analysis to identify aspects that require improvement. To process a large volume of reviews, sentiment analysis is conducted to classify user reviews and identify complaint patterns as well as frequently highlighted negative aspects.

According to [3], classification using the Support Vector Machine (SVM) method is more accurate than other classifiers, performs well on nonlinear data, and has a low risk of overfitting. Machine learning methods can be combined with feature selection to reduce the number of features and improve model accuracy. A comparison of Chi-Square, Information Gain, and Principal Component Analysis feature selection methods found that Chi-Square and Information Gain achieved the highest scores [4]. The performance of the SVM model is influenced by hyperparameters. Hyperparameter tuning such as kernel function (k), regularization constant (C), gamma (γ), and polynomial degree (p) can improve SVM performance [5].

Sentiment analysis research on the Vidio application has been conducted previously. Applied the KNN method and obtained 70% accuracy through manual calculation and 50% accuracy using RapidMiner tools [6]. Decision Tree method and achieved the best accuracy of 97.3% with an 80:20 data split ratio [7]. Applied KNN and Naïve Bayes Classifier (NBC), showing that KNN performed better with an accuracy of 93%, recall of 88%, precision of 93%, and F1-Score of 90% compared to NBC with an accuracy of 81%, recall of 91%, precision of 81%, and F1-Score of 86% [8].

In previous sentiment analysis studies on the Vidio application, the review data used was imbalanced. To address this issue, the most appropriate evaluation metric is balanced accuracy [9]. However, balanced accuracy results have not been presented, and feature selection has not been applied, meaning all features were

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

used without considering their relevance to model performance. Therefore, this study applies the SVM classification method and feature selection by comparing the Chi-Square and Information Gain feature selection methods on Vidio user reviews. In addition, hyperparameter tuning based on balanced accuracy is performed using Grid Search to improve model performance.

2 LITERATURE REVIEW

2.1 Feature Selection

Feature selection is a machine learning technique used to select the minimum number of features required to represent data accurately, making computation more efficient. The use of relevant feature selection can improve accuracy, shorten the machine learning process by reducing dimensionality and removing noise, and produce simpler concepts [8]. In this study, a comparison of feature selection effectiveness is conducted. The feature selection methods compared for sentiment analysis are Chi-Square and Information Gain as individual feature selection techniques.

2.1.1 Chi-Square Feature Selection

According [10], Chi-Square feature selection aims to select features using Chi-Square statistical values to measure the dependency between words and their classes. The Chi-Square test function for a word against a category is determined using equation (1) [11],

$$\chi^2(t_u, c_k) = \frac{N\left(A_{c_k}D_{c_k} - C_{c_k}B_{c_k}\right)^2}{\left(A_{c_k}+C_{c_k}\right)\left(B_{c_k}+D_{c_k}\right)\left(A_{c_k}+B_{c_k}\right)\left(C_{c_k}+D_{c_k}\right)}, \quad (1)$$

where t_u is the u-th word, c_k is the k-th class, K is the number of classes, and N is the number of training documents. Meanwhile, A_{c_k} is the number of documents in class c_k containing t_u, B_{c_k} is the number of documents not in class c_k containing t_u, C_{c_k} is the number of documents I class c_k not containing t_u, and D_{c_k} is the number of documents not in class c_k not containing t_u.

The single Chi-Square value of a word is obtained by summing the Chi-Square values of the word across all K classes using equation (2):

$$\chi^2(t_u) = \sum_{k=1}^{K} \chi^2(t_u, c_k). \quad (2)$$

2.1.2 Information Gain Feature Selection

Information Gain feature selection identifies features that contain the most information based on a particular class. The best attribute is determined by calculating entropy. Entropy measures class uncertainty using the probability of occurrence of certain events or attributes. The entropy function before observing attribute A within classification classes is obtained using equation (3) [12]:

$$H(C) = -\sum_{k=1}^{K} P(c_k) \log_2 P(c_k), \quad (3)$$

where the c_k is the k-th class, K is the number of classes, and $P(c_k)$ is the probability of class c_k.

The entropy function after observing the u-th word is obtained using equation (4):

$$H(C|t_u) = \sum_{a \in values\ t_u} \left(P(a)\left(-\sum_{k=1}^{K} P(c_k|a) \log_2 P(c_k|a)\right)\right), \quad (4)$$

where a is the value of t_u, $P(a)$ is the probability of value a for attribute A, and $P(c_k|a)$ is the probability of class c_k given value a.

After obtaining entropy values, Information Gain for the u-th word can be calculated using equation (5):

$$IG(t_u) = H(C) - H(C|t_u). \quad (5)$$

2.2 Word Representation Using TF-IDF

Word representation is performed by transforming textual data into vectors so that it can be processed by machine learning models. [13] stated that Term Frequency–Inverse Document Frequency (TF-IDF) is a feature extraction method in which word frequencies are recalculated by considering how frequently those words appear across all documents. TF-IDF is obtained using the following equations:

$$\text{TF}(t_u, d_i) = \frac{n_{d_i,t_u}}{N_{d_i}}, \quad (6)$$

$$\text{IDF}(t_u) = \log \frac{N}{df(t_u)+1}, \quad (7)$$

$$\text{TF-IDF}(t_u, d_i) = \text{TF}(t_u, d_i) \times \text{IDF}(t_u), \quad (8)$$

where n_{d_i,t_u} is the number of occurrences of the u-th word in the i-th document, N_{d_i} is the total number of words in the i-th document, and $df(t_u)$ is the number of documents containing the u-th word. The resulting TF-IDF vector is then normalized using the Euclidean norm as shown in equation (9):

$$\text{TF-IDF}_{\text{norm}}(t_u, d_i) = \frac{\text{TF-IDF}(t_u, d_i)}{\sqrt{\text{TF-IDF}(t_1, d_i)^2 + \text{TF-IDF}(t_2, d_i)^2 + \cdots + \text{TF-IDF}(t_m, d_i)^2}}, \quad (9)$$

where $u = 1,2,3, \ldots, m$.

The resulting TF-IDF vector is then used as input for the Support Vector Machine classification method.

2.3 Word Representation Using TF-IDF

Support Vector Machine is a statistical method used for classification. SVM is a supervised learning method that analyzes data and recognizes patterns, used for classification and regression analysis [14]. According to [15], SVM aims to find the optimal separating function (hyperplane) that divides one class from another. The optimal hyperplane is determined by maximizing the distance (margin) between the hyperplane and the nearest data points (support vectors) from each class. The components of SVM include the optimal hyperplane, positive class hyperplane, negative class hyperplane, and margin.

The optimal hyperplane is defined in equation (10) [16]:

$$(\mathrm{w} \cdot \mathrm{x}) + b = 0, \quad (10)$$

where x is the input data vector, w is the weight vector, and b is a scalar bias term.

The nearest point from the positive class (positive hyperplane) is defined as:

$$(\mathrm{w} \cdot \mathrm{x}_{i+}) + b = 1, \quad (11)$$

where x_{i+} is the positive support vector

The nearest point from the negative class (negative hyperplane) is defined as:

$$(\mathrm{w} \cdot \mathrm{x}_{i-}) + b = -1, \quad (12)$$

where x_{i-} is the negative support vector.

According to [17], for all data i, the negative hyperplane can be formulated as:

$$(\mathrm{w} \cdot \mathrm{x}_i) + b \leq -1, \quad (13)$$

and the positive hyperplane can be formulated as:

$$(\mathrm{w} \cdot \mathrm{x}_i) + b \geq 1. \quad (14)$$

From equation (13) and (14), equation (15) is obtained:

$$y_i[(\mathrm{w} \cdot \mathrm{x}_i) + b] \geq 1 \quad (15)$$

where y_i is the label of the i-th document, x_i is the i-th input vector from x, and $\mathrm{x}_i \in R^n$, $y_i \in \{-1, +1\}$ for $i = 1,2, \ldots, N$.

The distance between the two hyperplanes (margin) is obtained by calculating the projection length of the difference between the support vectors of each class onto the vector w as shown in equation (16):

$$\rho = \frac{\mathrm{w}}{||\mathrm{w}||} \cdot (\mathrm{x}_{i+} - \mathrm{x}_{i-}). \quad (16)$$

Substituting equation (11) and (12) into equation (16), we obtain:

$$\rho = \frac{\mathrm{w}}{||\mathrm{w}||} \cdot (\mathrm{x}_{i+} - \mathrm{x}_{i-}) = \frac{(\mathrm{w} \cdot \mathrm{x}_{i+}) - (\mathrm{w} \cdot \mathrm{x}_{i-})}{||\mathrm{w}||} = \frac{1 - b - (-1 - b)|}{||\mathrm{w}||} = \frac{2}{||\mathrm{w}||}. \quad (17)$$

Thus, the margin ρ is:

$$\rho = \frac{2}{||\mathrm{w}||}. \quad (18)$$

The largest margin can be obtained by maximizing $\frac{1}{||\mathrm{w}||}$, which is equivalent to minimizing $||\mathrm{w}||$. The problem of finding the hyperplane with the largest margin can be formulated as a quadratic programming problem in equation (19):

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}||\mathrm{w}||^2 \\ &\text{subject to} && y_i[(\mathrm{w} \cdot \mathrm{x}_i) + b] \geq 1, \forall i. \end{aligned} \quad (19)$$

In practice, two classification classes cannot always be perfectly separated. As a result, an optimal hyperplane cannot always be found. To overcome this, SVM is designed using the soft-margin technique. According to [18], the constraint function is modified by adding a slack variable (ξ_i) as follows:

$$y_i[(\mathrm{w} \cdot \mathrm{x}_i) + b] \geq 1 - \xi_i, \forall i. \quad (20)$$

To minimize the slack ξ_i variables, a regularization parameter C is introduced to control tolerance for misclassification errors. Thus, equation (19) becomes:

$$\begin{aligned} &\text{minimize} && \frac{1}{2}\left|\left|\mathrm{w}\right|\right|^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{subject to} && y_i[(\mathrm{w}\cdot \mathrm{x}_i)+b] \geq 1-\xi_i, \forall i \\ & && \xi_i > 0, \forall i, \end{aligned} \tag{21}$$

where $C > 0$.

A larger value of C gives a larger penalty for classification errors.

The optimization problem in equation (21) can be solved using the Lagrange Multiplier method:

$$L(\mathrm{w}, b, \xi, \alpha, \beta) = \frac{1}{2}\left|\left|\mathrm{w}\right|\right|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i(y_i[(\mathrm{w}\cdot \mathrm{x}_i)+b]-1+\xi_i) - \sum_{i=1}^{N}\beta_i\xi_i, \tag{22}$$

where α_i and β_i are Lagrange multipliers. The optimal values of the multipliers are obtained by partially differentiating L with respect to w, b, and ξ and setting them equal to zero, resulting in:

$$\mathrm{w} = \sum_{i=1}^{N}\alpha_i y_i \mathrm{x}_i, \tag{23}$$

$$\sum_{i=1}^{N}\alpha_i y_i = 0, \tag{24}$$

$$C = \sum_{i=1}^{N}\alpha_i + \sum_{i=1}^{N}\beta_i. \tag{25}$$

Substituting equation (23), (24), dan (25) into equation (22) yields:

$$L(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j(\mathrm{x}_i \cdot \mathrm{x}_j). \tag{26}$$

Equation (26) is used to find the best hyperplane by transforming it into a maximization problem to determine α_i. The problem is solved using the Sequential Minimal Optimization (SMO) heuristic method:

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j(\mathrm{x}_i \cdot \mathrm{x}_j) \\ &\text{subject to} && \sum_{i=1}^{N}\alpha_i y_i = 0, \forall i \\ & && 0 \leq \alpha_i \leq C, \forall i. \end{aligned} \tag{27}$$

After obtaining α_i values, equation (23) is used to compute w. To obtain b, equations (11) and (12) can be rewritten as:

$$b_s = y_s - \mathrm{w}^T\mathrm{x}_s, \tag{28}$$

where b_s is the bias term for the support vector and y_s is the support vector label.

After obtaining w and b, the input vector x can be classified using:

$$f(\mathrm{x}) = \mathrm{w}\cdot \mathrm{x} + b. \tag{29}$$

Substituting equation (23) into equation (29):

$$f(\mathrm{x}) = \sum_{i=1}^{N}\alpha_i y_i \mathrm{x}_i \cdot \mathrm{x} + b. \tag{30}$$

The classification class is determined as:

$$sign(f(\mathrm{x}) = \begin{cases} +1, & \sum_{i=1}^{N}\alpha_i y_i \mathrm{x}_i \cdot \mathrm{x} + b \geq 0 \\ -1, & \sum_{i=1}^{N}\alpha_i y_i \mathrm{x}_i \cdot \mathrm{x} + b < 0. \end{cases} \tag{31}$$

In real-world applications, datasets are not always linearly separable. To solve nonlinear problems, SVM is modified using kernel functions. This approach is known as the kernel trick. Data is transformed into a space where it can be separated linearly. The kernel trick maps low-dimensional nonlinear data into a higher-dimensional space [19].

In nonlinear SVM, input data x is first mapped using $\Phi(\mathrm{x})$ into a higher-dimensional vector space so that the classes can be separated linearly by a hyperplane [20]:

$$\begin{aligned} &\Phi: \mathbb{R}^d \to \mathbb{R}^q, d < q \\ &\mathrm{x} \mapsto \Phi(\mathrm{x}). \end{aligned} \tag{32}$$

With this mapping, the equations for w and b become:

$$\mathrm{w} = \sum_{i=1}^{N}\alpha_i y_i \Phi(\mathrm{x}_i), \tag{33}$$

$$b_s = y_s - \mathrm{w}^T\Phi(\mathrm{x}_s). \tag{34}$$

The SVM learning process depends on the dot product $\Phi(\mathrm{x}_i) \cdot \Phi(\mathrm{x}_j)$. Based on Mercer's theorem, this can be replaced with a kernel function $K(\mathrm{x}_i, \mathrm{x}_j)$:

$$K(\mathrm{x}_i, \mathrm{x}_j) = \Phi(\mathrm{x}_i) \cdot \Phi(\mathrm{x}_j). \tag{35}$$

Thus, equation (27) becomes:

$$\text{maximize} \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(\mathrm{x}_i, \mathrm{x}_j). \quad (36)$$

Classification of input data becomes:

$$f(\Phi(\mathrm{x})) = \sum_{i=1}^{N_s} \alpha_i y_i K(\mathrm{x}_i, \mathrm{x}_j) + b. \quad (37)$$

The classification result is:

$$sign\left(f(\Phi(\mathrm{x}))\right) = \begin{cases} +1, & \sum_{i=1}^{N_s} \alpha_i y_i K(\mathrm{x}_i, \mathrm{x}_j) + b \geq 0 \\ -1, & \sum_{i=1}^{N_s} \alpha_i y_i K(\mathrm{x}_i, \mathrm{x}_j) + b < 0. \end{cases} \quad (38)$$

where N_s is the number of support vectors.
Commonly used kernel functions are shown in Table 1.

Table 1. Kernel Functions [16]

Kernel Name	$K(\mathbf{x}_i, \mathbf{x}_j)$
Linear	$\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$
Polynomial	$\left(\gamma\left(\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)\right) + r\right)^p$
Gaussian Radial Basis Function (RBF)	$\exp\left(-\gamma \left\|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\right\|^2\right)$
Sigmoid	$\tanh\left(\gamma\left(\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)\right) + r\right)$

2.4 Model Performance Evaluation

Model performance evaluation in this study uses the balanced accuracy metric. Balanced accuracy is used to evaluate how well the model predicts binary classes [21]. Balanced accuracy is calculated as:

$$Balanced\ Accuracy = \frac{1}{2} \cdot \left(\frac{TP}{TP + FN} + \frac{TF}{TN + FP}\right), \quad (39)$$

where TP is the number of positive samples correctly predicted as positive, TN is the number of negative samples correctly predicted as negative, FP is the number of negative samples incorrectly predicted as positive, and FN is the number of positive samples incorrectly predicted as negative. In addition to balanced accuracy, accuracy, precision, recall, and F1-score are also calculated to complement model evaluation.

2.5 Model Performance Evaluation

Hyperparameter tuning is conducted to improve SVM model performance. This study uses Grid Search to identify the optimal hyperparameter combination. The hyperparameters used include kernel type (linear, polynomial, Gaussian RBF, sigmoid), regularization constant ($C \in \{\frac{1}{100}, \frac{1}{10}, 1, 10, 100\}$), gamma ($\gamma \in \{\frac{1}{100}, \frac{1}{10}, 1, 10, 100\}$), polynomial degree ($p \in \{2, 3, 4, 5\}$), and offset parameter ($r \in \{-1, 0, 1\}$).

3 RESEARCH METHODS

3.1 Data

The data collected in this study consists of Vidio application reviews obtained from Google Play Store between July 1, 2024 and November 30, 2024. Data collection was conducted through web scraping using Python with the google-play-scraper library, resulting in a total of 4,670 reviews.

The Vidio reviews were manually labeled as positive or negative. To reduce subjective judgment, labeling was conducted by three individuals who reviewed each entry one by one. Reviews were only labeled as positive or negative because neutral sentiment is difficult to determine. Labeling was based on the intent and meaning of the review, and ratings were also considered when there was uncertainty in determining the sentiment label.

3.2 Text Preprocessing

Text preprocessing applied to the Vidio review dataset includes case folding to convert all uppercase letters to lowercase, punctuation removal to eliminate all non-alphabetic characters, normalization to convert informal words into standard forms, tokenization to split text into smaller units, stemming to extract root words, and stopwords removal to eliminate less meaningful words.

3.3 Data Splitting

In this step, the dataset is partitioned into two subsets: training data and testing data with an 80% and 20% ratio, respectively.

3.4 Feature Selection

3.4.1 Chi-Square Feature Selection

Chi-Square feature selection applied to training data to reduce less relevant features across all documents. This process selects words that best represent the dataset by testing the independence of each term with its class. The best features are determined by calculating Chi-Square values for each word using equations (1) and (2).

3.4.2 Information Gain Feature Selection

Information Gain feature selection is applied to the training data to reduce less relevant features across all documents. The best features are determined by calculating the Information Gain value for each word using equations (3) through (5).

3.5 Word Representation Using TF-IDF

One of the most widely used word representation methods is TF-IDF. TF-IDF is used to measure the relative importance of a word in a document. Word representation is applied to the training data after the feature selection stage. Each word must be transformed into a numeric vector so that it can be processed by the classification model. TF-IDF vector calculation is performed using equations (6) through (9).

3.6 Classification Using Support Vector Machine

The $\mathbf{TF\text{-}IDF}_{\text{norm}}(t_u, d_1) = \begin{bmatrix} \text{TF-IDF}_{\text{norm}}(t_1, d_1) \\ \text{TF-IDF}_{\text{norm}}(t_2, d_1) \\ \vdots \\ \text{TF-IDF}_{\text{norm}}(t_m, d_1) \end{bmatrix}$ vector obtained from the training data used as the input vector for the SVM classification model. The SVM method aims to find the best hyperplane that separates two data classes with maximum margin. The hyperplane determined by calculating $\mathbf{w}$ using (33) and b using (34).

