ISSN 2355-0082

IJNMT

International Journal of New Media Technology

VOLUME
12
No. 1

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2355-0082

EDITORIAL BOARD

EDITORIAL ADDRESS

Universitas Multimedia Nusantara (UMN)
Jl. Scientia Boulevard
Gading Serpong
Tangerang, Banten - 15811
Indonesia
Phone. (021) 5422 0808
Fax. (021) 5422 0800
Email : ultimacomputing@umn.ac.id

IJNMT (International Journal of New Media Technology) is a scholarly open access, peer-reviewed, and interdisciplinary journal focusing on theories, methods and implementations of new media technology. Topics include, but not limited to digital technology for creative industry, infrastructure technology, computing communication and networking, signal and image processing, intelligent system, control and embedded system, mobile and web based system, and robotics. IJNMT is published regularly twice a year (June and December) by Faculty of Engineering and Informatics, Universitas Multimedia Nusantara in cooperation with UMN Press.

Call for Papers

International Journal of New Media Technology (IJNMT) is a scholarly open access, peer-reviewed, and interdisciplinary journal focusing on theories, methods and implementations of new media technology. Topics include, but not limited to digital technology for creative industry, infrastructure technology, computing communication and networking, signal and image processing, intelligent system, control and embedded system, mobile and web based system, and robotics. IJNMT is published twice a year by Faculty of Engineering and Informatics of Universitas Multimedia Nusantara in cooperation with UMN Press.

Ultimatics : Jurnal Teknik Informatika is the Journal of the Informatics Study Program at Universitas Multimedia Nusantara which presents scientific research articles in the fields of Analysis and Design of Algorithm, Software Engineering, System and Network security, as well as the latest theoretical and practical issues, including Ubiquitous and Mobile Computing, Artificial Intelligence and Machine Learning, Algorithm Theory, World Wide Web, Cryptography, as well as other topics in the field of Informatics.

Ultima Computing : Jurnal Sistem Komputer is a Journal of Computer Engineering Study Program, Universitas Multimedia Nusantara which presents scientific research articles in the field of Computer Engineering and Electrical Engineering as well as current theoretical and practical issues, including Edge Computing, Internet-of-Things, Embedded Systems, Robotics, Control System, Network and Communication, System Integration, as well as other topics in the field of Computer Engineering and Electrical Engineering.

Ultima InfoSys : Jurnal Ilmu Sistem Informasi is a Journal of Information Systems Study Program at Universitas Multimedia Nusantara which presents scientific research articles in the field of Information Systems, as well as the latest theoretical and practical issues, including database systems, management information systems, system analysis and development, system project management information, programming, mobile information system, and other topics related to Information Systems.

ISSN 2355-0082

FOREWORD

Greetings!

IJNMT (International Journal of New Media Technology) is a scholarly open access, peer-reviewed, and interdisciplinary journal focusing on theories, methods and implementations of new media technology. Topics include, but not limited to digital technology for creative industry, infrastructure technology, computing communication and networking, signal and image processing, intelligent system, control and embedded system, mobile and web based system, and robotics. IJNMT is published regularly twice a year (June and December) by Faculty of Engineering and Informatics, Universitas Multimedia Nusantara in cooperation with UMN Press.

In this June 2024 edition, IJNMT enters the 1st Edition of Volume 12 No 1. In this edition there are ten scientific papers from researchers, academics and practitioners in the fields covered by IJNMT. Some of the topics raised in this journal are: Analysis of User Experience in Go Kreasi Application Using User Experience Questionnaire (UEQ), Implementation of Decision Support System for Analyzing the Suitability of Plantation Crops, Development and Implementation of a Corrosion Inhibitor Chatbot Using Bidirectional Long Short-Term Memory, Beyond Traditional Methods: The Power of Bi-LSTM in Transforming Customer Review Sentiment Analysis, Mapping User Dissatisfaction in Mobile Banking Applications Using Ensemble Clustering LDA and LSA, Web Design and Development of 'Peduli PMI' System for Managing Complaints and Protection of Indonesian Migrant Workers, Analysis of Student Performance Differences in Computer Network Courses with Learning Modes in Multimedia Nusantara University, E-Commerce Product Review Sentiment Analysis: A Comparative Study of Naïve Bayes Classifier and Random Forest Algorithms on Marketplace Platforms, Personalized Restaurant Recommendations: A Hybrid Filtering Approach for Mobile Applications, Monte Carlo Algorithm Applications in Shrimp Farming: Monitoring Systems and Feed Optimization.

On this occasion we would also like to invite the participation of our dear readers, researchers, academics, and practitioners, in the field of Engineering and Informatics, to submit quality scientific papers to: International Journal of New Media Technology (IJNMT), Ultimatics : Jurnal Teknik Informatics, Ultima Infosys: Journal of Information Systems and Ultima Computing: Journal of Computer Systems. Information regarding writing guidelines and templates, as well as other related information can be obtained through the email address ultimaijnmt@umn.ac.id and the web page of our Journal here.

Finally, we would like to thank all contributors to this June 2025 Edition of IJNMT. We hope that scientific articles from research in this journal can be useful and contribute to the development of research and science in Indonesia.

June 2025,

David Agustriawan, S.Kom., M.Sc., Ph.D.
Editor-in-Chief

ISSN 2355-0082

TABLE OF CONTENT

ISSN 2355-0082

Analysis of User Experience in Go Kreasi Application Using User Experience Questionnaire (UEQ)

Audi Martya Putri Winengku[1], Haya Utami[2], Suryaningsih Sinaga[3], Cornelius Mellino Sarungu[4]
[1]Information System, Bina Nusantara University, Malang, Indonesia
[2,3,4]Information System, Bina Nusantara University, Jakarta, Indonesia
[1]audi.winengku@binus.ac.id, [2]haya.utami@binus.ac.id, [3]suryaningsih.sinaga@binus.ac.id,
[4]cornelius.sarungu@binus.ac.id

Accepted 15 June 2023
Approved 9 June 2025

***Abstract*— This study aims to analyze the user experience of the Go Kreasi mobile application. The Go Kreasi application is a mobile application that provides tutoring services for students. The application was developed by PT. Ganesha Operation is still under development and used by 100 thousand users in Indonesia. The application has several features, including Visual, Audio, and Kinesthetic Ability Test (VAK), GO Assessment (GOA), Learning videos, Theory E-Books, Magic Books, Empathy, Racing, TOBK (Computer-Based Try Out), EPB, Leaderboard, and M3. [2] There are a number of complaints in the GO Kreasi app reviews, such as inappropriate color selection and confusing icon placement. This contributes to the app's low rating, which is 2.7 out of 5 on the Apple Store and 2.3 out of 5 on the Play Store. Many users stated that the app is less than satisfactory, which is one of the main reasons for this study. This study aims to analyze the user experience of the Go Kreasi mobile application using the User Experience Questionnaire (UEQ). UEQ measures user experience based on six dimensions: satisfaction, efficiency, stimulation, aesthetics, clarity, and identity. The research method is an online survey of 58 respondents who answered 26 questions about the Go Kreasi application. The researchers conducted a demographic analysis of the respondents. Most of the respondents were aged 17 years old and were students at the high school level. The results of the study showed that the overall UX of the application is good. The results showed that the dimensions of clarity and identity had the highest scores, while the dimensions of efficiency had the lowest. However, some aspects need to be improved to improve the user experience. These aspects include the ease of use, the accuracy of the information, and the dependability of the application. The researchers concluded that the Go Kreasi application has the potential to be a useful tool for students. From the research, we expected the Go Kreasi application can be better and more evolved in the future.**

***Index Terms*—user experience; user experience questionnaire; mobile learning application.**

I. INTRODUCTION

At present, technological advances are developing very rapidly so many things are converted to digital. Advances in technology also have an impact on various fields, including the field of education. [14] Tutoring is one of the mandatory things for students to support the knowledge they have. Apart from that, this is also useful for preparing for the selection of the dream college of every SMA/SMK student. Selection of tutoring is very important so that the expected goals are achieved. [20] Fluency in tutoring is the main thing that students look for in choosing a place for tutoring, including the application of tutoring as a learning support medium. Learning applications are considered practical for students in conducting learning outside of school hours.

One of the tutoring applications Go Kreasi has already been used by more than one hundred thousand users in Indonesia. This application made by Ganesha Operation allows students to study online and offline through media such as pictures, theoretical explanations, and interesting videos, do tryouts, and see the results of student progress. This application is made so that students can learn from anywhere and anytime. It is hoped that students will have more time to prepare themselves for taking the college selection exams they want. In addition, students can also evaluate their learning outcomes by looking at the progress of this application.

Many positive impacts have been given by the application of tutoring their students. The use of online tutoring applications can have a positive and significant impact on student achievement [3][15]. Students who are interested in a lesson must have the student's interest in taking the lesson. [19] The Go Kreasi application is considered to have the potential to be used on an ongoing basis in the future. This application is still under development, and currently, the app has received a lot of criticism and feedback from users, many of whom have said that the quality of the app is still unsatisfactory. So, we hope the research

is providing solutions and input to PT. Ganesha Operation so that the application becomes better and more useful in the future. Adaptive learning systems adjust the content and difficulty of the learning experience based on the individual learner's needs and progress. [4][14]

A. A. Theoretical Basis

a) Mobile Applications: often referred to as mobile applications or mobile applications, are specially designed and developed software packages that can run on mobile devices such as smartphones or tablets. Users can download and install mobile applications through the app store on certain mobile platforms, such as the Google Play Store for Android or the App Store for iOS.

Mobile applications can have a variety of functions, ranging from everyday uses such as social networking, communication, productivity, and entertainment, to applications specifically for business, education, health, banking, and others. Mobile applications can be used online or offline depending on the type and functionality of the application. The benefits of mobile learning are [8]:

- Flexibility
- Personalization
- Engagement
- Accessibility
- Cost-effectiveness
- Support for informal learning
- Improved communication and collaboration
- Enhanced motivation and engagement
- Increased learner satisfaction

b) User Experience: commonly abbreviated as UX is the feeling of users when using digital products, in another sense the result of the interaction of each user who visits or uses applications and websites. This feeling can be seen from the comfort of users in using digital products more easily and pleasantly. It is also an important aspect that focuses on the user experience, including in terms of response, emotion, and perception of the application or itself in the development of digital products. in today's modern era, such as websites, applications on smartphones, software on computers, and others.

UX is not just about making a product look good. It is also about making sure that the product is functional and meets the needs of the users. UX designers need to understand the users' needs and goals to create a product that is useful and delightful. [9]

c) The User Experience Questionnaire (UEQ): a method in the form of a questionnaire that is used to measure the user experience of a product. [1]

User Experience describes the user's subjective feelings towards the product they use. Every user has different feelings and impressions even when using the same product. [17]

The UEQ is used to evaluate different products and services, and the development of new versions of the UEQ that are more tailored to the needs of users, that the UEQ is a valid and reliable tool for evaluating user experience, and that it can be used to evaluate a wide range of products and services. [12]

Using the UEQ method can help determine and test whether the application/website has a good User Experience because the UEQ method uses several scenarios that help test the application. [18]

UEQ has 26 question items with six scales:

- Attractiveness: Measures the user's impression of interest in the product.
- Perspicuity: A measure of how easily users can understand and learn about the product.
- Efficiency: Measuring interactions between users and products quickly and efficiently.
- Dependability: Benchmarks of user feelings/interaction when using the product.
- Stimulation: Measures the user's enjoyment and motivation to use the product.
- Novelty: A measure of how creative and innovative the product is that can attract users' interest in using the product.

B. Literature Reviews from Similar Research

In the first study, Ahmad Husein Tarigan, Hergy Pradana, and Raihan Najmi Hersian used the User Experience Questionnaire (UEQ) method to analyze the UX of the MyTelkomsel application. The results show that the UX of the MyTelkomsel application has good value. However, several aspects need improvement, such as ease of use and accuracy of information. [5]

In the second study, Aditya Kusumadwiputra and Adela Zahwa Firdaus Suherman used the UEQ method to analyze the UX of the PeduliLindungi application. The results of the study show that the UX of the PeduliLindungi application still needs to be improved, especially in terms of the accuracy of the information. [6]

In the third study, Cici Damayanti Munthe and Sondang Sartika Siahaan used the UEQ method to analyze the UX of the Auto2000 application. The results of the study show that the UX of the Auto2000 application has good value. However, some aspects need improvement, such as ease of use and aesthetics. [7]

From those three studies, it can be concluded that. A good UX will make users feel satisfied and

comfortable in using the product so that it can increase the number of downloads and user satisfaction.

The difference between the third research and previous research is that the third research will examine mobile and web applications. This is done to get a more complete picture of the UX of the Auto2000 application. In addition, the third study will also propose a new user interface design based on the results of the evaluation carried out. Those studies are expected to contribute to the development of better digital products.

II. METHODOLOGY

A. Research step

The framework begins with the description and explanation of this Go Kreasi application. This research is based on User Experience with the User Experience Questionnaire (UEQ) Method. The results of this research later aim to help the Go Kreasi application have a UI/UX that suits its users which can later speed up running the application.

In this research, the first thing to do is to identify the problem. It is necessary to find out what problems occured in the application so the solutions or corrective actions can be proposed correctly. In solving existing problems, a solution method will be used by referring to library sources, literature, and journals. In determining the number of samples to be used, the random sampling technique and the solvency method.

A collection of samples will be collected in the form of a questionnaire with the UEQ method, and the results will be analyzed. The sample size should be as large as possible, the more samples that are collected, the more representative they will be, and the more generalizable the results will be.[13] At this stage, 20 to 30 respondents are needed, this follows the requirements for the minimum number of respondents in the UEQ method.[18]

The research step is shown in Fig. 1.

B. Stages of analysis

The stages in conducting data analysis in this study include:

1) Create a questionnaire according to the UEQ template: This questionnaire contains questions about demographic data. A questionnaire based on a literature study on UEQ was then re-created in the form of a Google Form and distributed to respondents via social media to fill out. An example of the questionnaire used in this study can be seen on the attachment page.

2) Collecting data: These data are based on answers to questionnaires by respondents. After that, a recapitulation of the answers given was carried out.

3) Convert the questionnaire: Convert into a weighted value for each answer by grouping each positive item and negative item. Here is an example of the conversion scale of UEQ in Table I.

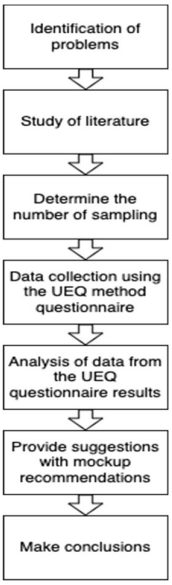


Fig. 1. Research step.

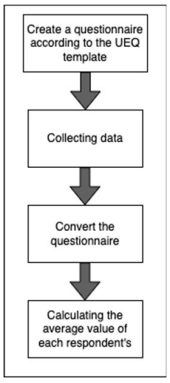


Fig. 2. Stage of analysis.

TABLE I. Example of Conversion Scale of UEQ.

Item	Conversion Scale							Item
	1	2	3	4	5	6	7	
inferior	-3	-2	-1	0	+1	+2	+3	valuable

This grouping is intended so that each item can be assessed on a scale range from -3 to +3. Values of -3, -2, and -1 indicate negative answers, values of 0 are neutral, while +1, +2, and +3 indicate positive answers. After the responses are obtained from the respondents, the answers are then converted into weighted values on the same scale, namely from -3 (strongly agree with the

negative statement) to +3 (strongly agree with the positive statement) [2].

4) Calculating the average value of each respondent's scale: Calculate with the data that has been converted into the weight of the answer value then grouped into 6 UEQ scales and items per scale. Each UEQ scale and its items are calculated on average by adding up all items on each UEQ scale and dividing by the number of each item.

III. Result and Discussion

A. Current Business Process

The current business process is to access this application, users need to register as students in online tutoring programs. The program provides several features including Visual, Audio, and Kinesthetic Ability Test (VAK), GO Assessment (GOA), Learning videos, Theory E-Books, Magic Books, Empathy, Racing, TOBK (Computer-Based Try-Out), EPB, Leaderboard, and M3. Users access the application by registering with the program separately outside of this application, then entering their mobile number, the registration number obtained when registering for the program, and receiving an OTP code via email, SMS, or WA based on the registered mobile number. After entering the OTP code, users can access the features of the Go Kreasi application. There are three menus in the Go Kreasi application, namely the Home menu, the 3B menu (Belajar, Berlatih, Bertanding), and the Social menu. The Home menu is the initial page when the user enters the application. Menu 3B is a page containing questions and theories that students can work on. The Social Menu is a menu for socializing with fellow GO students. This research focuses on the Home menu and 3B menu.

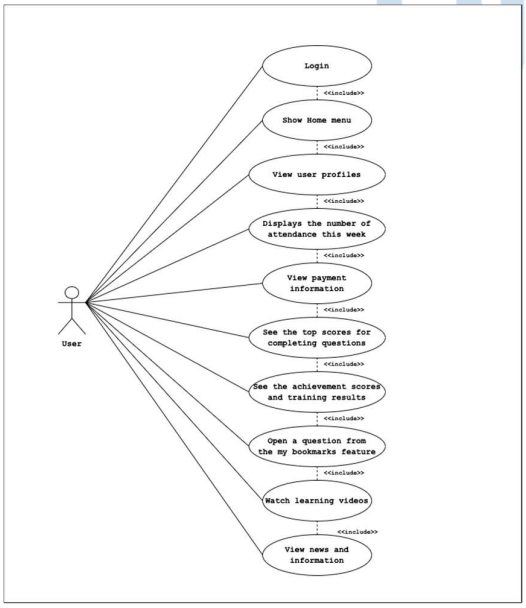


Fig. 3. Home menu use case diagram.

Home menu: The use case diagram describes the activities or activities that can be carried out by the user when accessing the GO Kreasi application on the home menu as the main menu of the Go Kreasi application. It is shown in Fig. 3.

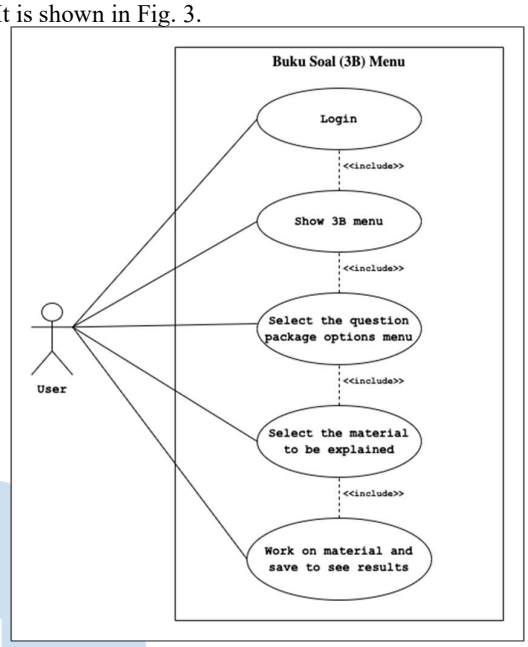


Fig. 4. Questions Book menu use case diagram (3B).

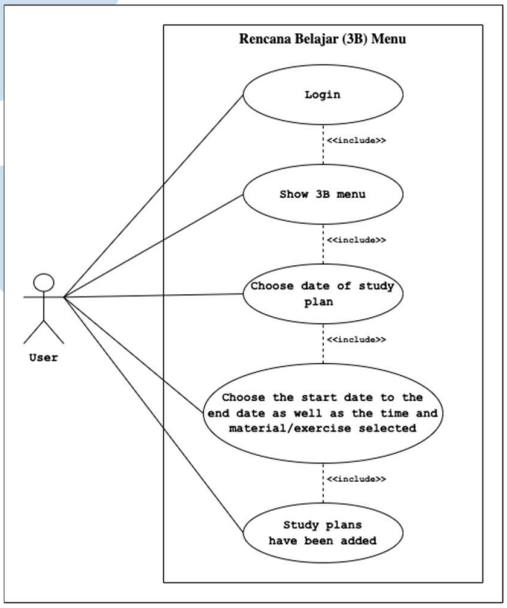


Fig. 5. Study Plan use case diagram (3B).

Buku Soal in 3B menu: The user opens menu 3B, the Buku Soal submenu, then selects the desired question package. Then select the material to be worked on. If you have chosen, then you can start working on the questions and immediately save them to see the results of the work evaluation. It is shown in Fig. 4.

Rencana Belajar in 3B menu: 3B menu: User opens menu 3B, the buku soal submenu, then selects the desired question package. Then select the material to be worked on. If you have chosen, then you can start working on the questions and immediately save them to see the results of the work evaluation. It is shown in Fig. 5.