3.7 Determining the Class of Testing Data

The class of the testing data is determined by substituting the testing input vectors into the sign function in equation (37), and the class is determined using equation (38). Testing data must also undergo text preprocessing and feature selection before being transformed into TF-IDF vectors.

3.8 Model Performance Evaluation

Model performance evaluation is carried out by calculating performance metrics. For imbalanced data, model evaluation uses balanced accuracy as defined in equation (39). Afterward, hyperparameter tuning is conducted to improve the performance of the existing SVM model.

4 RESULTS AND DISCUSSION

4.1 Data

The dataset used in this study consists of Vidio application reviews collected from Google Play Store between July 1, 2024 and November 30, 2024, totaling 4,670 reviews. The collected data was labeled based on sentiment into positive or negative categories. Of the 4,670 reviews, 3,431 reviews (73.5%) were labeled as negative sentiment, while 1,239 reviews (26.5%) were labeled as positive sentiment. This indicates that the dataset is imbalanced. Sample data is shown in Table 2.

Table 2. Sample User Review Data of the Vidio Application

Reviews	Sentiment
Kebanyakan iklan👎👎	Negative
Mantap	Positive
Download ini untuk nonton byon 4 😑	Positive
Beli Paket Byon Combat Vol.4 Ga bisa di tonton , kocak	Negative

4.2 Text Preprocessing

Text preprocessing was applied to the labeled dataset. The preprocessing steps included case folding to convert uppercase letters into lowercase, punctuation removal to remove non-alphabetic characters, normalization to convert informal words into standard forms, tokenization to split text into smaller units, stemming to extract root words, and stopword removal to remove less meaningful words. Reviews that contained only stopwords or non-alphabetic characters became empty and were removed, reducing the number of reviews to 4,584. Samples of preprocessing results are shown in Table 3.

Table 3. Sample Results of Text Preprocessing for Vidio Application Reviews

Reviews	Text Preprocessing	Sentiment
Kebanyakan iklan👎👎	'iklan'	Negative
Mantap	'mantap'	Positive
Download ini untuk nonton byon 4 😑	'download', 'tonton', 'byon'	Positive
Beli Paket Byon Combat Vol.4 Ga bisa di tonton , kocak	'beli', 'paket', 'byon', 'combat', 'volume', 'tidak', 'tonton', 'kocak'	Negative

4.3 Data Splitting

The data was divided into 80% training data (3,667 reviews) and 20% testing data (917 reviews).

4.4 Feature Selection

4.4.1 Chi-Square Feature Selection

Chi-Square feature selection was performed on the training data using equations (1) and (2). Sample results of Chi-Square values for each word are presented in Table 4.

Table 4. Sample Single Chi-Square Values for Each Word in Vidio Application Reviews

t_u	$\chi^2(t_u)$
tidak	1.282,747
bagus	678,6296
mantap	578,9626
⋮	⋮
broadcast	0,001989
bicara	0,001989

From Table 4, the words with the highest Chi-Square values are "tidak", "bagus", and "mantap", indicating that these words are the most effective for determining classification classes. Meanwhile, the words with the lowest Chi-Square values are "cerita", "broadcast," and "bicara", indicating that these words are less representative for determining classification classes.

4.4.2 Information Gain Feature Selection

Information Gain feature selection was performed on the training data using equations (3), (4), and (5). Sample results of Information Gain values for each word are shown in Table 5.

Table 5. Sample Information Gain Values for Each Word in Vidio Application Reviews

t_u	$IG(t_u)$
tidak	0,1064
langgan	0,041977
bagus	0,040818
⋮	⋮
broadcast	$1{,}36\times 10^{-7}$
bicara	$1{,}36\times 10^{-7}$

From Table 5, the words with the highest Information Gain values are "tidak", "langgan", and "bagus", indicating that these words are the most effective for determining classification classes. The words with the lowest Information Gain values are "broadcast" and "bicara", indicating that these words are less representative for classification.

4.5 Classification Using Support Vector Machine

The testing dataset consists of 917 reviews. The training dataset that has undergone text preprocessing, feature selection, and TF-IDF word representation produces input vectors **x** for the SVM classification method. The default model performance results for each feature selection method are shown in Table 6.

Table 6. Default Performance Results of Vidio Application Reviews for Each Feature Selection Method

Feature Selection	*Accuracy*	*Precision*	*Recall*	F1-*Score*	*Balanced accuracy*
Chi-Square	94,77%	92,67%	93,63%	93,13%	93,63%
Information Gain	93,02%	92,64%	88,44%	90,28%	88,44%

Hyperparameter tuning was then performed on $C, r, p, \gamma,$ and kernel parameters defined by the researcher. The best model performance results for each feature selection method are presented in Table 7.

Table 7. Best Performance Results of Vidio Application Reviews for Each Feature Selection Method

Feature Selection	Best parameter	*Accuracy*	*Precision*	*Recall*	F1-*Score*	*Balanced accuracy*
Chi-Square	Linear $C = 1$	95,2%	92,88%	94,78%	93,78%	94,78%
Information Gain	*Polynomial* $C = 10, r = 1, p = 3, \gamma = 0{,}1$	94,33%	91,49%	94,2%	92,72%	94,2%

Based on the balanced accuracy metric in Table 7, the best model performance was achieved using Chi-Square feature selection with a balanced accuracy of 94.78%, using a linear kernel and hyperparameter $C =$ 1. In contrast, the Information Gain method yielded a slightly lower Balanced Accuracy of 94.20%. Notably, this method required a much more complex configuration to reach its peak performance, utilizing

a Polynomial Kernel with parameters $C = 10$, $r = 1$, $p = 3$, and $\gamma = 0{,}1$. The comparison highlights that Chi-Square not only outperforms Information Gain in accuracy but does so with a significantly simpler model architecture.

4.6 Sentiment Analysis Results of Vidio Application Reviews

The classification results of the testing data using the best tuned model are shown in Table 8.

Table 8. Classification Results of Vidio Application Review Testing Data After Hyperparameter Tuning

Reviews	Actual Sentiment	Prediction Sentiment
Top markotop dah	Positive	Positive
Download krn pngen nnton zona merah.. tp trnyata lemotnya bukan maen...	Negative	Negative
Aplikasinya bagus Cukup menghibur.	Positive	Positive

Based on sentiment analysis results of Vidio user reviews, it was found that out of 917 reviews, 670 reviews (73.06%) were classified as negative sentiment, while 247 reviews (26.94%) were classified as positive sentiment.

The classification performance comparison between SVM with Chi-Square feature selection and SVM with Information Gain feature selection shows a difference in balanced accuracy. SVM with Chi-Square feature selection achieved a slightly higher balanced accuracy of 94.78% compared to SVM with Information Gain feature selection which achieved 94.2%. Although the difference in peak Balanced Accuracy is 0.58% (94.78% vs 94.20%), the practical significance of the proposed method lies in its computational efficiency.

5 CONCLUSION

Based on the results and discussion in this study, the following conclusions can be drawn:

1. Sentiment analysis results of Vidio application user reviews using SVM classification with Chi-Square feature selection produced 73.06% negative sentiment and 26.94% positive sentiment. Meanwhile, using Information Gain feature selection produced 72.19% negative sentiment and 27.81% positive sentiment.
2. The Chi-Square method demonstrated superior stability compared to Information Gain. In the default SVM mode (without hyperparameter tuning), Chi-Square achieved a baseline Balanced Accuracy of 93.63%, significantly outperforming Information Gain, which only reached 88.44%. This 5.19% gap indicates that features selected by Chi-Square are fundamentally more separable and less dependent on complex parameter optimization.
3. Through hyperparameter tuning, the SVM model based on Chi-Square proved to be the optimal approach, achieving the highest Balanced Accuracy of 94.78%. Crucially, this peak performance was attained using a computationally efficient Linear Kernel ($C = 1$). In contrast, the Information Gain method yielded a lower Balanced Accuracy of 94.20% despite requiring a complex Polynomial Kernel ($C = 10$, $r = 1$, $p = 3$, $\gamma = 0{,}1$) to converge. This confirms that the proposed method offers a superior trade-off between accuracy and computational cost.

For future research, deep learning classification methods such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Bidirectional Gated Recurrent Unit (Bi-GRU), and others may be applied.

REFERENCES

[1] H. Nurhayati-Wolff, “Number of users of video streaming (SVoD) in Indonesia from 2017 to 2027,” Statista.

[2] Google Play Store, “Vidio: Sports, Movies, Series.”

[3] A. Palanivinayagam, C. Z. El-Bayeh, and R. Damaševičius, “Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review,” *Algorithms*, vol. 16, no. 5, p. 236, Apr. 2023, doi: 10.3390/a16050236.

[4] M. Iqbal, M. M. Abid, M. N. Khalid, and A. Manzoor, “Review of feature selection methods for text classification,” *International Journal of Advanced Computer Research*, vol. 10, no. 49, pp. 138–152, Jul. 2020, doi: 10.19101/IJACR.2020.1048037.

[5] R. Guido, M. C. Groccia, and D. Conforti, “A hyper-parameter tuning approach for cost-sensitive support vector machine classifiers,” *Soft comput.*, vol. 27, no. 18, pp. 12863–12881, Sep. 2023, doi: 10.1007/s00500-022-06768-8.

[6] M. Fudhail Ferio Supeli and S. Setiaji, “Klasifikasi Sentimen Positif Dan Negatif Pada Aplikasi Vidio Dengan Algoritma K-Nearest Neighbor,” *Indonesian Journal Computer Science*, vol. 2, no. 1, pp. 7–15, Apr. 2023, doi: 10.31294/ijcs.v2i1.1874.

[7] I. L. Kharisma, D. A. Septiani, A. Fergina, and K. Kamdan, “Penerapan Algoritma Decision Tree

untuk Ulasan Aplikasi Vidio di Google Play," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 218–226, Sep. 2023, doi: 10.25077/TEKNOSI.v9i2.2023.218-226.

[8] A. Maulana, "Analisis Sentimen Menggunakan Algoritma K-Nearest Neighbor dan Naive Bayes pada Aplikasi Vidio," Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 2024.

[9] scikit-learn developers, "SVC - scikit-learn 1.8.0 documentation," https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html.

[10] I. S. Thaseen, Ch. A. Kumar, and A. Ahmad, "Integrated Intrusion Detection Model Using Chi-Square Feature Selection and Ensemble of Classifiers," *Arab. J. Sci. Eng.*, vol. 44, no. 4, pp. 3357–3368, Apr. 2019, doi: 10.1007/s13369-018-3507-5.

[11] C. F. Suharno, M. A. Fauzi, and R. S. Perdana, "Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square," *Systemic: Information System and Informatics Journal*, vol. 3, no. 1, pp. 25–32, Aug. 2017, doi: 10.29080/systemic.v3i1.191.

[12] E. Odhiambo Omuya, G. Onyango Okeyo, and M. Waema Kimwele, "Feature Selection for Classification using Principal Component Analysis and Information Gain," *Expert Syst. Appl.*, vol. 174, p. 114765, Jul. 2021, doi: 10.1016/j.eswa.2021.114765.

[13] R. N. Waykole and A. D. Thakare, "A Review of Feature Extraction Methods for Text Classification," *International Journal of Advance Engineering and Research Development*, vol. 5, no. 4, Apr. 2018.

[14] D. Oktavia, Y. R. Ramadahan, and M. Minarto, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 407–417, Aug. 2023.

[15] M. A. Reddy, T. T. Kumar, and G. S. S. R. Krishna, "Malaria Cell-Image Classification using InceptionV3 and SVM," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 8, pp. 6–10, Aug. 2021.

[16] R. A. Ariyanto and N. Chamidah, "Sentiment Analysis for Zoning System Admission Policy Using Support Vector Machine and Naive Bayes Methods," *J. Phys. Conf. Ser.*, vol. 1776, no. 1, p. 012058, Feb. 2021, doi: 10.1088/1742-6596/1776/1/012058.

[17] N. M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," in *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA 2016)*, Yogyakarta, Mar. 2016.

[18] Naiyang. Deng, Yingjie. Tian, and Chunhua. Zhang, *Support vector machines : optimization based theory, algorithms, and extensions*. CRC Press, Taylor & Francis Group, 2013.

[19] Md. S. Reza, U. Hafsha, R. Amin, R. Yasmin, and S. Ruhi, "Improving SVM performance for type II diabetes prediction with an improved non-linear kernel: Insights from the PIMA dataset," *Computer Methods and Programs in Biomedicine Update*, vol. 4, p. 100118, 2023, doi: 10.1016/j.cmpbup.2023.100118.

[20] A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support Vector Machine: Teori dan Aplikasinya dalam Bioinformatika," https://asnugroho.net/papers/ikcsvm.pdf.

[21] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 February 2026

Activation Function Sensitivity in LSTM-Based Peak Stock Price Forecasting for High-Volatility Financial Time Series

Rizal Rafi Nugraha[1)], Budiman[2)], Imannudin Akbar[1)]
[1)] Information System, Faculty of Technology and Informatics, Universitas Informatika dan Bisnis Indonesia
[2)] Informatics, Faculty of Technology and Informatics, Universitas Informatika dan Bisnis Indonesia
Email: rizal.rn21@student.unibi.ac.id; budiman@unibi.ac.id; imannudin@unibi.ac.id

Accepted:	Accepted After Revision:	Published:
27 February 2026	10 March 2026	12 March 2026

Abstract

Stock price prediction remains an intriguing task due to the high volatility and complex temporal dependencies present in financial time-series data. Accurate prediction of the highest stock price is particularly important for investors seeking to identify market peaks and optimize trading strategies. This study investigates the effectiveness of Long Short-Term Memory (LSTM) networks in forecasting DELL's highest stock price by analyzing the impact of different activation functions. Historical stock price data from 2016 to 2024 were used, and several preprocessing techniques, including data normalization and chronological train-test splitting, were applied. The LSTM models were trained for 100 epochs and evaluated using Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE). The main contribution of this research is a comparative analysis of the sensitivity of LSTM prediction performance to different activation functions, namely ReLU, ELU, Sigmoid, and Tanh, in the context of high-volatility financial time-series data. The experimental results show that the LSTM model using the ReLU activation function achieved the best performance, with an RMSE of 0.557942, MSE of 0.311300, and MAE of 0.338773, outperforming the other activation functions. These findings demonstrate that activation function selection significantly influences LSTM forecasting performance. The results provide practical insights for financial analysts and investors in selecting appropriate deep learning configurations for more reliable stock price prediction.

Keywords: Long Short-Term Memory (LSTM), stock price prediction, activation functions, ReLU, financial forecasting

1 INTRODUCTION

The stock market plays a crucial role in the global economy, serving as a barometer of companies' financial health and influencing investment decisions. [1]. Among the numerous factors that drive stock market activity, stock price prediction remains a critical area of interest [2]. Investors rely on accurate stock price predictions to make informed decisions that maximize returns and minimize risk. [3]. However, predicting stock prices is inherently complex due to the volatility and unpredictability of financial markets, which are influenced by a myriad of factors such as economic conditions, company performance, and market sentiment. [4], [5].

In particular, predicting the highest price of a stock, such as DELL's, is challenging. [6]. The highest price often represents the peak value within a trading period, which is crucial for investors aiming to capitalize on short-term price fluctuations. [7]. Traditional forecasting methods often fail to account for the temporal dependencies and non-linear relationships inherent in stock price movements. [8], [9]. Thus, there is a need for more advanced methodologies capable of capturing these complex patterns to improve prediction accuracy. [10].

The primary challenge in predicting stock prices lies in the inherent complexity of financial data. [11]. Stock prices are influenced by numerous external factors, including market sentiment, economic indicators, and geopolitical events, making them highly volatile and complex to predict [12]. Furthermore, the stock price data itself often exhibits non-linear behavior and temporal dependencies, which traditional models such as linear regression or time-series forecasting methods struggle to capture effectively. [13].

Additionally, choosing the appropriate model and optimization techniques remains a significant hurdle. While machine learning models, such as Long Short-Term Memory (LSTM) networks, have shown promise in handling sequential data, such as time series, their performance can vary with configuration, including the selection of activation functions [14]. The challenge of

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

selecting the optimal activation function to improve the model's predictive accuracy further complicates the application of LSTMs to stock price prediction [15].

Despite the growing adoption of Long Short Term Memory (LSTM) models in stock price forecasting, relatively limited attention has been given to the role of activation functions in influencing the predictive performance of these models. Most active studies chiefly focus on optimizing hyperparameters such as the number of layers, hidden units, learning rates, and dropout rates to improve forecasting accuracy. However, the selection of activation functions an essential component that determines how neural networks learn complex nonlinear relationships has not been systematically examined in the context of financial time series prediction. This limitation is peculiarly pertinent in extremely unstable markets, where the ability of the model to capture intense price fluctuations is critical. Consequently, there remains a research gap in understanding how different activation functions influence the learning dynamics and prediction accuracy of LSTM models for stock price forecasting, peculiarly for predicting peak stock prices in singular companies such as DELL.

This research aims to tackle these challenges by applying an LSTM model to predict DELL's highest stock price, using historical price data from 2016 to 2024. Specifically, the study aims to evaluate the performance of the LSTM model with various activation functions to determine which yields the highest prediction accuracy. The evaluation will be based on key metrics including Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE). Furthermore, the research will provide practical insights for investors on leveraging the optimized model to enhance decision-making in stock price predictions, ultimately helping them maximize returns and mitigate financial risks. This research not only advances deep learning techniques for financial forecasting but also provides valuable tools for investors to navigate the complexities of stock price prediction in a volatile market.

2 LITERATURE REVIEW

Stock price prediction has been a significant area of research due to its complexity and importance in financial decision-making. Numerous studies have applied Long Short-Term Memory (LSTM) networks to forecast stock prices, demonstrating the model's ability to capture temporal dependencies in financial data. However, the effect of different activation functions on the performance of LSTM models in predicting stock prices, particularly for individual stocks such as DELL, has not been widely explored.

M. Kumaresan et al. (2023) explored the use of LSTM for stock price prediction. However, they focused primarily on hyperparameters such as batch size, learning rate, and the number of hidden layers, without delving into the impact of activation functions [16]. Their research found that the LSTM performed well overall. However, they did not assess how different activation functions might affect the model's ability to predict the highest stock prices of companies like DELL. Further studies are required to understand how activation functions influence model performance in such contexts.

Similarly, Zian Wang (2024) compared LSTM with ARIMA for stock price forecasting and highlighted LSTM's advantages, including a 92% reduction in error rates. [17]. However, like the previous study, it did not consider the effects of activation functions on model performance. Instead, the research focused on the general applicability of LSTMs across multiple stocks, such as the S&P 500. However, it did not provide insights into how activation functions specifically impact the accuracy of DELL stock price predictions.