B. Respondent Analysis in Demographics Result

Based on 58 respondents who participated in filling out the questionnaire on the Go Kreasi application. The age range of respondents who filled out the questionnaire was from 10 years old to 18 years old. Most respondents were respondents aged 17 years with a total of 17 respondents. Demographic results based on the age category of all respondents who have filled out the questionnaire for the Go Kreasi application are attached to Fig. 6.

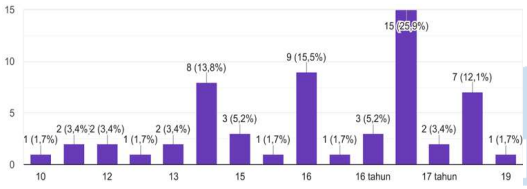


Fig. 6. Graphic of respondents based on age.

Respondents who used the application at the high school level were 43 people (74.1%), at the junior high school level there were 10 people (17.2%), and at the elementary level there were 5 people (8.6%). Demographic results based on the category of educational level when using the application from all respondents who have filled out the questionnaire for the Go Kreasi application are attached in Fig. 7.

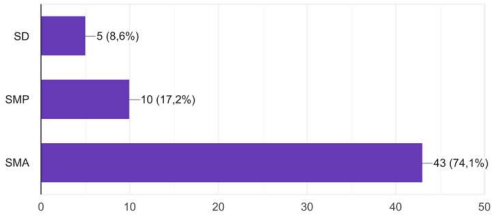


Fig. 7. Graphic of respondents based on education level.

C. UEQ Questionnaire Analysis Results of Home Menu

The questionnaire on the GO Kreasi application Home menu was compiled based on the UEQ template and distributed to respondents. The answers to the questionnaire were then converted into weighted values for each question item. Conversion begins by grouping each positive item and negative item. The results of the average value of the respondent's value scale are converted to the weight of the answer scores on the 6 UEQ scales and items per scale. The average value of all respondents on each questionnaire item is calculated by adding up all the respondents' values for each item divided by the number of respondents. The following are the results of these calculations in Table II.

TABLE II. Result of Calculation for Home menu based on 26 Items.

Item	Mean	Left	Right	Scale
1	1,59	annoying	enjoyable	Attractiveness
2	1,88	not understandable	understandable	Perspicuity
3	1,95	dull	creative	Novelty
4	1,88	difficult to learn	easy to learn	Perspicuity
5	2,28	inferior	valuable	Stimulation
6	1,59	boring	exciting	Stimulation
7	1,86	not interesting	interesting	Stimulation
8	1,60	unpredictable	predictable	Dependability
9	0,74	slow	fast	Efficiency
10	1,19	conventional	inventive	Novelty
11	1,76	obstructive	supportive	Dependability
12	2,10	bad	good	Attractiveness
13	1,60	complicated	easy	Perspicuity
14	1,47	unlikable	pleasing	Attractiveness
15	1,60	usual	leading edge	Novelty
16	1,79	unpleasant	pleasant	Attractiveness
17	2,10	not secure	secure	Dependability
18	1,71	demotivating	motivating	Stimulation
19	1,71	does not meet expectations	meets expectations	Dependability
20	1,71	inefficient	efficient	Efficiency
21	1,74	confusing	clear	Perspicuity
22	1,79	impractical	practical	Efficiency
23	1,76	cluttered	organized	Efficiency
24	1,53	unattractive	attractive	Attractiveness
25	1,93	unfriendly	friendly	Attractiveness
26	2,02	conservative	innovative	Novelty

The results of the average value of each scale from all respondents are added up with all the scale values of the respondents and divided by the number of respondents. The following results can be seen in Table III.

TABLE III. The Average of Each Scale from All Respondents on The Go Kreasi Home Application Menu.

UEQ's Scale	Average
Attractiveness	1,736
Perspicuity	1,776
Efficiency	1,500
Dependability	1,793
Stimulation	1,858
Novelty	1,690

Based on the results of the analysis of the UEQ questionnaire that has been distributed, it can be calculated the value of the average of all items on each scale from all respondents on the Home menu of the Go Kreasi application. The results of the overall average value are in Table IV and Fig. 8.

TABLE IV. Comparison Interval on The UEQ Home Menu.

Scale	Mean	Comparison interval for the UEQ scale
Attractiveness	1,74	Good
Perspicuity	1,78	Good
Efficiency	1,50	Above Average
Dependability	1,79	Excellent
Stimulation	1,86	Excellent
Novelty	1,69	Excellent

IV. CONCLUSION

Based on the results of the discussion in this study, it can be concluded that:

1. This study succeeded in conducting an analysis of the user experience evaluation [21] for the Home menu and the 3B Go Kreasi Application menu.
2. The results of data processing from 58 respondents using the scale comparison interval on the UEQ questionnaire for each item to measure the value of user experience, show that:
   a) Evaluation on the Go Kreasi Application Home menu received a *good* category for the attractiveness scale (1.74) and clarity (1.78), the *above average* category for the efficiency scale (1.50), and the excellent category for the accuracy scale (1.79), stimulation (1.86) and novelty (1.69).
   b) Evaluation on the 3B menu for the Go Kreasi application received a *good* category for an efficiency scale (1.71), an *above average* category for a clarity scale (1.38), and an excellent category for an attractiveness scale (1.93), accuracy (1.98), stimulation (2.00) and novelty (1.68).
   c) Evaluation of the Home and 3B menus of the Go Kreasi application has produced an average value on a scale with good and above average categories, so the application is said to be good.
3. Based on the results of the user experience assessment questionnaire on the Home menu of the Go Kreasi Application, the GO Kreasi application has a good user experience quality.

From the results of the questionnaire, suggestions were given in the form of a mockup of the GO Kreasi application.

REFERENCES

[1] M. Schrepp, (2023, September 12). "All you need to know to apply the UEQ " in User Experience Questionnaire Handbook. [online]. Available: https://www.ueq-online.org/Material/Handbook.pdf

[2] Ganesha Operation [online].Available: https://ganeshaoperation.com/

[3] A.Raraswati, in Pengaruh Pemanfaatan Aplikasi Bimbingan Belajar Online Terhadap Prestasi Belajar Akuntansi Peserta Didik. 2023. pp 10.

[4] Abubakari, Mussa and Nurkhamid, Nurkhamid and Hungilo, Gilbert. "Evaluating an e-Learning Platform at Graduate School Based on User Experience Evaluation Technique" in Journal of Physics: Conference Series. 2021.

[5] Tarigan, Pradana and Hersian, in Analisis Usability System Aplikasi Mytelkomsel Menggunakan Metode User Experience Questionnaire (UEQ). 2019.

[6] Kusumadwiputra and Suherman "Analisis Pengalaman Pengguna Pada Aplikasi PeduliLindungi di Jabodetabek dengan Menggunakan User Experience Questionnaire (UEQ)". 2023.

[7] Munthe & Siahaan, "Analisis User Experience Pada Aplikasi Auto2000 Menggunakan Metode User Experience Questionnaire (UEQ)". 2022.

[8] R.A Bates, "Mobile Learning: A Guide for Creating and Delivering Effective Content. John Wile & Sons". 2021

[9] Gothelf and Seiden, "User experience design: A practical guide to creating useful and delightful products". O'Reilly Media. 2022.

[10] Siti, Fasabuma and Tolle, Wijoyo, "Analisis pengalaman pengguna aplikasi pemesanan tiket bioskop menggunakan user experience questionnaire (UEQ) dan heuristic evaluation (HE)"[online]. Available: http://j-ptiik.ub.ac.id

[11] B. Vallendito, "Pemodelan User Interface dan User Experience Menggunakan Design Thinking". Universitas Islam Negeri Maulana Malilk Ibrahim Malang, 2020.

[12] Schrepp, M., Rauschenberger, A., & Schubert, T., "The User Experience Questionnaire (UEQ): A review of recent research". 2022.

[13] Nizar, A. (2019, Juli 17). "Menentukan jumlah sampel dalam penelitian" [online].Available: https://www.uinsyahada.ac.id/bagaimana-menentukan-jumlah-sampel-dalam-penelitian/3/

[14] Bay Atlantic University. (2022, Juni 3). "How does Technology impact student learning" [online]. Available: https://bau.edu/ blog/technology-impact-on-learning/#:~:text=A%20very%20important%20technological%20impact,has%20had%20on%20student%20learning.

[15] Alimah, Azkia D. "Pengaruh Pembelajaran Mobile Menggunakan Aplikasi "Sistem Kehidupan Vertebrata (3)" terhadap Kemampuan Kognitif Peserta Didik pada Materi Sistem Koordinasi." Bioedusiana, vol. 3, no. 1, 2018, pp. 15-21, doi:10.34289/277902.

[16] Norman, Don, The design of everyday things: Revised and expanded edition. Basic books, 2013.

[17] H. B. Santoso, M. Schrepp, R. Y. K. Isal, A. Y. Utomo, and B. Priyogi, "Measuring User Experience of the Student-Centered e-Learning Environment," J. Educ. Online, vol. 13, no. 1, pp. 58–79, 2016.

[18] N. G. Ameniar, H. Prastawa, and Z. F. Rosyada, "Evaluasi User Experience Menggunakan Metode User Experience Questionnaire (UEQ) dan Penerapan Kansei Engineering Pada Aplikasi Cinepolis Cinemas Indonesia," Industrial Engineering Online Journal, vol. 11, no. 4, Sep. 2022.

[19] F. Jabnabillah, W. Reza. "Pengaruh Penggunaan Aplikasi Geogebra Terhadap Minat Belajar Siswa Pada Pembelajaran Matematika", 2022.

[20] Dotson, R. (2016) "Goals Setting to Increase Academic Performance" [online]. Available: https://files.eric.ed.gov/fulltext/EJ1158116.pdf

[21] Kusumo, R. H. P., and Beni Suranto "Evaluasi User Experience Sistem Informasi Manajemen Tugas Akhir (SEKAWAN) Informatika Universitas Islam Indonesia Menggunakan Metode User Experience Questionnaire (UEQ)",. 2023.

[22] IEEE Author Center. IEEE Reference Guide [Online].Available:https://ieeeauthorcenter.ieee.org/wp-content/uploads/ IEEE-Reference-Guide.pdf.

[23] Zhang, H. (2020, Jan 29)"7 Principles of Icon Design"[onlinre] Available: https://uxdesign.cc/7-principles-of-icon-design-e7187539e4a2

ISSN 2355-0082

Implementation of Decision Support System for Analyzing the Suitability of Plantation Crops

Zilvanhisna Emka Fitri[1], Ahmad Dandi Irawan[2], Abdul Madjid[3] , Arizal Mujibtamala Nanda Imron[4]
[1,2] Department of Information Technology, Politeknik Negeri Jember, Jember, Indonesia
[3] Department of Agriculture Production, Politeknik Negeri Jember, Jember, Indonesia
[4] Department of Electrical Engineering, Universitas Jember, Jember, Indonesia
zilvanhisnaef@polije.ac.id, abdul_madjid@polije.ac.id, arizal.tamala@unej.ac.id

Accepted 19 April 2024
Approved 15 November 2024

***Abstract*— The productivity of plantation crops is a priori dependent on the suitability of the land and the quality of the land used. The objective of determining land suitability is to increase the amount of crop production, thereby preventing crop failure. The process of land evaluation entails the assessment of land performance with the objective of predicting the potential and limiting factors for crop production. This allows for the identification of alternative types of agriculture. The application of the Fuzzy Mamdani method to the land suitability assessment website, based on rainfall parameters, soil pH and planting depth, can provide a land class assessment while making recommendations for plantation crops as an alternative type of agriculture. The system accurately recommends with 100% success rate for very suitable classes (S1) and suitable (S2) categories.**

***Index Terms*— agriculture precision; decision support system; fuzzy mamdani; suitability of plantation crop**.

I. INTRODUCTION

Plantation commodities represent a significant component of the Indonesian economy, contributing to the country's foreign exchange earnings during the global pandemic. A review of the export value of plantation commodities in 2021 reveals that the total export value of plantations reached US$ 40.71 billion, equivalent to IDR 583.21 trillion (assuming 1 US$ = IDR 14,327) [1]. The productivity of plantation crops is a priori dependent on the suitability of the land and the quality of the land used. The objective of determining land suitability is to increase the amount of crop production, thereby preventing crop failure.

The term "land" encompasses the physical environment, including climate, topography/relief, soils, hydrology, and the state of natural vegetation. These elements collectively influence land use [2]. The process of land evaluation entails the assessment of land performance with the objective of predicting the potential and limiting factors for crop production. This allows for the identification of alternative types of agriculture [3]. The suitability of plants for cultivation is determined by their ability to thrive in specific types of land. The FAO has developed a classification system for land suitability that distinguishes between four categories: S1 (Very Suitable), S2 (Moderately Suitable), S3 (Marginal Suitable), and N (Unsuitable). [4].

Some of the plantation crops cultivated in Indonesia are palm oil, rubber, coconut, coffee, cocoa, cloves [5], pepper, tea, nutmeg, sugarcane, cinnamon, tobacco, etc. [1]. The Central Bureau of Statistics (BPS) of East Java Province reports that the most prevalent plantation crops cultivated in the region are coconut, rubber, coffee, cocoa, sugarcane, tea, and tobacco. The research on evaluation of land suitability for plantation crops that is referenced in this research study employs a variety of methods. These include an evaluation of land suitability for robusta coffee plants based on both physical environmental factors and economic factors, followed by an analysis of feasibility using the R/C ratio method. [6], A geographic information system (GIS) was employed to evaluate the suitability of the land for the cultivation of sugarcane [4], cocoa [7], and tobacco[2].

Decision support system method such as spatial decision tree algorithm for optimising oil palm land suitability with 23 rules [8]. In addition to the decision tree method, other techniques, such as fuzzy logic, are employed for the assessment of land suitability for the cultivation of sugarcane [9], rubber and oil palm plant suitability [10], the selection of horticultural crops [11], the determination of rubber plant quality [12] and the identification of superior cinnamon seedlings [13].

One of the most employed fuzzy methods is the Mamdani fuzzy method, which utilises the largest-smallest value to derive the output value. The stages of the Mamdani Fuzzy Inference System Logic Model, from the formation of the Fuzzy set to the affirmation process (deFuzzyfication), demonstrate a correlation between the input variables [14]. This correlation allows the model to determine the suitability of plantation crops.

II. THEORY

The research stages comprise a literature study, problem identification, data collection, application design and development, application testing and result analysis, as illustrated in Figure 1.

Fig. 1. Flow of research stages

A. Literature review and problem identification

A literature study is a preliminary stage of research, during which materials relevant to the topic under investigation are collected. These materials may take the form of previous theories or hypotheses related to the research in question. Additionally, the problem identification process may involve interviews and direct observations with farmers and experts who are knowledgeable about the subject matter.

B. Data collection

This research data came from the soil laboratory of Jember State Polytechnic with an expert, Ir. Abdul Madjid, MP as the head of the soil laboratory. The data on land suitability for plantation crops is presented in Table 1.

TABLE I. LAND SUITABILITY DATA

Factors	S1 (Very Suitable)	S2 (Moderately Suitable)	S3 (Marginal Suitable)	N (Unsuitable)
Drainage	Good	Rather Fast - Good	Rather Fast - Good	Somewhat inhibited - inhibited
Inundation	No inundation	At least 2 months inundation	Less than 4 months inundation	No permanent inundation
Salinity (mmhos/cm)	<1500	1500 - 2500	<4000	>4000
pH of soil	6 - 7	5.5 - 6 7 - 8	4.5 – 5.5 8 – 8.5	3.5 – 4.5 >8.5
Soil fertility	High	High - Medium	High - Low	Low
Number of stones	<5%	<25%	<50%	<75%
Texture	Clay	Loam, silty, silty clay	Clay, silty sand, clay	Clay, silty sand
Depth (cm)	>100	100 - 75	75 - 50	50 - 25

Source :Soil Laboratory Politeknik Negeri Jember

Table 1 presents a land suitability classification with four classes: S1 (Very Suitable), S2 (Suitable), S3 (Marginal Suitable) and N (Unsuitable). These classes are derived from 8 parameters, including drainage, inundation, salinity, soil pH, soil fertility, stone percentage, soil texture and planting depth. The land

TABLE II. LAND SUITABILITY BASED ON PLANTATION CROP CHARACTERISTICS

Plant type	Parameters	Land Suitability Classes			
		S1	S2	S3	N
Robusta coffee	Rainfall Rate	2000 - 3000	1750 - 2000 3000 - 3500	1500 - 1750 3500 - 4000	<1500 >4000
	pH of soil	5.3 – 6	6 – 6.5 5 – 5.3	>6.5 <5	-
	Depth (cm)	>100	75 - 100	50 - 75	<50
Palm oil	Rainfall Rate	1700 - 2500	1450 - 1700 2500 - 3500	1250 - 1450 3500 - 4000	<1250 >4000
	pH of soil	5 – 6.5	4.2 – 5 6.5 – 7	>7	-
	Depth (cm)	>100	75 - 100	35 - 55 50 - 75	>55 <50
Rubber	Rainfall Rate	2500 - 3000	2000 - 2500 3000 - 3500	1500 - 2000 3500 - 4000	<1500 >4000
	pH of soil	5 – 6	6 – 6.5 4.5 – 5	>6.5 <4.5	-
	Depth (cm)	>100	75 - 100	50 -75	<50
Cacao	Rainfall Rate	1500 – 2500	>2500 - 3000	>3000 - 4000 1250 - < 1500	>4000 <1250
	pH of soil	5.5 – 6.5	>6.5 – 7.5 5 - < 5.5	>7.5 – 8.5 4.5 - < 5	< 4
	Depth (cm)	>100	75 - 100	50 - <75	<50
Cloves	Rainfall Rate	1500 - 2500	2500 - 3000	1250 - 1500 3000 - 4000	<1250 >4000
	pH of soil	5 – 7	4 – 5 7 – 8	< 4 > 8	-
	Depth (cm)	>100	75 - 100	50 - 75	<50
Nutmeg	Rainfall Rate	2000 - 4500	1800 - 2000 4500 - 4800		<1800 >4800
	pH of soil	5 – 7	4 – 5 7 – 8	< 4 > 8	-
	Depth (cm)	>100	75 - 100	50 - 75	<50
Cinnamon	Rainfall Rate	2000 - 2500	1300 - 2000 2500 - 3000	1000 - 1300 3000 - 4000	<1000 >4000
	pH of soil	5 – 7	4 – 5 7 – 8	< 4 > 8	-
	Depth (cm)	>100	75 - 100	50 - 75	<50
Arabica coffee	Rainfall Rate	1200 - 1800	1000 - 1200 1800 - 2000	2000 - 3000 800 - 1000	<800 >600
	pH of soil	5.6 – 6.6	6.6 – 7.3	< 5.5 > 7.4	-
	Depth (cm)	>100	100 - 150	50 - 100	<50
Tea	Rainfall Rate	2500 - 4000	1800 - 2500 4000 - 5000	1300 - 1800 5000 - 6000	<1300 >6000
	pH of soil	4.5 - 5	4 – 4.4 5.1 – 5.5	3.5 - 3.9 5.6 – 6.5	<3.5 >6.6
	Depth (cm)	>150	100 - 149	40 - 99	<40

Source :Soil Laboratory Politeknik Negeri Jember

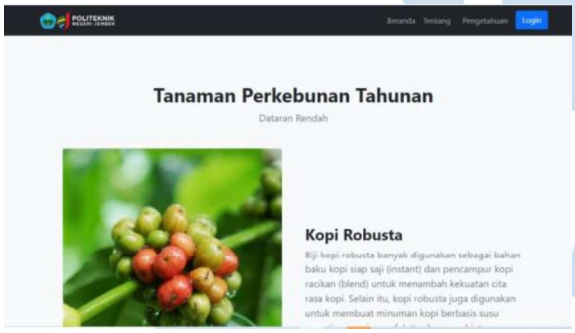
suitability data for plantation crops based on three parameters such as rainfall, soil pH and depth of planting are presented in Table 2. The plantation crops we use are robusta coffee, arabica coffee, oil palm, rubber, cocoa, cloves, nutmeg, cinnamon and tea.

C. App design and build

The application design consists of several menus such as dashboard home, about the system, plant knowledge page, rule page and calculation page as shown in Figure 2. The development of the application commences with the construction of the fuzzy Mamdani method flowchart (Figure 3), which comprises the following sequence of operations:

a. The first step in this process is to input the requisite data pertaining to the crops, specifically the precipitation levels, the pH of the soil and the depth of the planting.