Jaya Bharathi M et al. (2024) proposed an advanced LSTM model for stock price forecasting, which achieved a high accuracy rate of 92% [18]. However, this research focused more on optimizing the overall architecture of the LSTM model and integrating Explainable AI (XAI) techniques, rather than examining the role of different activation functions in the LSTM model. The study did not explore how activation functions such as ReLU, ELU, Sigmoid, or Tanh could affect the model's predictive accuracy for specific stocks, including DELL.

Haoran Li (2024) emphasized the importance of LSTM architecture, particularly the number of neurons and dropout rates, in achieving accurate stock price predictions. [19]. However, the study did not consider the effect of activation functions on model performance, which could play a pivotal role in how well the model generalizes to unseen data and how it handles volatile market conditions, such as those seen in DELL's stock history.

Barathi Subramanian et al. (2024) investigated the limitations of traditional activation functions, such as sigmoid and tanh, when applied to small datasets. [20]. They proposed an alternative activation function, Sigmoid Tanh Squared (SST), to enhance learning and gradient flow, which could potentially improve stock price prediction accuracy when data is limited. However, their research did not specifically analyze the impact of these functions on DELL's stock, leaving room for further exploration into how these activation functions influence LSTM performance in predicting individual stock prices.

Jongjin Jung and Ji Yeon Kim (2024) examined the importance of model optimization for stock price forecasting, but did not discuss the effects of activation functions on model accuracy. [21]. Their study compared various parameter settings and highlighted the need for further research into how different activation functions can be fine-tuned to optimize LSTM performance, particularly for high-volatility stocks such as DELL.

In contrast, several other studies [22], [23], [24] explored LSTM applications for predicting stock prices

but did not delve into activation function optimization. Their work emphasized LSTM's ability to handle sequential data and capture temporal dependencies in stock price movements, but the choice of activation function was not a primary focus.

While previous studies have demonstrated the effectiveness of LSTM for stock price prediction, a noticeable gap remains in the literature regarding the impact of activation function selection on model performance. Most studies focus on optimizing hyperparameters, such as the number of layers, the number of neurons, and dropout rates. However, few have systematically investigated how different activation functions, including ReLU, ELU, Sigmoid, and tanh, influence LSTM models' ability to predict stock prices, particularly for individual stocks such as DELL.

The lack of studies specifically addressing the role of activation functions in financial forecasting models represents a significant gap in the existing literature. Understanding how these functions impact prediction accuracy is crucial, as activation functions are integral to the learning dynamics of neural networks. Different activation functions can significantly alter the model's ability to capture complex, non-linear patterns in stock price movements, and their effect might vary depending on the volatility and behavior of the specific stock being predicted.

This research aims to fill this gap by focusing specifically on the impact of activation functions, including ReLU, ELU, Sigmoid, and Tanh, in predicting the highest stock price of DELL. By comparing the performance of these activation functions in stock price forecasting, this study will provide valuable insights into how the choice of activation function can influence the accuracy and reliability of LSTM models in the financial domain.

3 RESEARCH METHODS

In this study, we applied a systematic approach to develop and evaluate a predictive model for forecasting DELL's highest stock price using Long Short-Term Memory (LSTM) networks. Figure 1 shows that the methodology is designed to address the challenges of time-series prediction, where capturing the temporal dependencies and non-linear patterns in financial data is crucial. The methods outlined below detail the steps for preprocessing the dataset, building the LSTM model, and evaluating its performance across various activation functions. By focusing on key stages such as data cleaning, normalization, and splitting, we ensure the model's robustness and accuracy, ultimately providing valuable insights for investors and analysts into stock price prediction.

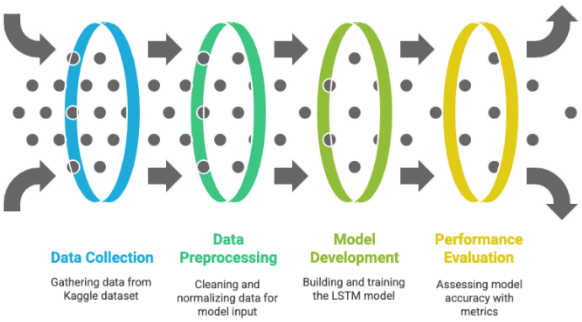


Figure 1. Proposed Method

3.1 Data Collection and Dataset Description

This research utilizes historical stock price data of DELL, obtained from Kaggle's "Microsoft Stock Data and Key Affiliated Companies" dataset. The dataset spans from August 17, 2016, to October 30, 2024, and contains 2,042 data points. The data includes key financial indicators: date, opening price, highest price, lowest price, closing price, adjusted closing price, and trading volume. For this study, the focus is on the "High" column, which records the highest stock price for DELL on each trading day. This column serves as the target variable for prediction, specifically predicting the highest price of DELL stock within the given time series.

3.2 Data Preprocessing

Data preprocessing in this study involved several critical steps to prepare the dataset for the LSTM model. First, data cleaning was performed to handle missing values by removing rows with null entries, ensuring data integrity. The dataset was then reduced to relevant features, specifically the "Date" and "High" columns, with the "Date" column formatted as datetime to maintain chronological order. Next, normalization was applied using Min-Max scaling, which scaled the "High" prices to the range [0,1], ensuring consistency for model input. The data was split into training and testing sets with a 70:30 ratio, maintaining chronological ordering to prevent future data leakage. The training dataset contained 1,429 data points, while the testing set contained 613.

The mathematical model used in this research is the Long Short-Term Memory (LSTM) network, which can be represented as follows:

$$h_t = \text{LSTM}(x_t, h_{t-1}, C_{t-1}) \quad (1)$$

where:

- h_t is the hidden state at time t
- x_t is the input at time t
- h_{t-1} is the previous hidden state
- c_{t-1} is the previous cell state.

The model's prediction for the highest stock price at time t, denoted as $\hat{y}$, is calculated as in (2):

$$\hat{y}_t = W_{\text{output}} h_t + b_{\text{output}} \quad (2)$$

where W_{output} is the weight matrix and b_{output} is the bias term. The model is trained using the Adam optimizer, and the loss is calculated in (3) using Mean Squared Error (MSE), which is minimized during training:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

where y_i represents the actual stock price and $\hat{y}_i$is the predicted value.

3.3 Model Development

The model developed in this study is an LSTM network, specifically designed for time-series prediction tasks due to its capability to capture long-term dependencies. The model architecture consists of a single LSTM layer with seven units, followed by a Dense layer to output the predicted highest stock price. The choice of activation functions includes ReLU, ELU, Sigmoid, and Tanh, each selected for their unique characteristics in influencing the model's learning behavior. The model was trained for 100 epochs using the Adam optimizer with a batch size of 32, employing early stopping to prevent overfitting. Performance evaluation was done using three metrics: Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE), to assess the model's prediction accuracy and robustness

.

3.4 Performance Comparison

To determine the most effective activation function for predicting DELL's highest stock price, the LSTM model's performance with each activation function was compared using RMSE, MSE, and MAE. The results of these evaluations provide insights into which activation function yields the most accurate predictions.

3.5 Experimental Setup

The experiments were conducted using Python, leveraging libraries such as Keras to build and train the LSTM model, Pandas for data manipulation, NumPy for numerical calculations, and Matplotlib for data visualization. The models were trained and evaluated on a local machine with sufficient computational power, including access to high-performance hardware for faster training using a Graphics Processing Unit (GPU).

4 RESULTS AND DISCUSSION

4.1 Results

To ensure the robustness and originality of the implementation, the LSTM model in this study was configured with a single LSTM layer consisting of seven hidden units followed by a dense output layer for regression prediction. The model was trained using the Adam optimizer with a batch size of 32 for 100 epochs , while the loss function was defined using Mean Squared Error (MSE). Early stopping was applied to reduce the risk of overfitting during training. In addition to the training and testing split (70:30), model validation was performed by monitoring the training loss and testing performance across epochs to ensure steady convergence of the model.

Although the dataset was obtained from Kaggle, the implementation in this research was independently developed developed by designing a customized LSTM configuration and consistently evaluating four activation functions (ReLU, ELU, Sigmoid, and Tanh) within the comparable experimental framework . The experimental results indicate that the ReLU activation function achieves superior predictive performance compared with the different functions . This can be explained by the ability of ReLU to mitigate the vanishing gradient problem and maintain efficient gradient propagation during backpropagation , enabling the model to learn nonlinear patterns and sharp fluctuations further efficaciously in unstable financial time series data. In contrast, Sigmoid and Tanh tend to suffer from gradient saturation , which slows the learning process and limits the model's ability to capture explosive changes in stock prices.

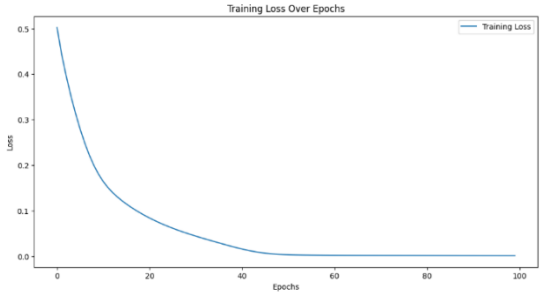


Figure 2. Training Loss ReLU Activation Function

The ReLU activation function demonstrated the best performance across all metrics. As shown in Figure 2, the model's training loss significantly decreased over the epochs, indicating effective learning. The RMSE, MSE, and MAE for the ReLU model were the lowest among the activation functions, indicating higher prediction accuracy and a better fit to historical data. This result suggests that ReLU is the most effective activation function for capturing the patterns in DELL's stock price.

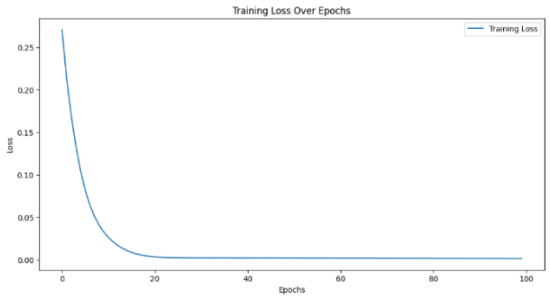


Figure 3. Training Loss ELU Activation Function

The ELU activation function showed a consistent decrease in training loss over the epochs, as illustrated in Figure 3. However, the performance on the test data was not as robust as ReLU. While the model performed well on the training data, it exhibited signs of overfitting, as evidenced by higher error metrics on the testing dataset. This suggests that ELU may struggle to generalize effectively to unseen data, rendering it less suitable for this stock price prediction task.

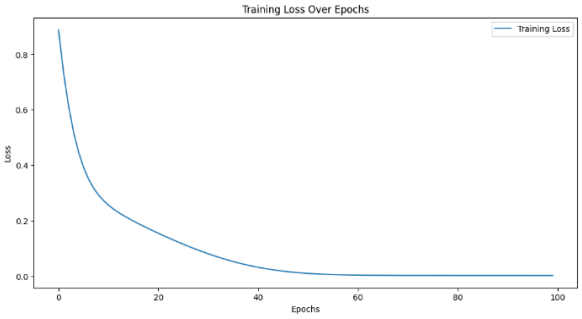


Figure 4. Training Loss Sigmoid Activation Function

The Sigmoid activation function, shown in Figure 4, exhibited slower convergence than ReLU and ELU, with a less pronounced reduction in loss over the epochs. The model trained with the Sigmoid function exhibited significant overfitting, leading to poorer performance on the test data. The prediction errors, as measured by RMSE, MSE, and MAE, were higher than those of the ReLU model, suggesting that Sigmoid is less suited for this particular prediction task.

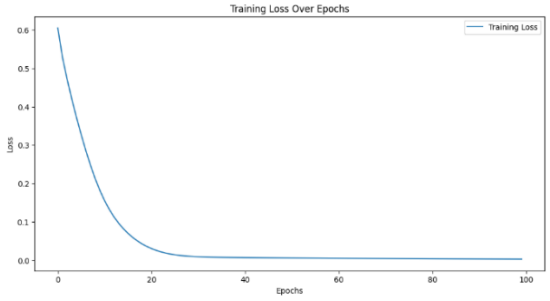


Figure 5. Training Loss Tanh Activation Function

The Tanh activation function, as seen in Figure 5, demonstrated the slowest improvement in training loss. Despite a steady reduction in loss during training, the model exhibited significant overfitting, with a substantial discrepancy between its training and testing performance. The error metrics for the Tanh model were the highest, indicating that this activation function does not effectively capture the complex patterns in stock price movements.

To further quantify each activation function's performance, the models were evaluated using three commonly used metrics: Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE). These metrics provide insight into the overall prediction accuracy and the magnitude of errors for each model. The results are summarized in Table 1.

Table 1. Comparison of Evaluation Metrics

Activation Function	R2 Score	RMSE	MSE	MAE
ReLU	0.843773	0.557942	0.3113	0.338773
ELU	0.2873	1.191694	1.420135	0.680826
Sigmoid	0.328764	1.156508	1.337512	0.654547
Tanh	-0.023218	1.427892	2.038876	0.843087

The table clearly shows that the ReLU activation function outperforms the other activation functions, achieving the lowest RMSE, MSE, and MAE values and the highest R^2 score. This indicates that the ReLU-based model is the most accurate at predicting DELL's highest stock price. In contrast, Tanh performed the worst, with the highest errors across all metrics, suggesting it is not an ideal choice for this type of financial prediction task.



Figure 6. Comparison of Prediction Results for The Highest Stock Price Of DELL Using Four Different Activation Functions In The LSTM Model

Figure 6 compares prediction results for DELL's highest stock price across four activation functions in the LSTM model: ReLU, ELU, Sigmoid, and Tanh. The graph clearly illustrates how each activation function impacts the model's ability to predict stock prices at various time points, with the predicted values compared against the actual values, represented by the yellow dashed line.

The ReLU activation function provided the most accurate predictions, aligning closely with the actual stock prices, especially during price surges. This indicates that the ReLU model effectively captures the fluctuations and key patterns in the stock price, making it the best-performing activation function in this study. On the other hand, the ELU model showed predictions that were generally close to the actual prices but lacked precision, particularly during sharp increases in stock prices. Although it followed the overall trend, it lagged behind actual prices during volatile periods, making it less effective than ReLU.

The Sigmoid activation function exhibited noticeable deviations from the actual stock prices, particularly during high-volatility periods. The Sigmoid model struggled to capture sharp rises and falls in stock prices, resulting in less accurate predictions than ReLU and ELU. Finally, the Tanh activation function showed the least accurate predictions, with significant discrepancies from the actual values. The Tanh model consistently underperformed during price surges, indicating it was unable to capture the complexity of stock price fluctuations.

In conclusion, Figure 6 demonstrates that ReLU offers the most accurate and reliable predictions for DELL's highest stock price, successfully tracking both trends and sharp price fluctuations. The findings from the model evaluation have significant implications for business and investment decision-making. Accurate prediction of stock prices, remarkably the highest price, plays a crucial role in informing trading strategies and investment decisions. By employing an LSTM model with the ReLU activation function, investors can potentially achieve better forecasting performance, enabling more informed decisions about buying and selling DELL stock. Accurate predictions of the highest stock price enable investors to identify peak market conditions and capitalize on price surges, thereby improving the timing of their trades.

For businesses, especially those in the financial sector, adopting machine learning models such as LSTMs with optimized activation functions can enhance risk management and more effectively allocate resources. By reducing prediction errors, the model enhances the accuracy of stock price forecasts, which in turn enables businesses to plan more effectively, minimize financial risks, and capitalize on market opportunities. The superior performance of the ReLU-based model underscores the importance of selecting the optimal configuration to improve financial predictions, ultimately leading to improved profitability in the long run. Moreover, such predictive models can be integrated into trading systems, providing real-time insights into market trends and helping businesses stay competitive in dynamic market environments.

4.2 Discussions

The results of this study highlight the significant influence of the choice of activation function on the performance of LSTM models in predicting DELL's highest stock price. The ReLU activation function consistently outperformed the other functions in terms of prediction accuracy. This aligns with previous research. [19] that emphasized the importance of optimizing the LSTM architecture to improve stock price prediction accuracy. ReLU effectively captured both general trends and sharp price surges, making it the most suitable activation function for predicting volatile stock prices, such as DELL's.

In contrast, the ELU, Sigmoid, and Tanh functions performed less effectively. Although ELU showed potential, it was prone to overfitting, especially during periods of high volatility. Both Sigmoid and tanh struggled to capture sudden price fluctuations, resulting in higher error rates compared to ReLU. This finding supports the limitations noted by [20], who found that traditional activation functions, such as Sigmoid and Tanh, were less effective for complex, non-linear stock price patterns.

While the LSTM model demonstrated strong performance with ReLU, its interpretability remains a challenge, as LSTM models are often considered "black-box" models. Future work could integrate Explainable AI (XAI) techniques, such as LIME (Local Interpretable Model-agnostic Explanations), to enhance model transparency, particularly in financial decision-making contexts. This would help investors better understand the factors influencing stock price predictions, thereby enhancing confidence in decision-making.

From a business perspective, accurately predicting stock prices is crucial for making informed investment decisions. The superior performance of the ReLU-based LSTM model demonstrates its potential as a reliable tool for predicting DELL's highest stock price, helping investors determine optimal entry and exit points. However, as with any model, the study's findings suggest the need for ongoing research to optimize LSTM models, particularly in terms of activation functions, to improve forecasting accuracy, especially in volatile markets.

Looking ahead, future research could explore the integration of external factors, such as market sentiment and macroeconomic indicators, that also influence stock prices. Incorporating such factors, along with advanced machine learning techniques such as reinforcement learning and ensemble methods, could enhance the robustness and predictive power of the LSTM model, making it even more valuable for real-world financial applications.

5 CONCLUSION

This study has demonstrated the significant impact of activation function selection on the performance of LSTM models in predicting DELL's

highest stock price. By applying and comparing four activation functions—ReLU, ELU, Sigmoid, and Tanh—we found that ReLU consistently outperformed the others in prediction accuracy. The best-performing model, using ReLU, achieved the following evaluation metrics: RMSE of 0.557942, MSE of 0.311300, and MAE of 0.338773. These results confirm that the ReLU activation function yields the most accurate forecasts for DELL's highest stock price, particularly in capturing both general trends and the sharp price fluctuations typically observed in stock markets.

The contribution of this research lies in its focus on the effect of activation functions on the performance of LSTM models for predicting individual stock prices, specifically for DELL. Previous studies have explored LSTM models for stock price prediction, but have often overlooked the impact of activation functions. By addressing this gap, this study provides valuable insights into how ReLU, in particular, can enhance the accuracy of stock price forecasts.

Furthermore, the findings from this study have practical implications for investors and financial analysts. By using LSTM models with an appropriate activation function, such as ReLU, stakeholders can make more informed investment decisions by optimizing buy, sell, or hold strategies based on more accurate stock price predictions. This research also underscores the importance of further investigating model interpretability and integrating external market factors to enhance the reliability and applicability of LSTM models in real-world financial forecasting.