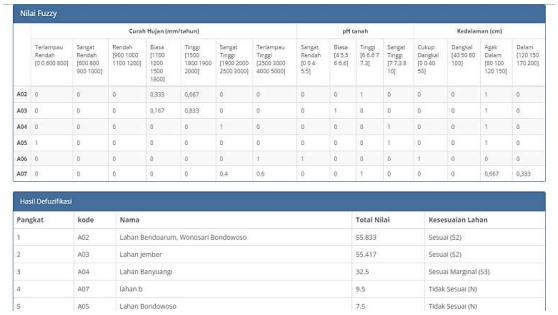
(a)



(b)

(c)

(d)



(e)

Fig. 2. illustrates an overview of the application design.

Figure 2 illustrates the application design, which comprises the following components: (a) home menu, (b) knowledge menu showing information about plantation crops used as data, (c) criteria menu containing parameters (rainfall, soil pH and soil depth) and land suitability classes, (d) rule menu used to create rules for land suitability classes and (e) calculation menu containing calculation tables from the fuzzy Mamdani method on selected rule options from the system.

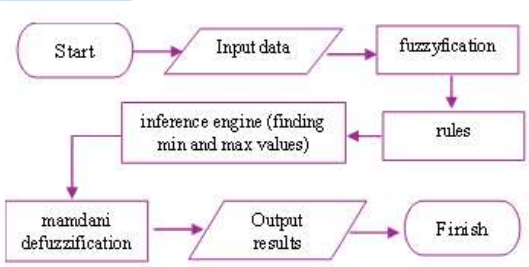


Fig. 3. Application design view

Figure 3 shows the stages of the mamdani fuzzy method which consists of input data, fuzzification, rule determination, inference engine, mamdani defuzzification and output results.

b. Fuzzification is the process of creating fuzzy sets from input variables and output variables.

a. Rainfall rate consists of 3 fuzzy sets with low, normal and high outputs.

$$\mu\ low\ (x) = \begin{cases} 0, & x \geq 1750 \\ \dfrac{1750 - x}{1750 - 1500}, & 1500 \leq x \leq 1750 \\ 1, & x \leq 10 \end{cases} \quad (1)$$

$$\mu\ normal\ (x) = \begin{cases} 0, & x \leq 1500\ atau\ x \geq 3000 \\ \dfrac{x - 1500}{1750 - 1500}, & 1500 \leq x \leq 1760 \\ 1, & 1750 \leq x \leq 2000 \\ \dfrac{3000 - x}{3000 - 2000}, & 2000 \leq x \leq 3000 \end{cases} \quad (2)$$

$$\mu\ high\ (x) = \begin{cases} 0, & x \leq 2000 \\ \dfrac{x - 2000}{3000 - 2000}, & 2000 \leq x \leq 3000 \\ 1, & x \geq 3000 \end{cases} \quad (3)$$

b. Soil pH consists of 3 fuzzy sets with low, normal and high outputs.

$$\mu\ low\ (x) = \begin{cases} 0, & x \geq 5.3 \\ \dfrac{6-x}{6-5.3}, & 5.3 \leq x \leq 6 \\ 1, & x \leq 5.3 \end{cases} \quad (4)$$

$$\mu\ normal\ (x) = \begin{cases} 0, & x \leq 5.3\ atau\ x \geq 7 \\ \dfrac{x-5.3}{6-5.3}, & 5.3 \leq x \leq 6 \\ 1, & 6 \leq x \leq 6.5 \\ \dfrac{7-x}{7-6.5}, & 6.5 \leq x \leq 7 \end{cases} \quad (5)$$

$$\mu\ high\ (x) = \begin{cases} 0, & x \leq 6.5 \\ \dfrac{x-6.5}{7-6.5}, & 6.5 \leq x \leq 7 \\ 1, & x \geq 7 \end{cases} \quad (6)$$

c. Planting depth consists of 3 fuzzy sets with outputs of shallow, slightly deep and deep.

$$\mu\ shallow\ (x) = \begin{cases} 0, & x \geq 50 \\ \dfrac{75-x}{75-50}, & 650 \leq x \leq 1000 \\ 1, & x \leq 50 \end{cases} \quad (7)$$

$$\mu\ slightly\ deep\ (x) = \begin{cases} 0, & x \leq 50\ atau\ x \geq 200 \\ \dfrac{x-50}{75-50}, & 50 \leq x \leq 75 \\ 1, & 75 \leq x \leq 100 \\ \dfrac{200-x}{200-100}, & 100 \leq x \leq 200 \end{cases} \quad (8)$$

$$\mu\ deep\ (x) = \begin{cases} 0, & x \geq 100 \\ \dfrac{x-100}{200-100}, & 650 \leq x \leq 1000 \\ 1, & x \leq 200 \end{cases} \quad (9)$$

The inference system generated 140 rules with three parameters (rainfall, soil pH and planting depth) and four output classes (S1, S2, S3 and N), as illustrated in Table 3.

TABLE III. EXAMPLE OF LAND AND SUITABILITY RULE

No.	Rules
1.	if Rainfall rate = low and Soil pH = very low and Planting depth = shallow enough then Land Suitability = Unsuitable (N)
4.	if Rainfall rate = low and Soil pH = very low and Planting depth = deep then Land Suitability = Marginal Suitable (S3)
5.	if Rainfall rate = low and Soil pH = low and Planting depth = shallow enough then Land Suitability = Unsuitable (N)
6.	if Rainfall rate = low and Soil pH = low and Planting depth = shallow then Land Suitability = Marginal Suitable (S3)
7.	if Rainfall rate = low and Soil pH = low and Planting depth = slightly deep then Land Suitability = Suitable (S2)
8.	if Rainfall rate = low and Soil pH = low and Planting depth = deep then Land Suitability = Very Suitable (S1)
9.	if Rainfall rate = low and Soil pH = Normal and Planting depth = shallow enough then Land Suitability = Unsuitable (N)
10.	if Rainfall rate = low and Soil pH = Normal and Planting depth = shallow then Land Suitability = Marginal Suitable (S3)
11.	if Rainfall rate = low and Soil pH = Normal and Planting depth = slightly deep then Land Suitability = Suitable (S2)
12.	if Rainfall rate = low and Soil pH = Normal and Planting depth = deep then Land Suitability = Very Suitable (S1)

c. Rule composition is a method of performing fuzzy system inference called the max method [15] with the equation:

$$\mu sf = \max\left(\mu sf[X_i], \mu kf[X_i]\right) \quad (10)$$

$\mu sf[X_i]$: The value of membership for a fuzzy solution up to rule i

$\mu kf[X_i]$: The value of membership in the fuzzy consequence of rule i.

d. The defuzzification process involves the input of a fuzzy set, which is obtained from the composition of fuzzy rules. The resulting output is a number within the domain of the fuzzy set. Defuzzification method on the composition of mamdani rules with centroid method (composite moment) with equation:

$$Z^* = \frac{\int z\mu(z)dz}{\int \mu(z)} \quad (11)$$

e. The calculation of land suitability and the analysis of the results related to the recommendations for plantation crops.

III. RESULT AND DISCUSSION

The calculation of land suitability can be based on a predetermined case study with three parameters (rainfall, soil pH and planting depth) as inputs, resulting in four outputs (Highly Suitable (S1), Suitable (S2), Marginal Suitable (S3) and Unsuitable (N)).

One example is a farmer who has land with :

1. Rainfall rate : 1700 mm/tahun
2. Soil pH : 6.5
3. Planting depth: 110 cm

a. A calculation of rainfall categories was conducted, utilising equations (1) and (2) to ascertain the rainfall value of 1700mm/year. The calculations are presented below.

$\mu\ low\ (x) = 0.2;\ \mu\ normal\ (x) = 0.8$

b. A calculation of the soil pH category was conducted, with the soil pH value determined to be 6.5, in accordance with the formulae (5) and (6). The calculation was as follows:

$\mu\ normal\ (x) = 0.6\ ;\ \mu\ high\ (x) = 0.4$

c. The depth category for plant planting is 110 cm, calculated using Equations 8 and 9 :

$\mu\ slightly\ deep\ (x) = 0.9\ ;\ \mu\ deep\ (x) = 0.1$

Subsequently, the most appropriate rule for determining the land suitability class should be identified by applying the maximum method with the formula equation (10) to the categories previously outlined :

a. Land value not suitable (N) = min (0) = 0

b. Land value in marginal suitability (S3) = min (0)
= 0

c. Land value suitable (S2) = min (0; 0.1; 0.2; 0.4; 0.6) = 0.6

d. Land value very suitable (S1) = min (0.1) = 0.1

Rule composition is an overall conclusion by taking the maximum membership level of each consequent of the implication function application and combining all the conclusions of each rule, so that the Fuzzy solution area is obtained as follows :

$$\mu sf = \max(\mu sf[X_i]) = \max(0.6)$$

The rule cut-off point is when $\mu\ agak\ dalam\ (x) = 0.6$ then the value of x can be determined as follows:

a. $\frac{200-x}{200-100} = 0.6$

$\frac{200-x}{100} = 0.6$

$x - 200 = -(0.6 * 100)$

$x - 200 = -60$

$x = 200 - 60$

$x = 140$

b. $\frac{200-x}{200-100} = 0.1$

$\frac{200-x}{100} = 0.1$

$x - 200 = -(0.1 * 100)$

$x - 200 = -10$

$x = 200 - 10$

$x = 180$

So we get the membership function of the solution area

$$\mu\ slightly\ deep\ (x) = \begin{cases} 0.1; x \leq 140 \\ \frac{200-x}{200-100}; 140 \leq x \leq 180 \\ 0.6; x \geq 180 \end{cases}$$

Defuzzification or affirmation is to convert Fuzzy sets into real numbers. The input of the affirmation process is a Fuzzy set, while the resulting output is a number in the Fuzzy set domain. The defuzzification method uses the centroid method using the formula equation (11):

$$Z = \frac{\int_{140}^{180}(0.6)xdx}{\int_{140}^{180}0.6dx} = \frac{0.3x^2]_{140}^{180}}{0.6x]_{140}^{180}} = \frac{3840}{24} = 160$$

Then the land suitability value is 160 with the appropriate class (S2) while for analysis recommending the yield of plantation crops obtained as in Table 4.

Plant type	Parameter		
	Rainfall rate (mm/year)	Soil pH	Planting depth (cm)
Palm oil	1450 - 1700	4.2 – 7	75 - 100
Cinnamon	1300 - 2000	4 – 8	75 - 100
Arabica coffee	1000 - 2000	6.6 – 7.3	100 - 150

If farmers have land with rainfall of 1700 mm/year, soil pH of 6.5 and planting depth of 110 then the recommended plantation crops are arabica coffee, cinnamon and oil palm with suitable land suitability category (S2).