In conclusion, this study makes a significant contribution to understanding how activation functions affect the performance of LSTM models for stock price prediction. It also lays the groundwork for future research that could explore the integration of explainable AI techniques and external factors, further refining the predictive power and interpretability of financial forecasting models.

REFERENCES

[1] M. A. Khan, H. Ali, H. Shabbir, F. Noor, and M. D. Majid, "Impact of Macroeconomic Indicators on Stock Market Predictions: A Cross-Country Analysis," *J. Comput. Biomed. Inform.*, vol. 8, no. 01, Art. no. 01, Oct. 2024, Accessed: Aug. 03, 2025. [Online]. Available: https://www.jcbi.org/index.php/Main/article/view/740

[2] N. Rouf *et al.*, "Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions," *Electronics*, vol. 10, no. 21, Art. no. 21, Jan. 2021, doi: 10.3390/electronics10212717.

[3] I. Botunac, J. Bosna, and M. Matetić, "Optimization of Traditional Stock Market Strategies Using the LSTM Hybrid Approach," *Information*, vol. 15, no. 3, Art. no. 3, Mar. 2024, doi: 10.3390/info15030136.

[4] K. Arora, A. Aggarwal, and K. K. Gola, "Predicting Stock Market Prices and Provide Recommendations," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 16, no. 3, Art. no. 3, Jul. 2024.

[5] T. Sahani, "Decoding Market Emotions: The Synergy of Sentiment Analysis and AI in Stock Market Predictions," *J. -Gener. Res. 50*, Dec. 2024, doi: 10.70792/jngr5.0.v1i1.47.

[6] M. Javed Awan, M. Shafry Mohd Rahim, H. Nobanee, A. Munawar, A. Yasin, and A. Mohd Zain Azlanmz, "Social Media and Stock Market Prediction: A Big Data Approach," *Comput. Mater. Contin.*, vol. 67, no. 2, pp. 2569–2583, 2021, doi: 10.32604/cmc.2021.014253.

[7] C. Li, W. Huang, W.-S. Wang, and W.-M. Chia, "Price Change and Trading Volume: Behavioral Heterogeneity in Stock Market," *Comput. Econ.*, vol. 61, no. 2, pp. 677–713, Feb. 2023, doi: 10.1007/s10614-021-10224-4.

[8] H. Pan, Y. Tang, and G. Wang, "A Stock Index Futures Price Prediction Approach Based on the MULTI-GARCH-LSTM Mixed Model," *Mathematics*, vol. 12, no. 11, Art. no. 11, Jan. 2024, doi: 10.3390/math12111677.

[9] E. G. A. Osman, "Integrating Deep Learning and Econometrics for Stock Price Prediction: An Empirical Study of Lstm and Traditional Time Series Models," Jul. 08, 2025, *Social Science Research Network, Rochester, NY*: 5340299. doi: 10.2139/ssrn.5340299.

[10] S. F. Ahmed *et al.*, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023, doi: 10.1007/s10462-023-10466-8.

[11] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. S., "Deep Learning for Stock Market Prediction," *Entropy*, vol. 22, no. 8, Art. no. 8, Aug. 2020, doi: 10.3390/e22080840.

[12] S. Maddodi and S. R. Kunte, "Market resilience in turbulent times: a proactive approach to predicting stock market responses during geopolitical tensions," *J. Cap. Mark. Stud.*, vol. 8, no. 2, pp. 173–194, Sep. 2024, doi: 10.1108/JCMS-12-2023-0049.

[13] D. Song and D. Song, "Stock Price Prediction based on Time Series Model and Long Short-term Memory Method," *Highlights Bus. Econ. Manag.*, vol. 24, pp. 1203–1210, Jan. 2024, doi: 10.54097/e75xgk49.

[14] W. Waheed, Q. Xu, M. Aurangzeb, S. Iqbal, S. H. Dar, and Z. M. S. Elbarbary, "Empowering data-driven load forecasting by leveraging long short-term memory recurrent neural networks," *Heliyon*, vol. 10, no. 24, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40934.

[15] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, "Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review," *Polymers*, vol. 16, no. 18, p. 2607, Sep. 2024, doi: 10.3390/polym16182607.

[16] M. Kumaresan, M. J. Basha, P. Manikandan, S. Annamalai, R. Sekaran, and A. S. Kumar, "Stock Price Prediction Model Using LSTM: A Comparative Study," in *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, Aug. 2023, pp. 1–5. doi: 10.1109/ASIANCON58793.2023.10270708.

[17] Z. Wang, "Stock price prediction using LSTM neural networks: Techniques and applications," *Appl. Comput. Eng.*, vol. 86, no. 1, pp. 294–300, Aug. 2024, doi: 10.54254/2755-2721/86/20241605.

[18] J. B. M and I. S, "Unlocking Market Trends: LSTM-based Stock Price Forecasting for Intelligent Investments," in *Advancements in Communication and Systems*, Malaviya National Institute of Technology Jaipur, A. K. Tripathi, V. Shrivastava, and National Institute of Technology Delhi, Eds., Soft Computing Research Society, 2024, pp. 627–633. doi: 10.56155/978-81-955020-7-3-55.

[19] H. Li, "Optimizing Stock Price Prediction: Exploring LSTM Architectural Parameters in Financial Forecasting," *Highlights Sci. Eng. Technol.*, vol. 85, pp. 1095–1100, Mar. 2024, doi: 10.54097/40px3f62.

[20] B. Subramanian, R. Jeyaraj, R. A. A. Ugli, and J. Kim, "Enhancing Sequential Model Performance with Squared Sigmoid TanH (SST) Activation Under Data Constraints," 2024, *arXiv*. doi: 10.48550/ARXIV.2402.09034.

[21] J. Jung and J. Kim, "A Performance Analysis by Adjusting Learning Methods in Stock Price Prediction Model Using LSTM," *J. Digit. Converg.*, vol. 18, no. 11, pp. 259–266, Nov. 2020, doi: 10.14400/JDC.2020.18.11.259.

[22] R. K. Vaish, "Stock Price Prediction Using LSTM Algorithm," *INTERANTIONAL J. Sci. Res. Eng. Manag.*, vol. 08, no. 05, pp. 1–5, May 2024, doi: 10.55041/IJSREM34831.

[23] M. Liu and Y. Zhao, "Stock Prediction Based on LSTM Model," in *2023 China Automation Congress (CAC)*, Chongqing, China: IEEE, Nov. 2023, pp. 1794–1798. doi: 10.1109/CAC59555.2023.10451610.

[24] R. Yang *et al.*, "Big data analytics for financial Market volatility forecast based on support vector machine," *Int. J. Inf. Manag.*, vol. 50, pp. 452–462, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.05.027.

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 February 2026

A Comparative Performance Analysis of Classification Algorithms for Hypertension Diagnosis

Imannudin Akbar, Titan Parama Yoga, Arnold Ropen Sinaga, Acep Hendra
Information System, Faculty of Technology and Informatics, Universitas Informatika dan Bisnis Indonesia
email: imannudin@unibi.ac.id; titanparama@unibi.ac.id; arnoldropen@unibi.ac.id; acephendra@unibi.ac.id.

Accepted:	Accepted After Revision:	Published:
26 February 2026	10 March 2026	12 March 2026

Abstract

Hypertension is a leading cause of cardiovascular diseases, strokes, and kidney failure, with early diagnosis being critical for prevention. Traditional diagnostic methods often face challenges such as human error and inconsistent measurements. While machine learning (ML) has been explored as a potential solution, previous studies have mainly focused on accuracy, often neglecting other important metrics like precision, recall, and F1-score, especially in imbalanced datasets. The primary purpose of this research is to address this gap by comprehensively comparing the performance of four machine learning algorithms - Naive Bayes, Support Vector Machines (SVM), Random Forest (RF), and XGBoost—to provide valuable insights for practical hypertension screening. The dataset consists of 1,985 records with 10 predictor features, including both categorical and continuous variables, and a binary target variable (Has_Hypertension: Yes/No) with a class distribution of 1,032 Yes and 953 No. The data undergoes preprocessing, including categorical encoding and feature scaling for SVM. Models are evaluated using a balanced set of metrics, including accuracy, precision, recall, and F1-score. The results show that RF/XGBoost perform best, with the highest F1 and accuracy, while SVM and Naive Bayes serve as competitive alternatives.

Keywords: Naive Baiyes, SVM, Random Forest, XGBoost, Hypertension

1 INTRODUCTION

Hypertension is a leading risk factor for cardiovascular diseases and is associated with a higher risk of stroke, heart disease, and kidney failure. Early diagnosis and management of hypertension are crucial in preventing these complications. However, traditional diagnostic methods, primarily based on manual blood pressure measurements, often suffer from inconsistencies due to human error, time variation, and equipment-related limitations. Consequently, there has been growing interest in leveraging machine learning (ML) models to automate the diagnosis of hypertension by using a wide array of patient data, such as age, BMI, blood pressure history, and lifestyle factors [1].

In this study, we utilize a dataset containing clinical and lifestyle information from individuals to predict the presence of hypertension. The dataset consists of multiple features including age, salt intake, stress score, blood pressure history, sleep duration, BMI, medication, family history, exercise level, smoking status, and the target variable has hypertension (Yes/No). This dataset provides a diverse set of features, which allows the application of various machine learning classifiers to assess their performance in diagnosing hypertension [2].

Recent studies have highlighted the growing role of machine learning in clinical prediction and health-data analysis [1],[3]. In comparative classification settings, algorithms such as Naive Bayes (NB), Support Vector Machines (SVM), Random Forest (RF), and XGBoost are widely used because they represent different modeling principles, ranging from probabilistic learning and margin-based classification to bagged trees and gradient-boosted ensembles [4],[8]. In healthcare datasets, these models are commonly assessed not only for overall accuracy, but also for robustness, interpretability, and their ability to generalize across heterogeneous clinical features [2], [3].

Despite these advancements, there are several gaps in existing literature. Many studies focus primarily on accuracy as the key metric for evaluating model performance, often ignoring other crucial metrics like precision, recall, and F1-score, which are critical when dealing with imbalanced datasets and the serious consequences of false negatives in medical diagnostics. Additionally, most studies do not utilize advanced

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

validation techniques like nested cross-validation, and the clinical significance of the false positive and false negative rates is often not adequately addressed.

To address these gaps, the primary purpose of this research is to comprehensively compare the performance of four well-known machine learning classifiers—Naive Bayes, Support Vector Machines (SVM), Random Forest, and XGBoost on a hypertension dataset. By evaluating these models using a balanced set of metrics, including accuracy, precision, recall, and F1-score, this study aims to provide actionable insights for healthcare providers regarding the most suitable models for practical hypertension screening. This approach ensures a deeper understanding of the models' strengths and weaknesses, focusing on both their predictive capabilities and clinical implications.

2 LITERATURE REVIEW

Machine learning (ML) has rapidly emerged as a powerful tool in healthcare, offering the potential to improve diagnosis, risk stratification, and clinical decision support. Recent literature in medical AI shows that data-driven models can help extract useful patterns from large-scale health records and structured clinical variables, thereby supporting more consistent and efficient prediction tasks [1], [2]. However, systematic evidence also suggests that the superiority of ML over conventional statistical approaches is not always guaranteed, making careful benchmarking and transparent evaluation essential [3].

Among the most commonly used classifiers in structured medical data are Support Vector Machines (SVM), Random Forest, and XGBoost. These methods are widely adopted because they can capture complex and non-linear relationships in classification tasks. Foundational work on Random Forest and XGBoost has demonstrated strong predictive capability, while widely used ML libraries have enabled their practical implementation in healthcare analytics and related applied studies [4], [5], [8].

Another critical aspect of machine learning in healthcare is data preprocessing, which ensures that the data are suitable for analysis. Techniques such as encoding, cleaning, and normalization are essential for improving algorithmic compatibility and model performance, especially for methods that are sensitive to feature scale differences. Min-Max normalization, for example, is widely used to rescale numerical features so that they contribute more equitably during model training [9], [10]. This step is particularly important in medical datasets that often contain a mixture of continuous and categorical variables.

Model evaluation is also a key stage in determining the effectiveness of machine learning methods for healthcare applications. Accuracy, precision, recall, and F1-score remain core metrics in binary classification, while cross-validation is commonly used to obtain a more reliable estimate of generalization performance [6], [7]. In addition, the Matthews Correlation Coefficient (MCC) is often recommended when class distributions are uneven because it accounts for all elements of the confusion matrix and provides a more balanced view of model quality [7].

From an applied perspective, healthcare machine learning studies in Indonesia also indicate growing interest in disease classification and the use of data-driven methods in health information environments [11]–[13]. This trend underscores the relevance of comparative studies that evaluate standard classifiers on clinically meaningful datasets. Therefore, comparing Naive Bayes, SVM, Random Forest, and XGBoost on a hypertension dataset remains methodologically relevant and practically valuable, especially as foundational concepts and implementation practices continue to be widely used in local and international contexts [14], [15].

3 RESEARCH METHODS

To ensure the reproducibility, objectivity, and robustness of our findings, this section delineates the systematic research methodology that underpins our empirical investigation, structuring the entire workflow from data acquisition to a final comparative performance analysis of the selected classification algorithms in a clear and logical progression.

a. Research Framework and Workflow

The research pipeline commences with the acquisition of the hypertension dataset, followed by an in-depth data preprocessing stage, a foundational step in any applied machine learning workflow designed to cleanse and prepare the data for modeling [9]. This stage is critical for ensuring algorithmic compatibility and optimal performance, involving standard procedures such as the encoding of categorical variables and the scaling of numerical features [10]. Subsequently, the prepared dataset is partitioned into training (80%) and testing (20%) sets. A stratified sampling technique is employed during this split to ensure that the class distribution of the target variable is maintained in both partitions, which is a crucial practice for preventing biased evaluation, particularly in clinical datasets that may be imbalanced [11]. The core of the methodology involves training four distinct machine learning models on the training data, where their performance is robustly estimated using a 10-fold cross-validation strategy. This technique is considered a standard for obtaining a reliable and

less biased estimate of a model's generalization capability [12]. Finally, the trained models are subjected to a definitive evaluation on the final, unseen test set to assess their real-world performance. This evaluation is conducted using a suite of standard classification metrics [13], leading to a comparative analysis from which conclusions about the most effective algorithm are drawn, a methodological approach common in comparative studies of clinical prediction models [14].

b. Data Collection
The first phase of this study is data collection, where we gather the essential information for our analysis. We compile a large set of anonymous patient records specifically related to hypertension. Each record contains various predictive details, including demographics, clinical data, and lifestyle habits, all linked to a final diagnosis. The quality and accuracy of this initial data are critical, as it forms the very foundation for training our machine learning models and testing how well they work. This study uses an anonymized, tabular hypertension dataset comprising N = 1,985 records with 10 predictor variables and a binary label (Has_Hypertension: Yes/No), exhibiting a class distribution of Yes = 1,032 and No = 953, and no missing values. Predictors include five continuous features—Age, BMI, Sleep_Duration (hours/day), Salt_Intake, Stress_Score (1–10)—and five categorical features—BP_History (Normal/Prehypertension/Hypertension), Medication (e.g., None/ACE Inhibitor/Other), Family_History (Yes/No), Exercise_Level (Low/Moderate/High), Smoking_Status (Non-Smoker/Smoker)—with consistent category labeling prior to encoding [15].

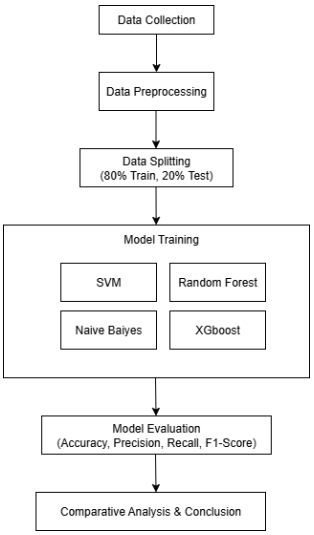


Figure 1. Research Methodology

c. Data Preprocessing
Raw data is seldom suitable for direct input into machine learning algorithms. The preprocessing stage is therefore critical for transforming the data into an optimal format and enhancing model performance [10].

1. Data Cleaning
Data cleaning is the process of identifying and rectifying data anomalies to enhance data quality and integrity. In a medical context, where data inaccuracies can lead to flawed conclusions, this step is particularly important [9].

2. Label Encoding
Machine learning algorithms require numerical input. Therefore, categorical features must be converted into a numerical format. Label Encoding is a technique that assigns a unique integer to each unique category within a feature [2].

3. Feature Scaling: Normalization.
Feature scaling adjusts the range of numerical features to bring them onto a common scale, which is essential for many algorithms. Normalization, specifically Min-Max scaling, is a primary technique for this purpose. By ensuring all features contribute equitably, normalization can significantly improve model convergence and performance [9].

d. Data Splitting
To ensure an objective evaluation, the data splitting phase methodically divides the processed dataset into two separate subsets. A standard 80/20 split is used, where 80% of the data forms the training set for the models to learn from. The other 20% is held back as the testing set, providing an unbiased measure of how well the models perform on new, unseen data. This separation is a fundamental practice to prevent overfitting—a common pitfall where a model learns the training data too well, including its noise, and consequently fails to generalize to new data—thereby giving a realistic estimate of real-world performance.

e. Model Training
Leveraging solely the 80% training partition, the model development phase serves as the primary computational stage. Four unique classifiers (Gaussian Naive Bayes, SVM, Random Forest, and XGBoost) are systematically trained, each employing its distinct optimization algorithm to derive a predictive function. This function aims to

model the complex correlations between input features and the diagnostic target. The stringent sequestration of the 20% testing data is a critical methodological control, guaranteeing that the subsequent performance evaluation is unbiased and a true measure of the models' generalization capabilities [4], [5].

f. Comparative Analysis and Conlusion
The study concludes with a comparative analysis where the performance scores of the four models are directly compared. This final step involves interpreting the results to identify the best-performing algorithm, focusing on which model offers the most optimal balance of accuracy, precision, and recall for diagnosing hypertension. A final conclusion is then drawn from this evidence to declare the most effective model and discuss the study's implications [7].

We use the following evaluation metrics to assess model performance: accuracy, precision, recall, and F1-score. Each of these metrics provides unique insights into the model's performance, particularly in scenarios with class imbalance, such as in hypertension diagnosis, where the costs of false positives and false negatives can be significant [7].

1. Accuracy
Accuracy measures the overall correctness of the model, defined as the proportion of correct predictions (both true positives and true negatives) out of all predictions. It is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where:
- TP = True Positives (correctly predicted positive cases)
- TN = True Negatives (correctly predicted negative cases)
- FP = False Positives (incorrectly predicted positive cases)
- FN = False Negatives (incorrectly predicted negative cases)

2. Precision
Precision is the proportion of true positive predictions relative to all positive predictions made by the model. It answers the question: "Of all the cases predicted as positive, how many were actually positive?" It is calculated as: [7].

$$\text{Precision} = \frac{TP}{TP + FP}$$

A high precision means that the model is not frequently classifying negative cases as positive, which is crucial in medical diagnostics to avoid unnecessary treatments [7].