The system was tested on 10 land sample datasets featuring rainfall, soil pH and planting depth parameters. The results were validated using the system output, which successfully recommended plants in the S1 (Very Suitable) and S2 (Suitable) classes, as detailed in Table 4.

Table 4 presents the results of the analysis, indicating that 10 land sample data sets correlate with the recommended plants in both S1 and S2 classes, with a 'valid' mark. This indicates that the system is effective in recommending plants according to the three parameters, with an accuracy rate of 100%.

TABLE IV. SYSTEM TESTING RESULTS

No.	Land and Parameters	Output Sistem		Validate results
		S1	S2	
1	Antirogo Rainfall : 1780 soil ph : 5.5 planting depth : 100	Cloves Oil palm Cocoa	Arabica coffee Cinnamon Nutmeg Robusta coffee	Valid
2	Rembangan Rainfall : 2100 soil ph : 5.7 planting depth : 105	Cloves Oil palm Cocoa Cinnamon Robusta coffee	Rubber	Valid
3	Dawuhan 1 Rainfall : 1300 soil ph : 5.7 planting depth : 100		Arabica coffee Cinnamon	Valid
4	Dawuhan 2 Rainfall : 1200 soil ph : 5.7 planting depth : 100		Arabica coffee	Valid
5	Senduro Rainfall : 1700 soil ph : 5 planting depth : 100	Cloves Oil palm	Arabica coffee Cocoa Cinnamon Robusta coffee	Valid
6	Tempurejo Rainfall : 1500 soil ph : 5.7 planting depth : 100	Cloves Cocoa	Arabica coffee Oil palm Cinnamon	Valid
7	Kebun Renteng PTPN XII Rainfall : 2700 soil ph : 5 planting depth : 100	Rubber Nutmeg	Cloves Oil palm Cocoa Cinnamon Tea Robusta Coffee	Valid
8	Gambir tea Rainfall : 4000 soil ph : 5.2 planting depth : 105	Nutmeg	Tea	Valid
9	Gambir tea Rainfall : 3500 soil ph : 4.6 planting depth : 105		Tea Nutmeg	Valid
10	Ijen Rainfall : 2200 soil ph : 5.3 planting depth : 100	Cloves Oil palm Cinnamon Nutmeg Robusta coffee	Rubber Cocoa Tea	Valid

Nevertheless, some plants remain red in colour, particularly within the S2 category. For example, the third data recommendation result indicates that the system has categorised cinnamon as land suitability class S2 (suitable). Cinnamon has a rainfall value of 2000–2500 mm/year, a soil pH of 5–7, and a planting depth of >100 cm. The system provides land suitability based on rainfall parameters, soil pH and planting depth in accordance with Rule 71. Where rainfall is normal, soil pH is normal and planting depth is rather deep, the system will recommend Class S2. Furthermore, as evidenced in Table 2, the data indicates that cinnamon

exhibits rainfall with two value ranges: 1300–2000 or 2500–3000, and planting depth with a value range of 75–100. If these two parameters are fulfilled, the system will suggest a recommendation within the S2 category.

Another example is the results of crop recommendations on the 9th data set, which indicate that tea should be a very suitable crop recommendation (S1). Tea is planted at an annual rainfall of 3500 mm, a pH of 4.6 and a planting depth of 105 cm, as indicated in Table 2. However, the system instead recommends tea plants into a suitable class (S2) according to fuzzy Mamdani calculations for land in the Gambier tea garden.

IV. CONCLUSION

The study revealed that rainfall and soil pH are adequate for determining crop recommendations based on land suitability, because the average depth of plantation crops is more than 100 cm. The system accurately recommends with 100% success rate for very suitable classes (S1) and suitable (S2) categories. However, further enhancements are necessary, particularly in incorporating rules or comparing with alternative calculation methods to develop a decision system for plantation crop land suitability.

REFERENCES

[1] Directorate General of Plantations, "Statistik Perkebunan Unggulan Nasional 2021-2023." Ministry of Agriculture Republic of Indonesia, 2024.

[2] R. D. Dewantara and D. Azis, "EVALUASI KESESUAIAN LAHAN PERKEBUNAN TEMBAKAU DI KABUPATEN ACEH TENGAH MENGGUNAKAN ANALISIS SISTEM INFORMASI GEOGRAFIS," *J. Pendidik. Geosfer*, vol. 6, no. 1, Jun. 2021, doi: 10.24815/jpg.v6i1.22099.

[3] T. M. Hartati, B. H. Sunarminto, and M. Nurudin, "Evaluasi Kesesuaian Lahan untuk Tanaman Perkebunan di Wilayah Galela, Kabupaten Halmahera Utara, Propinsi Maluku Utara," *Caraka Tani J. Sustain. Agric.*, vol. 33, no. 1, p. 68, Apr. 2018, doi: 10.20961/carakatani.v33i1.19298.

[4] T. Novita and A. W. Abdi, "EVALUASI KESESUAIAN LAHAN PERKEBUNAN TEBU DI KABUPATEN ACEH TENGAH DENGAN MENGGUNAKAN SISTEM INFORMASI GEOGRAFI," 2019.

[5] D. S. Simbolon and B. Sinaga, "Sistem Pendukung Keputusan Penentuan Kesesuaian Lahan Tanaman Cengkeh Dengan Metode Profile Matching," *J. Nas. Komputasi Dan Teknol. Inf. JNKTI*, vol. 4, no. 5, pp. 370–379, Oct. 2021, doi: 10.32672/jnkti.v4i5.3427.

[6] S. T. Dermawan, T. B. Kusmiyarti, and J. P. B. Sudirman, "Evaluasi Kesesuaian Lahan untuk Tanaman Kopi Robusta (Coffea canephora) di Desa Pajahan Kecamatan Pupuan Kabupaten Tabanan," vol. 7, no. 2, 2018.

[7] I. B. Suryawan, "Evaluasi Kesesuaian Lahan Untuk Beberapa Tanaman Pangan Dan Perkebunan Di Kecamatan Burau Kabupaten Luwu Timur Sulawesi Selatan," vol. 9, no. 1, 2020.

[8] A. Nurkholis and I. S. Sitanggang, "Optimization for prediction model of palm oil land suitability using spatial decision tree algorithm," *J. Teknol. Dan Sist. Komput.*, vol. 8, no. 3, pp. 192–200, Jul. 2020, doi: 10.14710/jtsiskom.2020.13657.

[9] D. A. Puryono, "Pemilihan Varietas Tebu Sesuai Lahan Menggunakan Metode Fuzzy Inferensi System Mamdani," *STIMIKA*, vol. 2, no. 1, pp. 83–90, 2015, doi: 10.31219/osf.io/4bdmf.

[10] M. Yusida, D. Kartini, R. A. Nugroho, and M. Muliadi, "IMPLEMENTASI FUZZY TSUKAMOTO DALAM PENENTUAN KESESUAIAN LAHAN UNTUK TANAMAN KARET DAN KELAPA SAWIT," *KLIK - Kumpul. J. ILMU Komput.*, vol. 4, no. 2, p. 233, Sep. 2017, doi: 10.20527/klik.v4i2.115.

[11] T. D. Puspitasari, M. F. N. Jadid, and A. Trihariprasetya, "Sistem Pendukung Keputusan Pemilihan Tanaman Holtikultura dengan Metode Fuzzy Sebagai Upaya Meningkatkan Ketahanan Pangan," 2018.

[12] P. Harianja, A. Saleh, and M. B. Akbar, "SISTEM PENDUKUNG KEPUTUSAN PENENTUAN KUALITAS TANAMAN KARET MENGGUNAKAN METODE FUZZY MAMDANI (STUDI KASUS : PTPN III MEDAN)," 2018.

[13] S. Silvilestari, "Expert System Logika Fuzzy Penentuan Proses Penanaman Bibit Unggul Kayu Manis dengan Metode Mamdani," *Build. Inform. Technol. Sci. BITS*, vol. 3, no. 3, pp. 141–147, Dec. 2021, doi: 10.47065/bits.v3i3.1014.

[14] J. Tindaon, M. R. Lubis, and I. O. Kirana, "Analisis Tingkat Kepuasan Pelayanan Masyarakat Pada Dinas Lingkungan Hidup Pematangsiantar Menggunakan Metode Fuzzy Mamdani," *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 2, pp. 173–184, 2022.

[15] S. Arifin, M. A. Muslim, and S. Sugiman, "Implementasi Logika Fuzzy Mamdani untuk Mendeteksi Kerentanan Daerah Banjir di Semarang Utara," *Sci. J. Inform.*, vol. 2, no. 2, p. 179, Feb. 2016, doi: 10.15294/sji.v2i2.5086.

ISSN 2355-0082

Development and Implementation of a Corrosion Inhibitor Chatbot Using Bidirectional Long Short-Term Memory

Nibras Bahy Ardyansyah[1], Dzaki Asari Surya Putra[1], Nicholaus Verdhy Putranto[2], Gustina Alfa Trisnapradika[2], Muhammad Akrom[2]

[1] Informatics Engineering Program, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

[2] Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

Accepted 21 August 2024
Approved 14 March 2025

***Abstract*— This research delves into the intricate phenomenon of corrosion, a process entailing material degradation through chemical reactions with the environment, causing consequential losses across diverse sectors. In response, corrosion inhibitors are a proactive measure to counteract this deleterious impact. Despite their paramount significance, public awareness regarding corrosion and inhibitors remains limited; necessitating intensified educational efforts. The primary focus of this study is developing a Chatbot system designed to disseminate information on corrosion, inhibitors, and related topics. Employing the Machine Learning Life Cycle model, a deep learning approach, specifically the Bidirectional Long Short-Term Memory (BLSTM) architecture, is utilized to construct an optimized Chatbot model. Post-training evaluation of the BLSTM model reveals noteworthy performance metrics, including a remarkable 99% accuracy rate and a substantial 96% validation accuracy over 100 epochs. Training and validation losses are reported as 0.1112 and 0.4228, respectively. In conclusion, the BLSTM algorithm is an effective tool for training and enhancing Chatbot models, ensuring commendable corrosion awareness and inhibition performance.**

***Index Terms*— corrosion; inhibitor; Chatbot; deep learning; LSTM.**

I. INTRODUCTION

Corrosion is a process of material degradation that occurs through chemical reactions with the surrounding environment, causing structural damage and a decrease in material quality[1], [2]. Although it occurs frequently in everyday life, corrosion often goes unnoticed. Lack of awareness of corrosion symptoms can result in ignorance of the potential hazards that may arise due to structural damage and material deterioration. Environmental factors such as water, air, and certain chemicals play a role in accelerating this process[3]. Contact between two dissimilar metals in an electrolyte, microorganism interaction, as well as stress-induced corrosion can also accelerate material deterioration[4]. Corrosion not only impacts the degradation of materials and infrastructure but also has a wider impact, penetrating various interrelated sectors and playing an important role in modern life. including economic, environmental, and security industries[5], [6]. Corrosion prevention can be done through various methods that have been proven effective, such as metal surface coating, cathodic protection, and the use of corrosion inhibitors [7], [8]. The use of corrosion inhibitors is an important step that can inhibit the rate of corrosion reactions by adding certain chemical compounds into a corrosive environment[5], [9]. This method helps extend the material's service life and reduces maintenance and replacement costs caused by corrosion damage. However, general knowledge about corrosion and corrosion inhibitors is still limited. Educational efforts are needed to improve public understanding of corrosion. An effective information delivery method that is accessible to a wide range of people is key to raising awareness about corrosion prevention [10], [11].