3. Recall (Sensitivity or True Positive Rate)
Recall measures the proportion of actual positive cases that were correctly identified by the model. It answers the question: "*Of all the actual positive cases, how many did the model correctly identify*?" It is calculated as:

$$\text{Recall} = \frac{TP}{TP + FN}$$

A high recall means that the model is good at identifying positive cases, which is particularly important in medical diagnostics to minimize missed diagnoses (false negatives).

4. F1-Score
The F1-score is the harmonic mean of precision and recall. It balances the two metrics, giving a single score that accounts for both false positives and false negatives. The F1-score is particularly useful when the class distribution is imbalanced, as it considers both false positives and false negatives. It is calculated as: [7].

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-score is valuable when a balance between precision and recall is needed, making it a more comprehensive evaluation metric in the context of medical classification, where both false positives and false negatives can have significant consequences. These metrics allow for a detailed evaluation of model performance, ensuring that both types of errors (false positives and false negatives) are considered, which is crucial in the context of hypertension diagnosis where the goal is to correctly identify at-risk patients while minimizing misclassifications [7], [12], [13].

4 RESULTS AND DISCUSSION

4.1 Result

a. Data Collection
This research is based on a public dataset containing 1,985 anonymized patient records. The dataset features 11 predictor attributes and one binary target variable, “Has_Hypertension”. Its predictor variables encompass a mix of

demographic (e.g., Age), clinical (e.g., BMI, BP History), and lifestyle factors (e.g., Salt Intake, Smoking Status). This heterogeneous data composition provides a realistic and challenging testbed for evaluating the selected classification algorithms

b. Data Preprocessing

- Data Cleaning

```
Missing values per column:
Age                   0
Salt_Intake           0
Stress_Score          0
BP_History            0
Sleep_Duration        0
BMI                   0
Medication          799
Family_History        0
Exercise_Level        0
Smoking_Status        0
Has_Hypertension      0
dtype: int64
```

Figure 2. Missing Value

The issue here is a missing value in the Medication column. This type of data is categorical, meaning it's a descriptive label rather than a number. The solution is we do imputation,where the process of filling in the missing value with a calculated or estimated one. The standard statistical approach is to use the mode. This involves identifying the most common value in the Medication column and using it to replace the missing entry. For instance, if 'Beta Blocker' appears more often than any other drug, it would be used to fill the blank.

```
Missing values after imputation:
Age                 0
Salt_Intake         0
Stress_Score        0
BP_History          0
Sleep_Duration      0
BMI                 0
Medication          0
Family_History      0
Exercise_Level      0
Smoking_Status      0
Has_Hypertension    0
dtype: int64
```

Figure 3. Imputation Result

- Label Encoding

The primary objective of Label Encoding within this dataset is to transform ordinal categorical variables—those with an intrinsic hierarchical structure—into a numerical representation. Following the imputation of missing values, this technique was applied to the dataset to convert textual categories into integer values. Consequently, categorical features such as BP_History, Medication, Family_History, Exercise_Level, and Smoking_Status were successfully encoded into numerical formats (e.g., 0, 1, 2, 3). Similarly, the binary target variable, Has_Hypertension, was encoded as 0 and 1. This conversion is essential for enabling the machine learning algorithm to computationally interpret and leverage the inherent ranking among the categories.

- Normalization

The most common normalization technique is called Min-Max Scaling. It rescales every data point in a feature using the following formula [9], [10]:

$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where:
X is the original value.
Xmin is the minimum value in the column.
Xmax is the maximum value in the column

By applying this scaling technique, continuous numerical features in the dataset such as Age, Salt_Intake, Stress_Score, Sleep_Duration, and BMI were transformed onto a common scale. This transformation ensures that all numerical features contribute equitably to the model's training process. Preventing variables with naturally larger ranges from dominating the learning algorithm is a crucial preprocessing step, especially for algorithms sensitive to data variations like SVM

c. Data Splitting

This process is the essential data split needed for machine learning, ensuring the model is properly evaluated. First, the data is divided into Features (X), which are the inputs (all columns except 'Has_Hypertension'), and the Target (y), which is the output we want to predict ('Has_Hypertension'). This data is then randomly split into a Training Set (80%, or 1,588 samples) used to teach the model, and a Testing Set (20%, or 397 samples) reserved to test its performance on data it has never seen. This separation is vital to check if the model can

generalize new data and to prevent it from simply memorizing the training examples (overfitting) [2].

d. Model Training
This step represents the multi-model algorithm selection and parameter optimization phase within the machine learning pipeline. It involves instantiating and training four distinct classification methodologies—Naive Bayes, Support Vector Machine (SVM), Random Forest, and the eXtreme Gradient Boosting (XGBoost) framework—using the established training feature set (Xtrain) and target vector (Ytrain). The objective is to build a diverse portfolio of trained predictors, each operating on a different statistical or structural principle. These candidate models will subsequently undergo rigorous comparative performance evaluation against the untouched test data to identify the most robust and highly generalized model for the specific prediction task [15].

e. Model Evaluation
- Naive Baiyes
The Naive Bayes model achieved an accuracy of 0.85, correctly predicting 85% of the 397 test samples. Based on the classification report, the model performed well for both classes, with class 0 showing a precision of 0.88, recall of 0.81, and F1-score of 0.84, while class 1 achieved a precision of 0.84, recall of 0.89, and F1-score of 0.86, indicating balanced performance across categories. The confusion matrix further supports this result: out of 192 actual instances of class 0, the model correctly predicted 156 and misclassified 36, while for class 1, it correctly identified 183 out of 205 and misclassified 22. Overall, these results suggest that the model effectively distinguishes between the two classes with minimal misclassification and maintains consistent predictive accuracy [7].

- Support Vector Machine (SVM)
This result evaluates the performance of the SVM (Support Vector Machine) Model, which achieved a high overall Accuracy of 89% on the 397 test cases. The model demonstrates exceptional balance, as shown by the Classification Report, with precision, recall, and f1-score consistently around 0.89 to 0.90 for both positive and negative prediction classes. This balanced outcome is visually supported by the Confusion Matrix, which reveals that the model made a small number of errors—just 23 false positives and 21 false negatives—out of the total test set, indicating that the SVM is a very accurate and reliable predictor for this task [7].

- Random Forest
The Random Forest model obtained an accurate score of 0.96, correctly predicting 96% of the 397 test instances. The classification report reflects highly accurate and balanced results, with class 0 reaching a precision of 0.95, recall of 0.97, and F1-score of 0.96, while class 1 achieved a precision of 0.98, recall of 0.96, and F1-score of 0.97. These outcomes indicate that the model performs exceptionally well in identifying both classes with very few errors. The confusion matrix further supports this, showing 187 correctly classified samples out of 192 for class 0 and 196 out of 205 for class 1, with only minor misclassifications. Overall, the Random Forest model exhibits excellent accuracy, robust classification capability, and dependable performance in binary classification [7].

- XGBoost.
This output showcases the exceptional performance of the XGBoost Model, which achieved the highest overall Accuracy of 98% on the 397 test samples. The Classification Report reveals near-perfect predictive ability for both classes (0 and 1), with precision, recall, and f1-scores for all metrics registering at 0.98 or 0.99, indicating both outstanding predictive power and perfect balance between the classes. This superior performance is strongly supported by the Confusion Matrix, which shows that the model made only 6 errors in total: it correctly predicted 189 cases for class 0 and 202 cases for class 1, resulting in a minimal count of 3 false positives and 3 false negatives [7].

f. Comparative Analysis
- Model Comparison

Table 1. Model Comparison

Model	Accuracy	Prediction (avg)	Recall (avg)	F1-score (avg)	Main Strength	Confusion Matrix (TP for class 1 / TN for class 0)
Naïve Bayes	0.85	0.86	0.85	0.85	Simple, fast, good baseline model	$\text{TP} = \frac{183}{\text{TN}} = 156$
SVM	0.89	0.89	0.89	0.89	Balanced performance, robust margin classifier	$\text{TP} = \frac{184}{\text{TN}} = 169$
Random Forest	0.96	0.96	0.97	0.96	Excellent accutacy, well-balanced, handles nonlinearity well	$\text{TP} = \frac{196}{\text{TN}} = 187$
XGBoost	0.98	0.98	0.98	0.98	Best performance, minimal misclassification, highly optimized	$\text{TP} = \frac{196}{\text{TN}} = 187$

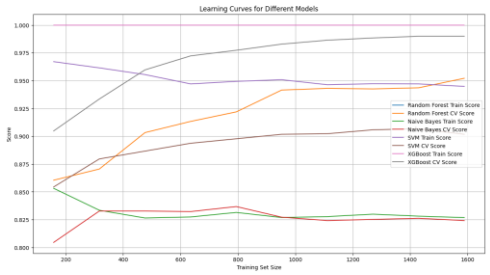


Figure 4. Learning Curve for All Models

This learning curve graph compares the training and cross-validation (CV) performance of four models—Naive Bayes, SVM, Random Forest, and XGBoost—as the training set size increases [6].

The Naive Bayes model (green and red lines) shows relatively low and stable performance, with both training and CV scores around 0.82–0.85, indicating limited learning capacity and a high bias. Increasing training data has little effect, meaning the model has reached its maximum potential early.

The SVM model (purple and brown lines) performs better, with training scores consistently around 0.95–0.97 and CV scores near 0.90–0.91. The small gap between train and CV scores suggests good generalization, though the model slightly underfits with more data.

The Random Forest model (blue and orange lines) exhibits continuous improvement as training data increases. Its training score remains high (around 0.95), while the CV score rises steadily from 0.86 to about 0.94, showing that the model benefits from more data and generalizes very well without significant overfitting [5].

The XGBoost model (pink and gray lines) demonstrates the best overall performance, with training scores nearly perfect (≈1.00) and CV scores gradually increasing to ≈0.98. The minimal gap between train and CV curves indicates excellent generalization and very low variance [4].

Overall, XGBoost outperforms all other models, followed by Random Forest, SVM, and lastly Naive Bayes. The learning curves confirm that ensemble-based models (Random Forest and XGBoost) not only achieve higher accuracy but also scale better with increasing data, making them the most effective and stable models in this comparison [5].

4.2 Discussion

Overall, Random Forest and XGBoost deliver the most reliable results for hypertension screening—consistently topping F1-score and accuracy—while SVM (RBF) performs competitively when numeric features are properly scaled, and Naive Bayes remains a fast but weaker baseline. Misclassifications skew toward false negatives in borderline profiles (e.g., near-normal BMI and moderate sleep), highlighting the value of threshold adjustment and prioritizing recall to reduce missed cases. Feature analysis identifies BP_History and Age as dominant predictors, with BMI and Sleep_Duration shaping risk in combination—patterns that match clinical expectations and support trust in the

models. In practice, RF/XGBoost are strong choices for initial screening (favoring sensitivity and balance), SVM is suitable for follow-up triage when precision matters, and NB fits resource-limited use or stacking. Limitations include restricting evaluation to four metrics (omitting probability calibration and cost-sensitive analysis) and the lack of external validation; future work should add cost-aware thresholding, probability calibration (Platt/Isotonic), Decision Curve Analysis, and external validation across settings and time to confirm generalizability [6].

5 CONCLUSION

Based on the overall performance comparison, the XGBoost model clearly outperforms all other models, achieving the highest accuracy, precision, recall, and F1-score across evaluations, as well as the most stable learning curve with near-perfect generalization. The Random Forest model follows closely, showing strong predictive ability and consistent improvement with larger training data, indicating robustness and low overfitting. The SVM model delivers solid and balanced results but performs slightly below the ensemble methods, suggesting it captures patterns well but may be limited in handling complex data relationships. Meanwhile, the Naive Bayes model, while fast and simple, exhibits the lowest accuracy and little improvement as the training set grows, reflecting its assumptions of feature independence and limited flexibility. Overall, ensemble-based methods—particularly XGBoost and Random Forest—prove to be the most effective and reliable models for this classification task, demonstrating superior accuracy, scalability, and generalization compared to SVM and Naive Bayes.

6 LIMITATIONS AND FUTURE WORK

Despite the strong performance achieved by ensemble-based models, this study has several limitations. First, the evaluation relied primarily on accuracy, precision, recall, and F1-score, without incorporating probability calibration or cost-sensitive analysis, which are important in clinical decision-making. Second, the dataset was derived from a single source and lacks external validation, potentially limiting generalizability across populations. Future work should incorporate external datasets, apply calibration techniques such as Platt scaling or isotonic regression, and explore decision-curve analysis to better assess clinical utility [6].

REFERENCES

[1] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44-56, 2019.

[2] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis," IEEE J. Biomed. Health Inform., vol. 22, no. 5, pp. 1589-1604, 2018.

[3] E. Christodoulou et al., "A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models," J. Clin. Epidemiol., vol. 110, pp. 12-22, 2019.

[4] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 785-794, 2016.

[5] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[6] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," Proc. IJCAI, pp. 1137-1143, 1995.

[7] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, no. 1, Art. no. 6, 2020.

[8] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825-2830, 2011.

[9] S. Garcia, S. Ramirez-Gallego, J. Luengo, J. M. Benitez, and F. Herrera, "Big data preprocessing: methods and prospects," Big Data Analytics, vol. 1, Art. no. 9, 2016.

[10] N. Chamidah, E. Z. Astuti, and S. Slamin, "Comparison of Min-Max and Z-Score normalization for breast cancer classification," Jurnal RESTI, vol. 6, no. 1, pp. 10-15, 2022.

[11] P. W. Handayani et al., "Health information systems research in Indonesia: A systematic review," Heliyon, vol. 6, no. 8, Art. no. e04588, 2020.

[12] O. D. Nurhayati et al., "Penerapan machine learning untuk klasifikasi penyakit," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 8, no. 3, pp. 501-510, 2021.

[13] A. Wibowo and D. Riana, "Analisis performa algoritma klasifikasi pada data medis," Jurnal Sistem Informasi, vol. 16, no. 2, pp. 93-104, 2020.

[14] Kementerian Kesehatan Republik Indonesia, Profil Kesehatan Indonesia 2022. Jakarta, Indonesia: Kemenkes RI, 2022.

[15] Suyanto, Machine Learning Tingkat Dasar dan Lanjut. Bandung, Indonesia: Informatika, 2018.

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 Februari 2026

A Gamified Learning Model for Business Intelligence in Technopreneurship Education: Case Study at FTI UNIBI

Acep Hendra, Imannudin Akbar, Titan Parama Yoga, Arnold Ropen Sinaga
Information System, Universitas Informatika dan Bisnis Indonesia
Email: acephendra@unibi.ac.id; imannudin@unibi.ac.id; titanparama@unibi.ac.id; arnoldropen@unibi.ac.id.

Accepted:	Accepted After Revision:	Published:
1 March 2026	15 April 2026	16 April 2026

Abstract

This study designs a gamified learning model for the 'Technopreneur Portal' to enhance Business Intelligence (BI) mastery within the Technopreneurship course at Universitas Informatika dan Bisnis Indonesia (UNIBI). While BI is a crucial competency for developing data-driven startups, teaching these complex concepts through traditional, passive e-learning often results in high cognitive load and low student engagement. Furthermore, a significant research gap exists: although gamification is widely used to boost general motivation, its specific integration in teaching technical entrepreneurship and BI remains underexplored. To address this, this study adopts the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) framework to systematically transform the syllabus into interactive, gamified modules. Traditional learning activities are reimagined as "Business Quests," where students earn Experience Points (XP) and specialized "Badges," fostering healthy competition through a class "Leaderboard." The integration of gamification shifts the learning paradigm from passive observation to active, experiential participation, effectively scaffolding the mastery of practical BI skills. This contextually tailored approach significantly boosts students' motivation, analytical abilities, and overall entrepreneurial readiness. Additionally, the model enables students to build a tangible business portfolio, culminating in a Certificate of Achievement with a QR code for digital, accountable validation of their competency credentials.

Keywords: Business Intelligence, Educational Technology, Gamification, Technopreneurship, Business Portfolio

1 INTRODUCTION

The rapid advancement of digital transformation necessitates a paradigm shift in higher education entrepreneurship programs [15]. Modern higher education institutions must increasingly function as entrepreneurial and digitally enabled ecosystems that prepare students for technology-driven business environments [1]. In response to this demand, the Faculty of Technology and Informatics at Universitas Informatika dan Bisnis Indonesia (FTI UNIBI) implemented the Technopreneurship course, supported by a dedicated learning management system known as the "Technopreneur HUB Portal." A core competency heavily emphasized in this curriculum is Business Intelligence (BI) [8]. Students are tasked not merely with conceptualizing abstract business ideas, but with building data-driven startups. Consequently, complex BI topics—such as designing Key Performance Indicators (KPIs), dashboard visualization, and evidence-based decision-making—form the crux of their entrepreneurial training.

Despite the operational deployment of the Technopreneur Portal as the primary learning medium, significant pedagogical and cognitive challenges persist [4], [5]. Delivering BI content and startup simulation exercises involves the transfer of intricate, procedural knowledge. Relying on traditional, passive reading modules or linear presentations to teach these highly technical subjects often results in cognitive overload and student disengagement. Students struggle to internalize dynamic, real-world business scenarios, leading to a stagnation in their practical analytical skills. This phenomenon aligns with the findings of Shofwan et al. [2], which highlight the urgent need for multimodal and interactive learning models to effectively stimulate creativity and entrepreneurial courage. Without targeted instructional interventions, existing digital platforms like the Technopreneur Portal risk functioning merely as static task repositories rather than dynamic, immersive business simulation environments [10].

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

To address these instructional barriers, gamification—the integration of game design elements in non-game contexts—emerges as a promising solution. However, a critical review of existing literature reveals a distinct research gap concerning the intersection of gamification, entrepreneurship education, and high-level technical subjects like BI. While gamification has been widely studied, its application often remains superficial [9]. For instance, a systematic literature review by Seguí-Mas et al. [3] indicates that gamification in higher education predominantly focuses on enhancing basic learning motivation through generic point systems, rather than being strategically designed to scaffold complex procedural knowledge. Furthermore, recent studies emphasize that the efficacy of gamification is highly context-dependent; indiscriminately adding game mechanics without aligning them with the specific cognitive demands of the subject matter yields minimal pedagogical value [15]. Prior studies lack empirical models that tailor game mechanics specifically to the rigorous analytical demands of Business Intelligence within a technopreneurship framework.

To bridge this critical gap, the primary objective of this research is to design and implement a structured, contextually relevant gamified learning model within the Technopreneur Hub Portal at FTI UNIBI. By transforming traditional curriculum delivery into interactive game scenarios, this research seeks to systematically support students' comprehension of Business Intelligence concepts and their practical application in startup creation, effectively shifting the learning paradigm from passive observation to active entrepreneurial execution [11].