Chatbot is a program designed to undergo human-computer interaction by utilizing Artificial Intelligence (AI) systems such as Natural Language Processing (NLP) [12], [13]. In addition, Chatbot can also be used as an effective tool to provide information related to technical issues such as corrosion. Implementing this Chatbot can create new opportunities to improve the accessibility of information regarding corrosion. Chatbots can accept various inputs, such as text, and provide responses based on pre-programmed patterns[14]. There are two main classifications in Chatbot, namely open domain and closed domain. Open-domain chatbots can respond appropriately to a wide range of general topics. In contrast, closed-domain chatbots can only provide responses related to specific topics and may not be effective in responding to other topics[15]. Chatbots have a high degree of flexibility, allowing for customization and training in multiple languages, thus meeting diverse and specific needs across different sectors. This capability allows chatbots to adapt to local languages, cultures, and user

preferences, ensuring effective and relevant interactions across multiple contexts. As a result of this flexibility, chatbots can function optimally in a wide range of industries, including customer service, education, and healthcare, providing solutions tailored to the evolving needs of the market [16]. In developing a *Chatbot*, a model that can undergo training and testing processes using machine learning algorithms, such as Neural Networks, is needed [17]. This journal discusses the implementation or development of NLP systems, namely the corrosion inhibitor *Chatbot* using the *Deep Learning* approach and the *Bidirectional Long Short-Term Memory* (BLSTM) algorithm.

Some methods commonly applied in Chatbot development involve Recurrent Neural Network and Long Short-Term Memory (RNN-LSTM) [18], Bidirectional Long Short-Term Memory (BLSTM), and Natural Language Processing (NLP) [19]. With a structure designed to remember information, the LSTM (Long Short-Term Memory) algorithm eases the interaction between users and computers, allowing for more natural and effective conversations. Due to its ability to handle sequential data and recognize complex patterns in text. LSTM enables natural language processing that makes it easier for users to interact with computer [20]. Therefore, the LSTM method is becoming very popular in chatbot development. With the use of LSTM, a chatbot can receive input and generate output based on previously learned patterns [21].

This research brings innovation by applying the LSTM algorithm to Chatbot development. The main focus of the Chatbot is to provide responses related to user questions regarding information and knowledge about corrosion inhibitors and provide effective suggestions for corrosion prevention. The LSTM model training process was conducted independently using the Python platform and Jupyter Notebook. The strategic decision to choose LSTM as the main algorithm was based on the results of previous studies that showed superior classification accuracy compared to other methods, such as RNN and K-Kearnest Neighbors (KNN) [22]. Thus, this research proposes using LSTM as a more reliable foundation for building a Chatbot that can provide innovative solutions to questions and challenges surrounding corrosion inhibitors.

II. METHODOLOGY

This research was conducted through a series of machine learning model development processes. Figure 1 overviews this research's machine learning model development process. The development begins with collecting relevant data on corrosion and corrosion inhibitors and then exploring the data to understand its patterns and characteristics. Next, data preparation is performed, including data cleaning and transformation, to ensure its quality. Once the data is ready, a splitting stage is performed to divide the data into training and testing sets. A suitable algorithm is selected and trained using the training data in the modelling stage. Then, the model is evaluated using the testing data at the evaluation stage to measure the performance and accuracy of the model. Once the model is deemed adequate, the next step is requirement chatbot, where we test the chatbot model that has been developed to ensure that it meets the needs and then deployment, which is the implementation of the model into Streamlit to enable the model to be used effectively by users.

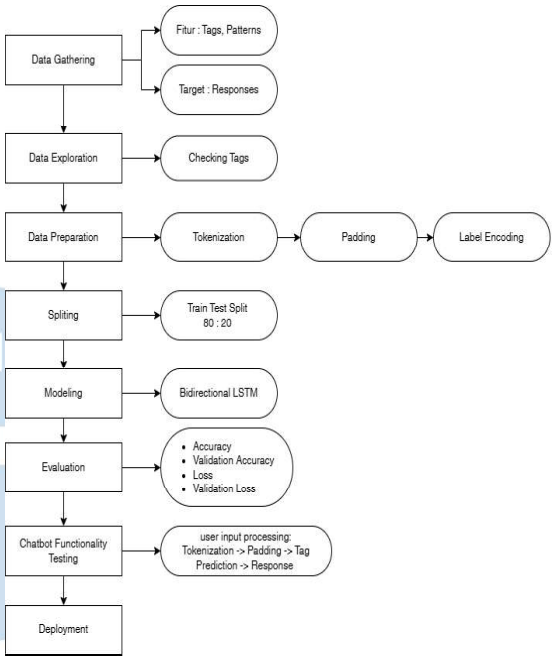


Fig. 1. ML model development Flowchart

A. Data Gathering

The initial stage of development involved data collection, which was conducted through research and the manual gathering of relevant data. This process was based on a literature review of corrosion and corrosion inhibitors from various reliable online sources. The choice of sources, such as Alodokter, HaloDoc, DrugBank, IDNmedis, Vinmec, PubChem, Drugs, Alomedika, Medicinka, and others, was driven by their relevance to the research topic. These platforms provide validated information on chemical substances, corrosion mechanisms, and inhibitors essential for building a high-quality dataset. The characteristics of the data were determined based on specific keywords related to corrosion and corrosion inhibitors. These keywords included corrosion, corrosion inhibitor, chemical substances, metal protection, and environmental impact. Data filtering used these keywords to ensure the collected information was relevant, accurate, and aligned with the research goals. This filtering process focused on extracting content that could serve as training patterns for the Chatbot. Additionally, the effort was made to include diverse

questions and answers related to corrosion, ensuring comprehensive coverage of user queries.

Once the data was collected, it was organized into a dataset to train the chatbot model algorithm. The dataset was stored in JavaScript Object Notation (JSON) format and designed with the following structure:

- Tags: To group similar text data and use the same output as a target to train the neural network.
- Patterns: A component containing input pattern data expected to match the user input. Patterns are used as predictors.
- Responses: This is part of the data that contains answers or outputs that will be sent based on index tags and patterns determined by the system.

B. Data Exploration

The process at this stage serves as the first qualification before conducting model development. The data collected includes information related to corrosion, corrosion prevention, inhibitors, types of inhibitors, inhibitor mechanisms, corrosion inhibitors, drugs, expired drugs, expired drugs, corrosion inhibitors, and names of chemical compounds. This data will be used as answers or responses in the dataset. The developed chatbot uses a single-response approach, where each question from the user will be answered with one short and direct response according to the context of the question asked.

C. Data Preparation

In this stage, data processing is carried out to convert raw data into data ready to be used in the machine learning model training process at a later stage. Data exploration and data preparation, also known as data preprocessing, aims to ensure that the data used in Chatbot model training is clean, consistent, and representative so that it can produce better final results[22], [23]. The stages involved transforming the data into a suitable dataset format, using the panda's library to facilitate manipulation or modification, performing tokenization to convert the text in the pattern into a sequence of numbers, adding padding to the sequence of numbers to have the same length, and converting tag labels into numbers using the LabelEncoder technique. The data that has passed these stages has been optimally prepared for use in model training.

D. Splitting

The next stage is splitting the preprocessed dataset into two main parts: training and validation data. This division aims to train the model with some data and test its performance using data that the model has never seen before. The training process is performed for 100 epochs to improve the accuracy and performance of the model. The train test split function divides the dataset with a proportion of 80:20, where 80% becomes training data, and 20% becomes validation data. The training data is used as input and output for the model, while the validation data is used to evaluate the model's performance during training.

E. Modeling

The next stage in development involves the modelling process, which includes selecting an effective algorithm to recognize and respond to user queries accurately. The deep learning algorithm chosen for training the Chatbot is LSTM, which is known for achieving a high level of accuracy, superior response capabilities, and quick adaptability [24]. This study enhances the LSTM model by implementing a Bidirectional LSTM (BLSTM) structure. BLSTM processes sequential data in both forward and backward directions, making it particularly effective in understanding the context of user input. BLSTM was selected due to its ability to capture relationships between preceding and succeeding words in a sentence, which is crucial for NLP tasks such as Chatbot development. This capability allows the Chatbot to generate more accurate and context-aware responses than standard LSTM models that process data in only one direction. The model structure includes an Embedding layer, Bidirectional LSTM, Dropout layer, Normalization layer, Dense layer, and an additional LSTM layer. The model is compiled using the Adam optimizer and the categorical_crossentropy loss function, with accuracy as the primary evaluation metric. The model is trained using preprocessed data to maximize its performance in effectively recognizing and responding to user queries.

F. Evaluation

In this evaluation stage, the results of the trained model are displayed using an evaluation matrix. The evaluation metrics used include accuracy, validation accuracy, loss, and validation loss. In this model, the loss function applied is categorical cross-entropy, while the optimization uses the Adam method. Furthermore, to monitor the model's performance at each epoch, visualization is performed by displaying training accuracy, validation accuracy, training loss, and validation loss graphs. This evaluation process provides a holistic picture of the extent to which the model can provide accurate and consistent responses. The accuracy and loss graphs provide insight into the model's performance during the training and validation. After the evaluation process, the best-trained model can be selected for use in the Chatbot implementation[25].