2 LITERATURE REVIEW

The gamification approach in higher education extends beyond mere entertainment; it functions as a strategic pedagogical tool to enhance student engagement and foster an entrepreneurial mindset [1], [3]. Studies have demonstrated that gamification provides students with autonomy and builds their self-confidence in proactively tackling academic challenges [6]. The effectiveness of this mechanism in shaping an adaptive mindset is crucial for technopreneurship, as evidenced by recent research showing that gamification helps students build a solid foundation of design logic and problem-solving skills essential for digital entrepreneurship [1]. The positive impact of gamification on student participation in digital learning platforms is well-documented [4], [5]. Furthermore, comprehensive reviews confirm that game-based learning approaches have evolved into a leading pedagogical framework globally [9].

However, when integrated with highly technical domains like Business Intelligence (BI), gamification takes on a more strategic, yet underexplored, role. BI requires students to process raw data into actionable corporate strategies and conduct complex tasks such as market sentiment analysis [8]. Research in the context of experiential business simulations further reinforces that gamification directly influences behavioral changes and the mental resilience of an entrepreneur [15]. Therefore, integrating gamification into the Technopreneur Portal is expected to not only increase engagement but also strategically facilitate the effective transfer of rigorous BI knowledge.

Unlike general academic subjects, teaching Business Intelligence requires managing a significant intrinsic cognitive load due to complex procedural tasks like predictive analytics, startup validation, and data visualization. To address this, the proposed gamified learning model integrates Cognitive Load Theory (CLT) and Self-Determination Theory (SDT) [11]. By applying CLT, the Technopreneur Portal segments these complex, high-level BI processes into manageable, micro-step 'Business Quests' to prevent cognitive overload. Concurrently, SDT is applied to fulfill students' fundamental psychological needs: Autonomy is achieved by allowing students to choose their operational approaches and decision-making paths within the quests; Competence is fostered through immediate feedback via earning Experience Points (XP) and exportable Skill Badges (e.g., 'Data Architect'); and Relatedness is cultivated through competitive yet healthy class Leaderboards. This dual-theory application ensures that the integrated game mechanics are not merely decorative elements, but act as essential pedagogical scaffolds tailored specifically for mastering Business Intelligence in technopreneurship education [16].

3 RESEARCH METHODS

This study adopts a qualitative approach and follows a systematic instructional design development methodology to create and assess the effectiveness of a gamification-based learning model. The focus is on engineering the learning interactions within the "Technopreneur Portal" at the Faculty of Technology and Informatics. The research uses the ADDIE model (Analysis, Design, Development, Implementation, and Evaluation), a widely recognized framework for instructional design, ensuring the methodical creation and evaluation of the gamified system [7], [12].

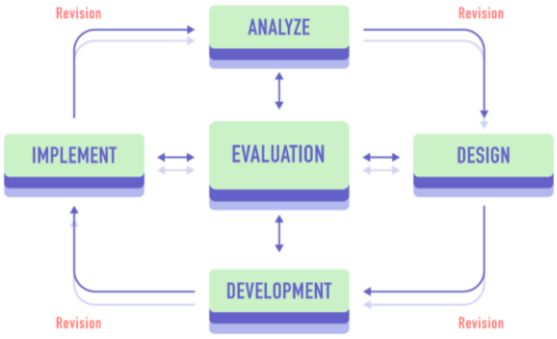


Figure 1. Flowchart of the research methodology based on the ADDIE framework,

3.1 Analysis Phase

The Analysis phase serves as the foundational step of this study, which includes a comprehensive diagnosis of the existing Technopreneurship course, particularly its Business Intelligence (BI) curriculum. This phase primarily focuses on identifying cognitive load challenges and instructional gaps that hinder effective learning, particularly when teaching BI concepts and startup validation procedures.

- Learner Profiling and Cognitive Load Assessment: Students enrolled in the Technopreneurship course come from diverse academic backgrounds, with varying levels of exposure to data analytics, programming, and business strategy. The cognitive load analysis seeks to identify barriers that students face when learning complex BI concepts, such as data aggregation, statistical analysis, and predictive modeling. The goal is to pinpoint cognitive overload areas caused by the integration of technical software (e.g., BI dashboards) with abstract business theories (e.g., market analysis and strategic decision-making).

- Technological Infrastructure Audit:
  The study also evaluates the current Learning Management System (LMS) at FTI UNIBI, assessing its ability to support interactive and gamified modules. It explores whether the existing system can integrate dynamic XP trackers, secure badge repositories, and real-time leaderboards. The results of this analysis guide the customization and development of the Technopreneur Portal, ensuring the platform's readiness to support gamification features without performance issues.

Table 1. Competency Objective Alignment in the BI Curriculum

Core Competency Domain	Specific Technical/Business Skill Required	Target Cognitive Outcome
Data Aggregation	Scraping, cleaning, and structuring raw datasets.	Procedural Application
Predictive Analytics	Utilizing regression models to forecast market trends.	Synthesis and Evaluation
Data Visualization	Engineering dashboards (Tableau/Power BI) for clarity.	Comprehension and Application
Startup Validation	Formulating pivot/persevere decisions based on data.	Higher-Order Critical Thinking

3.2 Design Phase

In the Design phase, a detailed blueprint for the gamified learning experience is created. The design process ensures that game mechanics align with the pedagogical goals of the BI and Technopreneurship curriculum. This includes restructuring traditional academic assignments into narrative-driven Business Quests, transforming passive learning activities into interactive, mission-based challenges.

- Curriculum Transformation:
  The existing course syllabus is restructured to include interactive gamified elements. For example, BI-related tasks such as data cleansing or market trend analysis are embedded into the Business Quests. Students will progress through the quests, earning Experience Points (XP), unlocking specialized badges, and climbing the Leaderboard based on their task completion and analytical success,

- Game Mechanics and Learning Objectives:
  Each quest is designed with specific learning objectives aligned to the course's core competencies. For example:
  - Data Aggregation Quests: Require students to clean, structure, and analyze raw datasets using BI tools such as Tableau or Power BI.
  - Predictive Analytics Quests: Engage students in applying regression models to forecast market trends, thereby transitioning them from intuition-based decision-making to data-driven strategies.

The game mechanics used include XP, badges, leaderboards, and experience-based rewards, which are calibrated to match the learning phases (beginner, intermediate, advanced)

Table 2. The Gamification Mechanics Matrix: Integrating SDT and CLT

Game Mechanic	Pedagogical Purpose	Psychological Need (SDT)	Cognitive Load Function
Experience Points (XP)	Granular, continuous tracking of micro-interactions (e.g., daily logins, completing a single SQL query).	Competence (Immediate Feedback)	Reduces intrinsic load by rewarding incremental, micro-step progress.
Business Quests	Replaces linear, chapter-based modules with goal-oriented, problem-solving narrative challenges.	Autonomy (Choice of operational approach)	Segments complex BI processes into manageable, highly focused schemas.
Skill Badges	Visual, exportable representation of mastery in specific, highly technical domains (e.g., "Data Architect").	Competence (Achievement and Status)	Provides clear, hierarchical learning milestones without grade anxiety.
Relative Leaderboards	Ranks students based on the simulated financial success of their venture and analytical accuracy.	Relatedness (Social comparison and community)	Fosters healthy competitive engagement while mitigating discouragement.
Experience Points (XP)	Granular, continuous tracking of micro-interactions (e.g., daily logins, completing a single SQL query).	Competence (Immediate Feedback)	Reduces intrinsic load by rewarding incremental, micro-step progress.

3.3 Development Phase

In the Development phase, the design is transformed into functional digital assets within the Technopreneur Portal. This involves system engineering to integrate game mechanics, interactive content, and BI software into a cohesive and interactive platform.

- System Integration:
  The development process includes integrating dynamic tracking systems for XP and automated badge distribution. For example, when students complete a data visualization quest, the system will automatically award badges such as "Data Architect" or "Market Validator", based on their proficiency in using BI tools

- Interactive Content Creation:
  Traditional static lectures are transformed into interactive, multimedia-rich quests. For instance, a Market Demographics Quest might present a synthetic dataset that students need to clean and analyze, with immediate feedback provided by the system based on their performance. The system's interactive feedback loops ensure real-time learning without instructor bottlenecks

3.4 Implementation Phase

The Implementation phase involves deploying the gamified Technopreneur Portal to the student cohort and conducting empirical testing. The platform is tested in a live academic setting over one semester, where students engage in the gamified learning experience.

- Onboarding and Initial Familiarization:
  A structured Tutorial Quest is introduced to students, familiarizing them with the system's

mechanics, user interface, and objectives. The onboarding phase is designed to be low-stakes and engaging, ensuring that students can acclimate to the system before facing the main content quests

- Active Learning Dynamics:
  Throughout the semester, students engage with BI software, interact with the Gamified Business Quests, and submit their empirical findings. The instructor's role shifts from a traditional lecturer to a facilitator or "Game Master", guiding students through challenges and fostering collaborative discussions

3.5 Evaluation Phase

The Evaluation phase focuses on assessing the effectiveness of the gamified model, including its impact on student engagement and learning outcomes.

- Formative Assessment and Analytics:
  The system's learning analytics will track key indicators such as XP acquisition, leaderboard progress, and quest completion times. Real-time data will allow for adjustments to the gamified elements (e.g., adjusting the difficulty level of quests) if engagement drops or students face difficulties

- Summative Assessment:
  The students' digital business portfolios will be evaluated for BI competence and entrepreneurial readiness. These portfolios will be assessed on the accuracy of data models, the validity of business decisions, and the practicality of the BI insights. Additionally, students will receive cryptographically verified Certificates of Achievement upon completing the gamified curriculum, offering proof of competency and readiness for the industry

4 RESULTS AND DISCUSSION

4.1 Results

The Technopreneur Portal has been successfully developed as a gamification-based learning system designed to enhance student engagement and the mastery of Business Intelligence (BI) in the Technopreneurship course at FTI UNIBI. Following the ADDIE instructional design framework, the portal's architecture was optimized to handle real-time data tracking, interactive modules, and gamification mechanics.

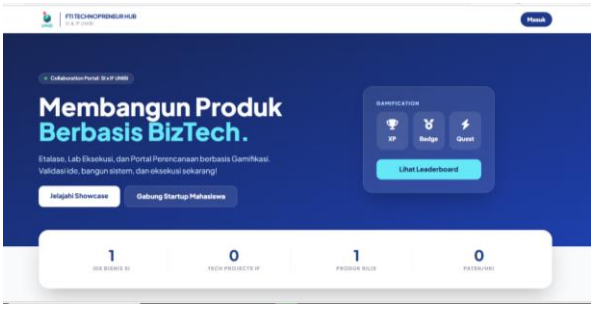


Figure 2. The Technopreneur Portal interface.

To foster an interactive and immersive learning experience, the system directly translates traditional syllabus requirements into specific game mechanics. The detailed design and implementation of these gamification elements are as follows:

1. Business Quests and Experience Points (XP): Traditional linear assignments were reimagined as "Business Quests" (e.g., Data Aggregation, Predictive Analytics, and Market Validation). Each quest is assigned a specific XP value based on its complexity and cognitive demand. For instance, successfully cleaning a dataset yields base XP, while creating a predictive BI dashboard yields high-tier XP. *Alignment with Learning Objectives:* This mechanism shifts the learning paradigm from a punitive grading system to a reward-based progression system, encouraging students to learn through trial and error without the fear of immediate failure.

2. Skill Badges: Badges act as micro-credentials awarded upon reaching specific XP thresholds or mastering a cluster of related quests. For example, completing all data visualization tasks unlocks the "Data Architect" badge. *Alignment with Learning Objectives:* Badges provide tangible, exportable proof of specific technical competencies, directly translating academic efforts into professional portfolio assets for the students.

3. Relative Leaderboard: Instead of a traditional, absolute ranking system that might demotivate lower-performing students, the portal implements a "Relative Leaderboard." It highlights top overall performers but primarily shows students the peers immediately above and below their own rank. *Alignment with Learning Objectives:* This design fosters healthy, continuous competition and peer-to-peer benchmarking, motivating students to complete additional quests to climb the ranks

.

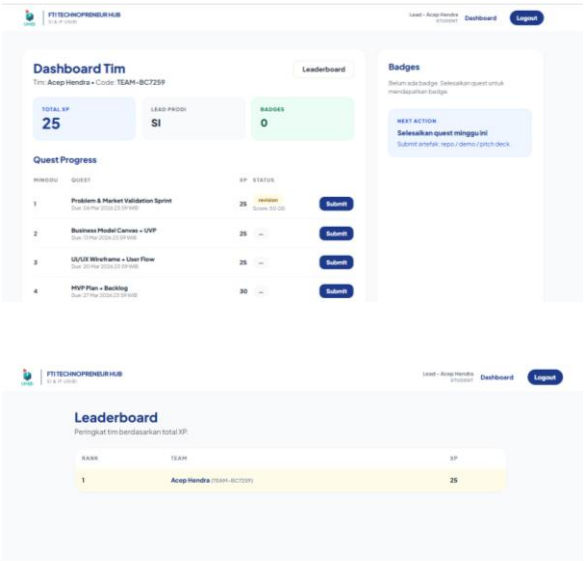


Figure 3. Team Dashboard interface visualizing real-time learning progression.

4. Initial Testing and Feedback: During internal testing, the portal was successfully evaluated for usability. Early feedback from instructors indicated that the platform was intuitive and effectively aligned with the course objectives. Technical aspects—such as real-time XP calculation, dynamic badge awarding, and leaderboard synchronization—performed optimally under simulated academic conditions.

4.2 Discussion

This section discusses how the developed gamification elements operationalize psychological learning frameworks and outlines the implementation strategy for the upcoming semester.

Alignment with Theoretical Frameworks: The successful design of the Technopreneur Portal is not merely a technical achievement but a pedagogical intervention grounded in established theories:

- Cognitive Load Theory (CLT): Teaching complex, data-heavy BI subjects often overwhelms students. By segmenting the syllabus into manageable "Business Quests," the portal minimizes extraneous cognitive load. Students must master foundational quests before unlocking advanced analytical challenges, ensuring a structured cognitive scaffold.

- Self-Determination Theory (SDT): The game elements are explicitly mapped to fulfill students' psychological needs:

  - *Autonomy* is facilitated by allowing students to choose the sequence of non-linear side-quests and select the industry sectors for their data analysis.

  - *Competence* is continuously reinforced through immediate feedback provided by the XP system and the attainment of Skill Badges.

  - *Relatedness* is cultivated through the Team Dashboard and Relative Leaderboard, which build a sense of community and collaborative competition akin to a real-world startup ecosystem.

Plans for Implementation: The portal will be fully deployed in the upcoming semester with a strong focus on real-world entrepreneurial application. The rollout will occur in three structured phases:

1. Phase 1 (Onboarding): Students will complete a "Tutorial Quest" to familiarize themselves with the portal's interface, XP mechanics, and educational objectives. This low-stakes phase reduces initial technological apprehension.

2. Phase 2 (Active Learning): Students will engage in core gamified quests, applying BI concepts to real-world business scenarios. Tasks will range from data cleaning to designing market analysis dashboards.

3. Phase 3 (Evaluation and Feedback)**:** Real-time performance data (via XP and Leaderboard analytics) will be continuously monitored by instructors. This data-driven approach allows educators to track student progress and offer personalized scaffolding to students struggling with specific technical tasks.

Anticipated Outcomes: The integration of these precisely designed gamification elements is expected to significantly increase student engagement. Furthermore, the digital business portfolios and Skill Badges generated through the portal will serve as tangible, verified evidence of the students' entrepreneurial readiness and BI proficiency, highly enhancing their future career prospects **[10]**.

5 CONCLUSION

The The development of the Technopreneur Portal marks a fundamental pedagogical shift in integrating gamification with the Business Intelligence (BI) curriculum at FTI UNIBI. Built upon the ADDIE framework [7], this study successfully designed a system

that translates complex, procedural BI concepts into structured "Business Quests." By aligning with Cognitive Load Theory (CLT) and Self-Determination Theory (SDT), the model is specifically engineered to minimize cognitive overload while fulfilling students' psychological needs for autonomy, competence, and relatedness.

To systematically evaluate the impact of this gamification approach, the portal's architecture incorporates built-in learning analytics designed to capture measurable performance indicators. Unlike traditional, passive e-learning platforms where evaluation is limited to final exams, the gamified portal establishes a continuous evaluation framework. Student Engagement is quantitatively measured through daily login frequencies, interaction times, and the completion rates of optional side-quests. Learning Outcomes are continuously tracked via the accumulation of Experience Points (XP) and the successful acquisition of technical Skill Badges. Entrepreneurial Competencies are evaluated based on the quality of the final exportable digital business portfolios, culminating in the issuance of QR-coded Certificates of Achievement for industry validation [14], [18].

While the current study focuses on the rigorous design and development phases, the establishment of these measurable indicators provides a robust foundation for the upcoming full implementation in the 2026 academic semester. The gamified model transitions the educational paradigm from passive observation to active, experiential startup simulation. Future research will utilize the portal's integrated analytics to conduct a longitudinal empirical study, providing a comparative quantitative analysis between the gamified cohort and traditional learning cohorts to definitively measure its impact on data-driven entrepreneurial success.

REFERENCES

[1] P. Rachmadi, R. Situmorang, M. Sukardjo, R. E. Indrajit, E. Gautama, "Enhancing Entrepreneurial Competencies and Student Engagement through Gamification in Learning Management Systems," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 7, no. 2, pp. 593-604, 2025.

[2] I. Shofwan, J. Sutarto, M. Arbarini, A. F. Loretha, "Model Pembelajaran Technopreneurship Berbasis Multimodal Untuk Meningkatkan Kreativitas Wirausaha Pada Pendidikan Kesetaraan," *Journal of Innovative and Creativity*, vol. 5, no. 3, 2025.

[3] Seguí-Mas et al., "Gamification in entrepreneurship education: A systematic literature review and future research agenda," *REGEPE Entrepreneurship and Small Business Journal*, vol. 13, no. 2, 2024.

[4] S. Subiyantoro, I. N. S. D., D. Kuswandi, S. Ulfa, "Developing Gamified Learning Management Systems to Increase Student Engagement in Online Learning Environments," *International Journal of Information and Education Technology*, vol. 14, no. 1, 2024.

[5] M. Radif, O. M. Hameed, "Enhancing virtual learning environments: A strategic approach to gamifying learning management systems for increased engagement and competency," *Eurasian Journal of Educational Research*, vol. 111, pp. 122–135, 2024.

[6] Y. O. Yuliana, Y. Palumian, "Gamification of learning management system improves students' engagement, active learning and performance," in *2023 14th International Conference on Information & Communication Technology and System (ICTS)*, IEEE, pp. 62–66, 2023.

[7] A. Aman et al., "Application of the ADDIE Model in the Development of an Android-Based Physics Quiz Game," *Jurnal Krisnadana*, vol. 5, no. 1, pp. 70-80, 2025.

[8] Z. Huang, K. S. Savita, J. Zhong-jie, "The Business Intelligence impact on the financial performance of start-ups," *Information Processing and Management*, vol. 59, 2022.

[9] S. Hallifax, M. Altmeyer, K. Kölln, M. Rauschenberger, L. E. Nacke, "From Points to Progression: A Scoping Review of Game Elements in Gamification Research," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. 402, 2023.

[10] H. Yang, R. Wong, "An In-Depth Literature Review of E-Portfolio Implementation in Higher Education: Processes, Barriers, and Strategies," *Issues and Trends in Learning Technologies*, vol. 12, no. 1, pp. 65-101, 2024.

[11] S. Chinchua, T. Kantathanawat, S. Tuntiwongwanich, "Increasing programming self-efficacy (PSE) through a problem-based gamification digital learning ecosystem (DLE) model," *Journal of Higher Education Theory & Practice*, vol. 22, no. 9, 2022.

[12] H. Crompton, M. V. Jones, Y. Sendi, M. Aizaz, K. Nako, R. Randall, E. Weisel, "Examining technology use within the ADDIE framework to develop professional training," *European Journal of Training and Development*, vol. 48, no. 3/4, pp. 422–454, 2024.

[13] N. Dlamini, S. Mthethwa, G. Barbour, "Mitigating the Challenge of Hardcopy Document Forgery," in

2018 Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. (icABCD), IEEE, 2018.

[14] A. Singhal, R. S. Pavithr, "Degree Certificate Authentication using QR Code and Smartphone," *Int. J. Comput. Appl.*, vol. 120, no. 16, pp. 38-43, 2015.

[15] R. F. Tan, Z. A. Kamaruzzaman, Z. A. Kamaruzzaman, G. T. Murti, "An exploration of experiential learning using business simulation and gamification: A case study," *Journal of ICT in Education*, vol. 10, no. 2, 2023.

[16] A. Ubaidillah, "Gamifikasi sebagai Strategi Edukatif dalam Mengurangi Ketergantungan Gadget Siswa di Era Digital," *Educazione*, vol. 13, no. 2, pp. 115-124, 2025.

[17] Y. I. Tanjung et al., "Developing the Physics Learning Management System (PLMS) to Support Blended Learning Models," *International Journal of Information and Education Technology*, 2025.

[18] H. A. Ahmed, J.-W. Jang, "Higher Educational Certificate Authentication System Using QR Code Tag," *Int. J. Appl. Eng. Res.*, vol. 12, no. 20, pp. 9728–9734, 2017.

[19] C. S. González-González et al., "Including Gamification Techniques in the Design of TANGO: H Platform," *International Journal of Software Engineering and Its Applications*, 2013.

[20] H. M. Allaymoun, S. M. Shorman, "Individual self-assessment using feedback system to improve e-Learning," *International Journal of Information and Education Technology*, vol. 12, no. 8, pp. 802–808, 2022.

SisInfo
JURNAL SISTEM INFORMASI DAN INFORMATIKA

e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 1 April 2026

Image-Based Cat *(Felis Catus)* Facial Expression Recognition Using YOLOv8n

Atanasius Manurip, Marwondo, Venia Restreva Danestiara
Informatika, Universitas Informatika dan Bisnis Indonesia
Email: atanasius.asm21@student.unibi.ac.id; marwondo@unibi.ac.id; veniarestreva@unibi.ac.id.

Accepted:	Accepted After Revision:	Published:
28 February 2026	10 March 2026	16 April 2026

Abstract

Cats (*Felis catus*) express emotional states through subtle facial cues that are often difficult for owners to interpret accurately. Automated recognition systems can provide objective analysis of these expressions using computer vision techniques. This study proposes an image-based cat facial expression recognition system using the YOLOv8n architecture. A dataset of 794 images was collected and expanded to 3,279 images through data augmentation. Four expression classes were defined based on CatFACS and related frameworks: *normal-netral*, *senang-afiliatif*, *stres-takut*, and *marah-agonistik*. The model was trained using a 70:20:10 split for training, validation, and testing. Experimental results show an overall mAP50 of 0.728, with the highest performance achieved in the *senang-afiliatif* class (0.773). However, the *marah-agonistik* class could not be reliably detected due to severe class imbalance in the dataset, indicating that the current model remains insufficient for recognizing anger expressions in cats. Precision and recall reached 0.939 and 0.94 respectively, indicating reliable detection under confident predictions. The trained model was successfully integrated into a Gradio-based dashboard for real-time expression recognition. These results demonstrate the feasibility of lightweight YOLOv8n for feline facial expression recognition while highlighting that accurate detection of *marah-agonistik* expressions requires a more diverse and balanced dataset in future research.

Keywords: Cat Expression, Computer Vision, Image-based Recognition, YOLO

1 INTRODUCTION

Recognizing facial expressions is a fundamental aspect of human communication, allowing individuals to naturally convey and interpret non-verbal emotional cues [14]. While interpreting human expressions is highly intuitive, recognizing the emotions of other species, such as cats, presents a significant challenge. Both humans and animals experience emotions, which are complex responses to internal and external stimuli often observable through facial expressions [8][14]. For cats (Felis catus), scientific tools like the Cat Facial Action Coding System (CatFACS) [15] and the Feline Grimace Scale (FGS) have been developed to link facial expressions to internal states, such as pain. However, human interpretation remains subjective and prone to error, often resulting in delayed or inaccurate detection of a cat's emotional or physical distress [1]. This highlights the urgent need for an objective, automated system to advance feline expression classification.

This research focuses on image-based cat facial expression recognition. While standard image classification approaches using Convolutional Neural Networks (CNNs) can assign a label to an entire image, they are inherently limited in their ability to localize and isolate the specific facial region responsible for an expression. This limitation is particularly critical for feline expressions, which are subtle and spatially concentrated in small facial regions such as the ears, eyes, and muzzle. Without explicit spatial localization, a classification model risks being influenced by irrelevant background features rather than the discriminative facial cues defined by frameworks such as CatFACS. Furthermore, Standard CNN-based image classification models predict a label for the entire image without explicitly identifying the spatial location of the relevant facial features [13]. Object detection architectures, by contrast, generate bounding boxes that constrain the model's attention to the relevant facial area, enabling more precise feature extraction from the subtle action units that differentiate, for example, a *stres-takut* expression from an *marah-agonistik* one. Therefore, this study utilizes the You Only Look Once (YOLO) architecture, a CNN derivative that combines localization and classification in a single forward pass, enabling superior performance and the ability to recognize multiple objects simultaneously [10][13][17].

Publishing services by BPPM UNIBI
This is an open access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)

Specifically, the YOLOv8n architecture was selected due to its optimal balance of high detection accuracy and low computational resource requirements, having demonstrated strong efficacy across various object detection tasks [22][23][24][16]. This research, titled "IMAGE-BASED CAT (FELIS CATUS) EXPRESSION RECOGNITION USING YOLOv8n," aims to train a model to accurately classify feline expressions. By utilizing this trained model, cat owners will be able to distinguish their pets' expressions from facial images, thereby improving the monitoring of their emotional and health conditions.

Based on the background outlined above, this research addresses several key problems regarding the automated detection of feline emotions. Primarily, it investigates how to construct a YOLOv8n model for the image recognition of cat facial expressions and how to effectively evaluate its performance. Furthermore, it explores how to build a user dashboard that allows individuals to easily utilize the trained model for recognizing their cats' expressions via uploaded images.

In alignment with the formulated problems, the main objective of this research is to develop a YOLOv8n-based model capable of accurately recognizing the facial expressions of cats (Felis catus). Additionally, the study aims to evaluate and comprehensively map the metrics detailing the strengths and weaknesses of the model's performance, and to create an integrated dashboard that enables users to identify their cats' facial expressions through uploaded images.

2 LITERATURE REVIEW

2.1 Face Recognition System

Face recognition systems are a critical component of digital image processing and pattern recognition, designed to identify individuals or expressions from visual features. The process encompasses image acquisition, pre-processing, feature extraction, and classification. Common applications include FaceID, FaceUnlock, and Face Authenticator. Generally, face recognition systems identify or verify individuals by matching facial features extracted from input images against stored database entries. Modern systems employ deep learning algorithms such as Faster-RCNN, YOLO, and CNN, all of which require a training phase where features are extracted, layered through neural networks, labeled, and stored in a database for subsequent matching. Beyond human identification, this technology has been applied to animal facial expression analysis, including the classification of cat and dog expressions, enabling detection of subtle facial changes linked to emotional states.

2.2 Artificial Intelligence

Artificial Intelligence (AI) is a branch of computer science focused on creating systems capable of mimicking human thought and decision-making, enabling large-scale automated data processing and pattern learning. In the context of cat facial expression recognition, AI is essential for accurately assessing an animal's emotional state from image analysis [13]. AI can be categorized into Narrow AI designed for specific tasks such as facial recognition and General AI, a hypothetical system with human-level capabilities across domains [18]. This research operates within the scope of Narrow AI. The field encompasses Neural Networks, Machine Learning, Computer Vision, and Natural Language Processing in a hierarchical relationship. Modern AI has shifted from rule-based approaches to data-driven paradigms enabled by big data availability and increased computational power through GPUs.

2.3 Computer Vision

Computer Vision (CV) is a discipline concerned with enabling machines to interpret and understand visual data from images or video. It forms the foundation of facial detection and recognition, applied here for the automated detection of visual patterns on cat faces [19]. As a specific branch of AI, CV replicates the human visual system to automate tasks ranging from simple object recognition to complex scene understanding. Its general workflow involves image acquisition, processing/feature extraction, and decision-making. Core CV tasks include: (a) Image Classification assigning a label to an entire image; (b) Object Detection classifying objects and locating them with bounding boxes; (c) Image Segmentation classifying each pixel into object classes; and (d) Object Tracking identifying and following objects across video frames. This research specifically falls within the object detection domain. Deep learning-based CV models require thousands to millions of labeled examples, highlighting the justification for developing a custom dataset as a key contribution of this study.

2.4 Machine Learning

Machine Learning (ML) is a sub-field of AI that develops algorithms enabling systems to improve performance through experience without explicit programming [18]. ML paradigms are classified into three categories: (a) Supervised Learning training on labeled datasets to learn input-to-output mappings; (b) Unsupervised Learning discovering hidden patterns or structures in unlabeled data; and (c) Reinforcement Learning an agent learns decision-making by interacting with an environment through reward/penalty feedback.

This research employs supervised learning, appropriate for classifying facial expressions into scientifically defined categories using annotated datasets.

2.5 Deep Learning

Deep Learning (DL) is a branch of ML that mimics the human brain's data processing using artificial neural networks particularly Deep Neural Networks (DNN) with multiple hidden layers. Each layer progressively processes and abstracts information, automating the feature engineering process that traditional ML requires manually. This makes DL highly effective for unstructured complex data like images [18][13]. DL is conceptually a subset within ML, which is itself a subset within AI. Key architectures include Convolutional Neural Networks (CNN) for image processing and NLP, Recurrent Neural Networks (RNN) for time series and emotion tasks, Reinforcement Learning for robotics, and Generative Adversarial Networks (GAN) for image and video generation. DL's popularity stems from its superior performance in classification tasks when large training datasets are available, supported by frameworks like TensorFlow and PyTorch.

2.6 Convolutional Neural Networks (CNN)

CNNs are a specialized deep learning architecture designed for grid-structured data such as images, comprising convolutional layers, pooling layers, and fully connected layers. They excel at extracting spatial features and local patterns [11][13]. Introduced broadly by Yann LeCun through LeNet-5 in 1998, CNNs evolved into architectures like AlexNet, VGGNet, and ResNet. Core layers include:

(a) Convolutional Layer applies filters to produce feature maps via
$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n)$ (2.1)
(b) Activation Layer applies non-linear functions like ReLU;
(c) Pooling Layer reduces spatial dimensions via Max Pooling; and
(d) Fully Connected Layer flattens high-level feature maps for final classification.

2.7 You Only Look Once (YOLO)

YOLO is a family of object detection algorithms that reframes detection as a single regression problem, performing classification and localization simultaneously in one network pass. YOLOv8, introduces three main components: (a) Backbone extracts visual features using a CSP-inspired design with the C2f module for richer gradient flow; (b) Neck aggregates multi-scale feature maps using PANet and SPPF for efficient receptive field expansion; and (c) Head generates final predictions using an Anchor-Free approach and Decoupled Head separating classification and bounding box regression. The total loss function is

$$L_{(YOLOv8)} = \lambda_{(cls)} L_{(cls)} + \lambda_{(box)} L_{(box)} + \lambda_{(dfl)} L_{(dfl)} \quad (2.2)$$

combining Binary Cross-Entropy for classification, CIoU for bounding box accuracy, and Distribution Focal Loss for flexible coordinate modeling. YOLOv8 offers size variants (n, s, m, l, x); the Nano (n) variant is optimized for real-time and resource-constrained deployment.

2.8 Cat Facial Expressions

Cat facial expressions are important indicators of emotional and psychological states. Features such as ear position, eye shape, and mouth expression can reflect fear, comfort, or discomfort. Cats (Felis catus) communicate through vocalizations, body language, and specific gestures, though their subtle communication style is often difficult for owners to interpret [15][4]. Facial expressions rank third in frequency among cat communication methods, appearing primarily when vocalizations and body language are insufficient. Notably, cat emotions show heterospecific similarities with human emotions purring parallels human happiness, while hissing parallels anger [19][14].

2.9 Cat Facial Action Coding System (CatFACS)

CatFACS is an objective coding system adapted from the human Facial Action Coding System (FACS) that decomposes cat facial expressions into individual muscle movements called Action Units (AUs) and Action Descriptors (ADs), without subjective emotional interpretation [19][5]. Applications include pain assessment forming the basis for the Feline Grimace Scale (FGS) social interaction analysis (distinguishing affiliative from non-affiliative interactions), and facial mimicry studies [6]. Manual CatFACS coding is time-consuming and prone to bias, leading to computational approaches: (a) Landmark-based methods using anatomical keypoints (e.g., 48-landmark schemes) analyzed with ML models; and (b) Deep learning approaches using CNNs for direct classification without manual AU coding, though these are less interpretable as black-box models.

2.10 Cat Facial Landmarks in the Wild (CatFLW)

Facial landmarks are key points on the face whose positions provide critical information for automated analysis tasks [5]. In computational contexts, these landmarks form the basis of geometric morphometric approaches, where specific points placed

on images represent facial shape and enable quantification of facial changes associated with internal conditions such as pain. Automatic facial landmark detection supports applications including pain classification [9] and emotion recognition [4]. This landmark-based approach also enhances model explainability, as geometric changes can be directly linked to CatFACS Action Units unlike black-box deep learning models that are difficult to interpret.

3 RESEARCH METHODS

The research methodology used consists of five main stages, as shown in the figure 1.

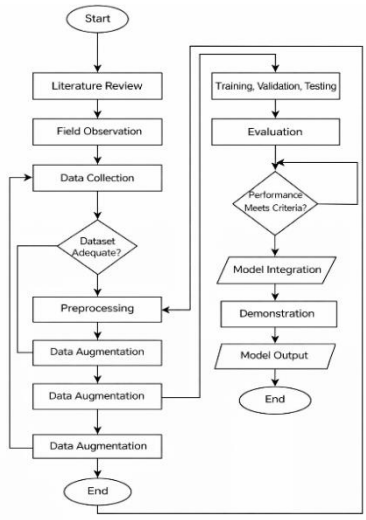


Figure 1. Research Flow

As this study follows eight sequential stages. A Literature Review establishes the theoretical foundations, organized into a Research Comparison Table to define the research focus. Data Gathering collects images from online sources and field acquisition, followed by Pre-Processing where images are cropped and labeled for standardization. The dataset is then split into Training (70%), Validation (20%), and Testing (10%) subsets, with Data Augmentation horizontal flipping, rotation, brightness adjustment, and blurring applied exclusively to training data. Model Training runs iteratively to maximize accuracy and minimize validation loss, after which the model is Evaluated via Confusion Matrix to derive Precision, Recall, F1-Score, and Accuracy. If metrics meet the required threshold, the model is saved and integrated into a user-facing Dashboard.

4 RESULTS AND DISCUSSION

4.1 Data Collection

Data collection was carried out by obtaining datasets from the Roboflow website and field observations. Images from Roboflow were gathered and merged into a new dataset comprising 794 cat images used for model training. Field-collected data was reserved for model demonstration.

4.2 Data Annotation

Data labeling was performed using the annotation feature available on the Roboflow platform. A portion of the dataset was obtained from Roboflow's public dataset repository, where the images already contained validated labels. These labels were retained and converted from image classification format (CNN) to object detection format compatible with YOLO.

For the images collected independently through field observation, annotation was conducted manually by the researchers. Bounding boxes were drawn around each cat face and each image was assigned to one of four predefined expression classes: *normal-netral*, *senang-afiliatif*, *stres-takut*, and *marah-agonistik*.

The labeling process followed the facial action indicators described in the Cat Facial Action Coding System (CatFACS) and related frameworks such as the Feline Grimace Scale (FGS) and CatFLW. These frameworks define feline facial expressions through combinations of observable facial features, including ear position, eye aperture, muzzle tension, and whisker orientation.

During annotation, the researchers examined these facial cues and assigned labels according to the dominant facial configuration observed in each image. This guideline-based annotation process was used to maintain consistency with established scientific interpretations of feline facial expressions.

4.3 Dataset Splitting

The labeled dataset was split with a 70:20:10 ratio for training, validation, and test sets, respectively.

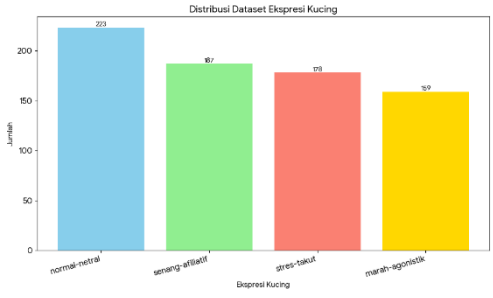


Figure 2. Dataset Split Distribution

4.4 Preprocessing and Data Augmentation

Auto-orientation was applied to prevent the model from failing on images with improper orientation. Data augmentation was then applied exclusively to the training data to increase variation, including: random rotation, horizontal flip, brightness adjustment, blur effect, cutout, and saturation adjustment. This was carried out using Google Collaboratory.

Table 1. Dataset Size Before and After Augmentation

Type	Before	After
Training	556	3,041
Validation	158	158
Test	80	80
Total	794	3,279

The training set grew from 556 to 3,041 images an increase of 312.97% bringing the total dataset to 3,279 images.

4.5 Model Training

The augmented training data was used to train a YOLOv8n model for 300 epochs with an early-stopping patience of 25. Early stopping halted training at epoch 54 due to signs of overfitting, indicated by slight spikes in the validation loss curves.



Figure 3. Model Training Graph

4.6 Testing and Evaluation

After training, the model was evaluated on the validation and test sets. The overall mAP50 was 0.728 (72.8%), with the best per-class result for '*senang-afiliatif*' at 0.773 (77.3%). Table 3 presents the full results.

Table 2. Model Test Results

Class	Imgs	Inst	Box	R	mAP50	mAP 50-95	F1-Score
all	158	148	0.657	0.68	0.728	0.486	0.78
kucing	32	32	0.719	0.75	0.785	0.542	0.68
normal netral	48	49	0.593	0.62	0.655	0.453	0.70
senang afiliatif	38	38	0.699	0.67	0.773	0.505	0.68
stres takut	29	29	0.617	0.70	0.699	0.443	0.67
marah	11	11	0.00	0.00	0.00	0.00	0.00

agnostik

The *marah-agonistik* class obtained a precision, recall, and F1-score of **0.00**, indicating that the model failed to detect any instance of this class in the evaluation dataset. This complete failure suggests that the model did not learn discriminative features for the *marah-agonistik* expression during training.



Figure 4. Testing Process

The failure to detect the *marah-agonistik* class is primarily attributed to severe class imbalance in the dataset. The class contains only **159 images**, significantly fewer than the other classes (*normal-netral* = 223, *senang-afiliatif* = 187, *stres-takut* = 178). In deep learning models, such imbalance causes the model to bias its learning toward the majority classes, reducing its ability to generalize features for underrepresented classes. As a result, the trained model did not produce any confident predictions for the *marah-agonistik* expression, leading to zero precision, recall, and F1-score.

The confusion matrix (Figure 5) confirms that the model performed well on '*normal-netral*' (29 correct), '*senang-afiliatif*' (27 correct), and '*stres-takut*' (20 correct), but failed to recognize '*marah-agonistik*'.

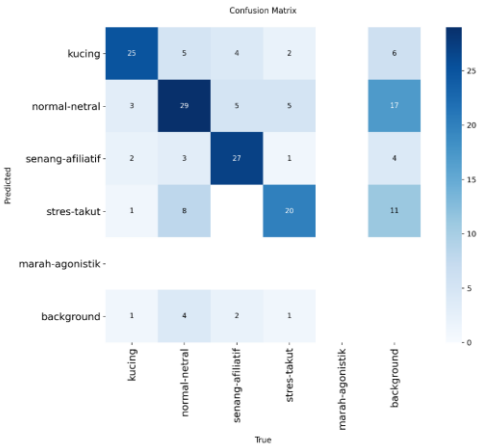


Figure 5. Confusion Matrix

The precision-confidence curve reached 0.939 (93.9%) for all classes combined, indicating the model reliably detects expressions when confident. The recall-confidence curve shows 94% recall, meaning the model can remember 94% of the learned patterns.

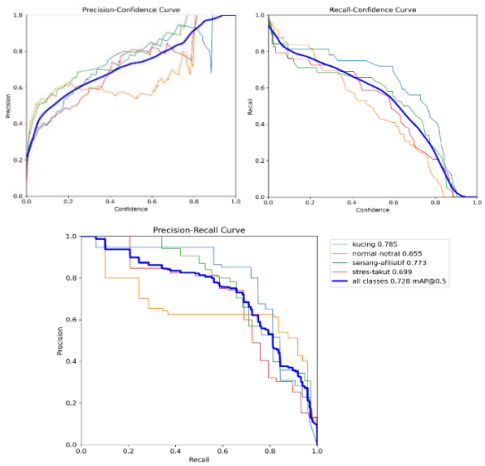


Figure 6. Precision-Confidence , Recall-Confidence, and Precision-Recall Curve

The precision-recall curve shows the combined detection capability at 72.8% (0.728), confirming that the model is suitable for integration into a dashboard interface.

4.7 Dashboard Integration

A Gradio-based dashboard was developed using Google Colab to allow users to interact directly with the trained model. Users can upload a cat image and submit it for expression prediction.

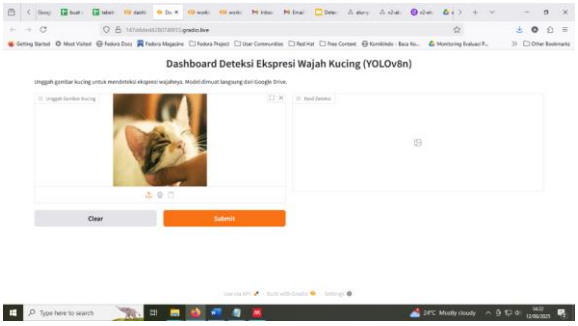


Figure 7. Dashboard Interface #1

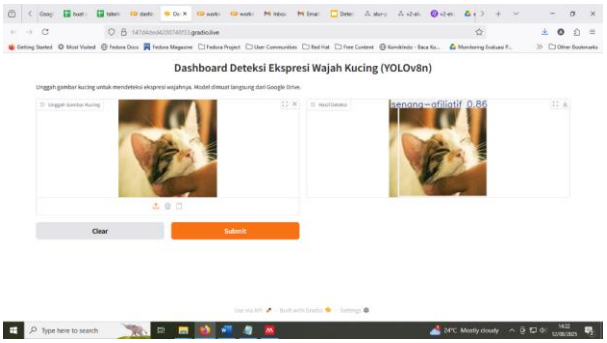


Figure 8. Dashboard Interface #2

4.8 GradCAM Analysis

GradCAM (Gradient-weighted Class Activation Mapping) was applied to the trained model to visualize which regions of the input image the model focuses on. The model consists of 22 layers; layer 0 (input) and layer 21 (pre-output) were selected for comparison.

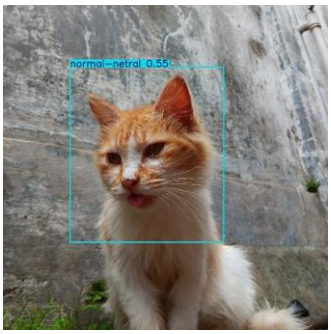


Figure 9. Model Detection Output



Figure 10. GradCAM Layer 0

4.9 Research Analysis

The model trained on 3,279 images across four expression classes successfully recognized three classes. The '*marah-agonistik*' class failed entirely due to class imbalance it contained only 159 images, compared to 223 (*normal-netral*), 187 (*senang-afiliatif*), and 178 (*stres-takut*), as shown in Figure 10. Augmentation widened this imbalance further.

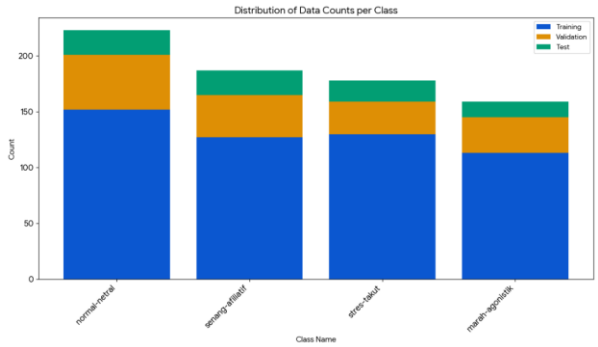


Figure 11. Class Distribution Comparison

Despite this limitation, the remaining three classes achieved mAP50 values above 50%: *senang-afiliatif* at 77.3%, *stres-takut* at 69.9%, and *normal-netral* at 65.5%. These results demonstrate that the model is viable for integration in real-world applications, provided additional data for the '*marah-agonistik*' class is collected to address the imbalance.

5 CONCLUSION

This study presents the development of an image-based cat facial expression recognition system using the YOLOv8n object detection architecture. The experimental results demonstrate that the proposed model is capable of recognizing several feline facial expression categories with satisfactory performance. Specifically, the model achieved reliable detection results for the *normal-netral*, *senang-afiliatif*, and *stres-takut* classes, with mAP50 values above 0.65 and an overall mAP50 of 0.728. These results indicate that the YOLOv8n architecture can effectively extract relevant facial features associated with subtle feline expressions and perform real-time detection when integrated into an application interface.

This research makes several contributions to the field of animal facial expression recognition. First, it demonstrates the feasibility of applying a lightweight object detection architecture (YOLOv8n) for recognizing subtle feline facial expressions from images. Second, the study constructs and organizes a dataset of cat facial images categorized into four expression classes (*normal-netral*, *senang-afiliatif*, *stres-takut*, and *marah-agonistik*), providing a structured dataset suitable for supervised learning in feline expression analysis. Third, the research implements a practical deployment scenario by integrating the trained model into a Gradio-based dashboard, enabling users to interact with the model and perform real-time expression recognition.

However, the evaluation results reveal that the model completely failed to detect the *marah-agonistik* class (F1-score = 0.00). This outcome indicates that the model was unable to learn discriminative features for this expression during training. Further analysis shows that this limitation is primarily caused by severe class imbalance in the dataset, where the number of *marah-agonistik* samples is significantly lower than the other expression classes. As a consequence, the model becomes biased toward the majority classes and fails to generalize to underrepresented categories.

Despite this limitation, the trained model was successfully integrated into a Gradio-based dashboard that allows users to upload cat images and obtain expression predictions in real time. This demonstrates the practical feasibility of implementing lightweight object detection models for assisting cat owners in monitoring their pets' emotional conditions through facial expression analysis.

Future research should focus on improving dataset balance and diversity, particularly by collecting additional samples representing the "*marah-agonistik*" expression. Expanding the dataset, improving annotation consistency, and exploring advanced techniques such as transfer learning, class re-weighting, or alternative deep learning architectures may further improve the robustness and generalization capability of the model for feline facial expression recognition.

REFFERENCE

[1]. J. Cheng, A. Malo, M. Garbin, B. P. Monteiro, and P. V. Steagall, "Construct validity, responsiveness and reliability of the Feline Grimace Scale in kittens," *Journal of Feline Medicine and Surgery*, vol. 25, no. 12, 2023. https://doi.org/10.1177/1098612X231211765

[2]. Vijayakumar and S. Vairavasundaram, "YOLO-based Object Detection Models: A Review and its Applications," *Multimedia Tools and Applications*, vol. 83, no. 35, pp. 83535–83574, 2024. https://doi.org/10.1007/s11042-024-18872-y

[3]. Peng, M. Sun, K. Zou, B. Zhang, G. Dai, and A. C. Tsoi, "Facial Expression Recognition-You Only Look Once-Neighborhood Coordinate Attention Mamba: Facial Expression Detection and Classification Based on Neighbor and Coordinates Attention Mechanism," *Sensors*, vol. 24, no. 21, 2024. https://doi.org/10.3390/s24216912

[4]. G. Martvel, N. Farhat, I. Shimshoni, and A. Zamansky, "CatFLW: Cat Facial Landmarks in the Wild Dataset," arXiv:2305.04232, 2023. http://arxiv.org/abs/2305.04232

[5]. G. Martvel, L. Scott, B. Florkiewicz, A. Zamansky, I. Shimshoni, and T. Lazebnik, "Computational investigation of the social function of domestic cat facial signals," *Scientific Reports*, vol. 14, no. 1, 2024. https://doi.org/10.1038/s41598-024-79216-2

[6]. G. Martvel, I. Shimshoni, and A. Zamansky, "Automated Detection of Cat Facial Landmarks," *International Journal of Computer Vision*, vol.

132, no. 8, pp. 3103–3118, 2024. https://doi.org/10.1007/s11263-024-02006-w
[7]. L. Scott and B. N. Florkiewicz, "Feline faces: Unraveling the social function of domestic cat facial signals," *Behavioural Processes*, vol. 213, p. 104959, 2023. https://doi.org/10.1016/J.BEPROC.2023.104959
[8]. M. E. Kret, J. J. M. Massen, and F. B. M. de Waal, "My Fear Is Not, and Never Will Be, Your Fear: On Emotions and Feelings in Animals," *Affective Science*, vol. 3, no. 1, pp. 182–189, 2022. https://doi.org/10.1007/s42761-021-00099-x
[9]. M. Feighelstein, I. Shimshoni, L. R. Finka, S. P. L. Luna, D. S. Mills, and A. Zamansky, "Automated recognition of pain in cats," *Scientific Reports*, vol. 12, no. 1, 2022. https://doi.org/10.1038/s41598-022-13348-1
[10]. M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," *Machines*, vol. 11, no. 7, 2023. https://doi.org/10.3390/machines11070677
[11]. M. L. Ali and Z. Zhang, "The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection," *Computers*, vol. 13, no. 12, 2024. https://doi.org/10.3390/computers13120336
[12]. M. M. Ambali Parambil, L. Ali, M. Swavaf, S. Bouktif, M. Gochoo, H. Aljassmi, and F. Alnajjar, "Navigating the YOLO Landscape: A Comparative Study of Object Detection Models for Emotion Recognition," *IEEE Access*, vol. 12, pp. 109427–109442, 2024. https://doi.org/10.1109/ACCESS.2024.3439346
[13]. M. M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," *Computation*, vol. 11, no. 3, 2023. https://doi.org/10.3390/computation11030052
[14]. M. Mendl, V. Neville, and E. S. Paul, "Bridging the Gap: Human Emotions and Animal Emotions," *Affective Science*, vol. 3, no. 4, pp. 703–712, 2022. https://doi.org/10.1007/s42761-022-00125-6
[15]. P. Kovačič, "Cat Facial Action Coding System (CatFACS) and Scientific Illustration," pp. 159–165, 2024. https://doi.org/10.55295/psl.2024.ii17
[16]. Q. Zhang, K. Ahmed, M. I. Khan, H. Wang, and Y. Qu, "YOLO-FCE: A feature and clustering enhanced object detection model for species classification," *Pattern Recognition*, vol. 171, 2025. https://doi.org/10.1016/j.patcog.2025.112218
[17]. R. Zhang and R. Ma, "Facial expression recognition method based on PSA—YOLO network," *Frontiers in Neurorobotics*, vol. 16, 2023. https://doi.org/10.3389/fnbot.2022.1057983
[18]. R. I. Mukhamediev, Y. Popova, Y. Kuchin, E. Zaitseva, A. Kalimoldayev, A. Symagulov, V. Levashenko, F. Abdoldina, V. Gopejenko, K. Yakunin, E. Muhamedijeva, and M. Yelis, "Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges," *Mathematics*, vol. 10, no. 15, 2022. https://doi.org/10.3390/math10152552
[19]. S. Broomé, M. Feighelstein, A. Zamansky, G. Carreira Lencioni, P. Haubro Andersen, F. Pessanha, M. Mahmoud, H. Kjellström, and A. A. Salah, "Going Deeper than Tracking: A Survey of Computer-Vision Based Recognition of Animal Pain and Emotions," *International Journal of Computer Vision*, vol. 131, no. 2, pp. 572–590, 2023. https://doi.org/10.1007/s11263-022-01716-3
[20]. S. Ren, M. Sun, B. Wang, M. Liu, and S. Men, "High Precision Infant Facial Expression Recognition by Improved YOLOv8," *IEEE Access*, vol. 13, pp. 39621–39630, 2025. https://doi.org/10.1109/ACCESS.2025.3543950
[21]. T. Rakhimzhanova, A. Kuzdeuov, and H. A. Varol, "AnyFace++: Deep Multi-Task, Multi-Domain Learning for Efficient Face AI," *Sensors*, vol. 24, no. 18, 2024. https://doi.org/10.3390/s24185993
[22]. X. Cao, Y. Su, X. Geng, and Y. Wang, "YOLO-SF: YOLO for Fire Segmentation Detection," *IEEE Access*, vol. 11, pp. 111079–111092, 2023. https://doi.org/10.1109/ACCESS.2023.3322143
[23]. X. Lang, Z. Ren, D. Wan, Y. Zhang, and S. Shu, "MR-YOLO: An Improved YOLOv5 Network for Detecting Magnetic Ring Surface Defects," *Sensors*, vol. 22, no. 24, 2022. https://doi.org/10.3390/s22249897
[24]. X. Ren, Y. Bai, G. Liu, and P. Zhang, "YOLO-Lite: An Efficient Lightweight Network for SAR Ship Detection," *Remote Sensing*, vol. 15, no. 15, 2023. https://doi.org/10.3390/rs15153771
[25]. Y. Bu, H. Ye, Z. Tie, Y. Chen, and D. Zhang, "OD-YOLO: Robust Small Object Detection Model in Remote Sensing Image with a Novel Multi-Scale Feature Fusion," *Sensors*, vol. 24, no. 11, 2024. https://doi.org/10.3390/s24113596
[26]. Y. Guo, S. Chen, R. Zhan, W. Wang, and J. Zhang, "LMSD-YOLO: A Lightweight YOLO Algorithm for Multi-Scale SAR Ship Detection," *Remote Sensing*, vol. 14, no. 19, 2022. https://doi.org/10.3390/rs14194801
[27]. Y. Qiu, F. Sha, and L. Niu, "DKA-YOLO: Enhanced Small Object Detection via Dilation Kernel Aggregation Convolution Modules," *IEEE Access*, vol. 12, pp. 187353–187366, 2024. https://doi.org/10.1109/ACCESS.2024.3515201
[28]. Y. S. Dosso, D. Kyrollos, K. J. Greenwood, J. Harrold, and J. R. Green, "NICUface: Robust

Neonatal Face Detection in Complex NICU Scenes," *IEEE Access*, vol. 10, pp. 62893–62909, 2022. https://doi.org/10.1109/ACCESS.2022.3181167

[29]. Y. Shao, Z. Xu, and Q. Zhu, "SH-YOLO: Enhanced Real-Time Detection of Laparoscopic Surgical Instruments in Computer-Aided Surgery Based on Star Operation and Hybrid Attention Mechanisms," *IEEE Access*, vol. 13, pp. 135179–135195, 2025. https://doi.org/10.1109/ACCESS.2025.3593635

[30]. Y. Song, Z. Xie, X. Wang, and Y. Zou, "MS-YOLO: Object Detection Based on YOLOv5 Optimized Fusion Millimeter-Wave Radar and Machine Vision," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 15435–15447, 2022. https://doi.org/10.1109/JSEN.2022.3167251

[31]. Y. Su, Q. Liu, W. Xie, and P. Hu, "YOLO-LOGO: A transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms," *Computer Methods and Programs in Biomedicine*, vol. 221, 2022. https://doi.org/10.1016/j.cmpb.2022.106903

[32]. Y. Wang, Y. Li, D. M. S. Kayes, H. S. Abdullahi, S. Gao, H. Zhang, Z. Song, and P. Lv, "Research on a Lightweight PCB Detection Algorithm Based on AE-YOLO," *IEEE Access*, vol. 12, pp. 109367–109379, 2024. https://doi.org/10.1109/ACCESS.2024.3439523

[33]. Y. Yang, Z. Liu, J. Chen, H. Gao, and T. Wang, "Railway Foreign Object Intrusion Detection Using UAV Images and YOLO-UAT," *IEEE Access*, vol. 13, pp. 18498–18509, 2025. https://doi.org/10.1109/ACCESS.2025.3533304