G. Requirement Chatbot

At this stage, the best Chatbot model is used for the user input function, which aims to provide interaction between the user and the trained model. The user enters sentence patterns as input, and this function will process, perform class prediction, and display random responses from the Chatbot based on the pre-trained model. The preprocessing process involves several

steps: character removal, conversion to lowercase, and tokenization. After getting the results from tokenization, the data is converted into a numeric sequence and padding is performed if needed. Next, the model performs tag label prediction, and the corresponding response is retrieved from the response dataset based on the tag label. The result of this interaction is to display the user input and a randomized response from the Chatbot.

H. Deployment

At this stage, the Chatbot model will be deployed for users to access online. In this research, the deployment process uses the Streamlit framework with the Python programming language. An environment with the library scikit-learn version 1.3.2, tensorflow 2.12.0, numpy 1.23, and Streamlit 1.29.0 is required to run the development process.

III. Result and Discussion

This research describes the development of a Chatbot to help improve public understanding of corrosion. This chatbot was developed using machine learning algorithms based on natural language processing (NLP) techniques. These methodologies enable the Chatbot to process natural language input effectively, providing accurate and contextually relevant responses to users' inquiries. The dataset used to train the Chatbot model algorithm is a manual dataset in the form of a JSON file. This dataset stores several power components, namely intents, tags, patterns, and responses, forming the foundation for the Chatbot's decision-making and response-generation capabilities. The structure of the dataset in Figure 2 allows the Chatbot to understand and provide appropriate responses to the questions asked by the user according to the specified topic. The model development involved using the collected dataset in JSON form and converting it into a data frame consisting of pattern and tag columns as the main columns. This process is done to organize the data to make it easier to process and structure. The data in the response column is used to provide answers that match the question based on the tags generated by the model, and they are randomly selected to increase the variety of responses. This dataset has 1386 rows representing diverse user inputs and is categorized into 298 unique tags. This extensive and well-curated dataset ensures the Chatbot's robustness and adaptability, allowing it to handle a broad spectrum of corrosion-related questions. By leveraging this comprehensive data, the Chatbot is better equipped to assist users, delivering clear and informative responses while enhancing the overall user experience.

Tokenization is applied to the pattern text, transforming it into a sequence of numerical representations. This step is critical in data preprocessing for natural language processing (NLP) models. Subsequently, padding is performed by adding zeros as prefixes or suffixes, standardizing the length of the numerical sequences across all data samples to maintain uniform input dimensions during the training process. Following this, the Label Encoding process is applied to the target variable. Specifically, the data in the tag column converts categorical labels into a numeric representation, typically in the form of binary vectors. The results of data preparation or preparation in x and y can be seen in Figure 3.

Fig. 2. JSON database

```
X shape =  (1387, 10)
y shape =  (1387,)
num of classes =  299
```

```
X
```

```
array([[339,   0,   0, ...,   0,   0,   0],
       [340,   0,   0, ...,   0,   0,   0],
       [339,  40,   0, ...,   0,   0,   0],
       ...,
       [  1,   2,   5, ...,   0,   0,   0],
       [  6,   4, 248, ...,   0,   0,   0],
       [  1,   2,   8, ...,   0,   0,   0]])
```

```
y
```

```
array([277, 277, 277, ..., 243, 243, 243])
```

Fig. 3. Data Preparation X dan Y

The Bidirectional LSTM deep learning algorithm is used to develop the Chatbot model on the data processed for the training process. In Figure 4, we can see the architectural structure of the model, which consists of several layers. The process starts with the input layer as the first layer, which receives a batch of sequences with a sequence length of 11. The second layer is the embedding layer, which receives the input from the previous layer and converts it into a vector with 100 dimensions. The third layer is a Bidirectional LSTM layer with 256 parameter units, which can generate a sequence of values in the input sequence and is equipped with a dropout to prevent overfitting. The

dropout layer is further used to prevent overfitting and improve the capabilities of the Chatbot. The normalization layer in each layer is used to optimize the training process. Then, the second LSTM layer is equipped with dropout and normalization to prevent overfitting and normalize the data. After that, a dense layer with 64 units and a ReLU activation function is used to provide non-linearity to the model. The dropout and normalization layers are again applied before the softmax activation function is used. The LSTM model was trained for 100 iterations (epochs) to achieve optimal results. The model was compiled using the Adam optimizer and the categorical_crossentropy loss function, with the evaluation matrix being accuracy.

```
_________________________________________________
Layer (type)                 Output Shape
=================================================
embedding (Embedding)        (None, 10, 100)

bidirectional (Bidirection   (None, 10, 256)
al)

dropout (Dropout)            (None, 10, 256)

layer_normalization (Layer   (None, 10, 256)
Normalization)

lstm_1 (LSTM)                (None, 128)

dropout_1 (Dropout)          (None, 128)

layer_normalization_1 (Lay   (None, 128)
erNormalization)

dense (Dense)                (None, 64)

layer_normalization_2 (Lay   (None, 64)
erNormalization)

dropout_2 (Dropout)          (None, 64)

dense_1 (Dense)              (None, 299)

=================================================
```

Fig. 4. LSTM Model Structure

The model training process starts by dividing the dataset into two main parts: training and validation data. This division aims to train the model using part of the data and test its performance using data that the model has never seen before. The training process is carried out for 100 epochs to improve the accuracy and performance of the model. In the training implementation, the train test split function divides the dataset with a proportion of 80:20, where 80% of the data becomes the training part and 20% becomes the validation part. The training data is used as input and output for the model, while the validation data is used to evaluate the model's performance during the training process. This process aims to allow the model to learn from the training data.

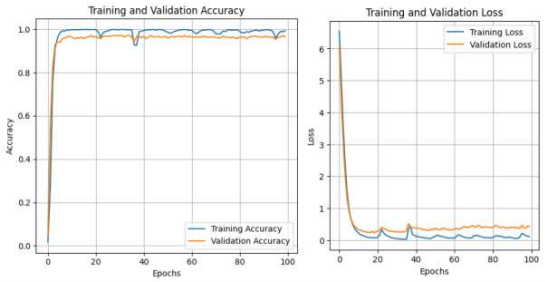


Fig. 5. Accuracy and loss graph of training and validation

TABLE I. ACCURACY AND LOSS METRICS, ALONG WITH VALIDATION OF LSTM TRAINING

Epoch	Traning Validation Result			
	Accuracy	*Val_Accuracy*	*Loss*	*Val_Loss*
10	0.9964	0.9676	0.2844	0.3861
20	0.9991	0.9640	0.0761	0.2735
30	0.9973	0.9712	0.0565	0.2676
40	0.9919	0.9604	0.1542	0.3889
50	0.9910	0.9676	0.1010	0.3320
60	0.9955	0.9676	0.0658	0.3302
70	0.9928	0.9640	0.0702	0.4597
80	0.9973	0.9568	0.0734	0.3977
90	0.9946	0.9640	0.0863	0.4090
100	0.9937	0.9640	0.1112	0.4228

Based on Table 1, after going through 100 epochs, the LSTM model achieved a high accuracy of 99% and a validation accuracy of 96.40% with a loss value of 0.0863 and a validation loss of 0.1112. These results show that the model performs excellently on training data and can predict well on validation data. This model's evaluation metrics include accuracy, validation accuracy, loss, and validation loss. The loss function applied to this model is sparse categorical cross-entropy, and the optimizer used is Adam. Model performance evaluation is done by visualizing the graphs of Training Accuracy, Validation Accuracy, Training Loss, and Validation Loss. In Figure 5, it can be seen that the training results show the ability of the model to learn the training data well. However, it should be noted that the validation accuracy is slightly lower than the training accuracy, which may indicate overfitting. Similarly, the loss graph shows that the validation loss reduces the training loss and validation loss values quickly, but the value fluctuates after a few epochs. Therefore, these results show that the model can understand the patterns in the data well and provide good predictions.

Fig. 6. Results of testing the chatbot

After preprocessing, the model predicts the tag label based on the input, and the corresponding response is taken from the pre-trained responses dataset. Figure 6 shows several test results by entering various questions, demonstrating that the Chatbot model can predict answers accurately. Unlike a Frequently Asked Questions (FAQ) page, which relies on static keyword-based search, this Chatbot uses a machine learning model based on Bidirectional Long Short-Term Memory (BLSTM). This allows the Chatbot to understand the context of user input and provide more dynamic, accurate, and context-aware responses. This capability makes the Chatbot a more effective and versatile tool for delivering information than traditional FAQ systems. Deployment is done so that users can use the Chatbot model created. The Chatbot model in this study was built into a web-based application using a framework or library in the Python programming language, Streamlit, to design the user interface. The results of the deployment stage can be seen in Figure 7, which shows the Chatbot web application interface. Users can enter a message or question into the column provided, and by pressing the send button, the Chatbot will provide an answer that matches the question. If the user wants to ask further questions, they can enter a new message and send it to the Chatbot again in the same way. Unlike previous studies that used LSTM models, this study implements Bidirectional LSTM (BLSTM). This innovation improves the model's ability to capture context in user queries, resulting in better accuracy and more relevant responses. This study outperforms previous research by achieving a higher validation accuracy of 96.40%, compared to 82.67% in earlier studies that used LSTM. Despite some indications of overfitting, the model remains capable of providing accurate and appropriate responses.

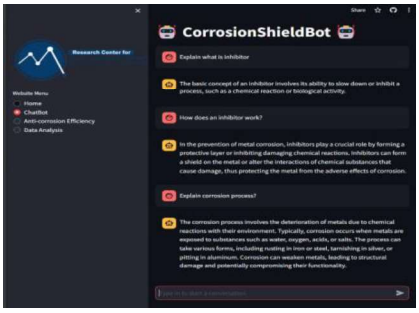


Fig. 7. Chatbot website interface

IV. CONCLUSION

Implementing the Long Short-Term Memory (LSTM) algorithm on the corrosion inhibitor Chatbot website has proven instrumental in enhancing the understanding of corrosion. The Chatbot training model, utilizing the LSTM algorithm, has demonstrated impressive performance, achieving an accuracy rate of 99% and a minimal loss value of 0.1112. Furthermore, integrating the Chatbot into a web-based application using the Streamlit framework has expanded its accessibility, allowing users to access it online and locally. The successful implementation of the Long Short-Term Memory (LSTM) algorithm on the corrosion inhibitor Chatbot website proved instrumental in improving the understanding of corrosion. The Chatbot training model, which utilizes the LSTM algorithm, has demonstrated commendable performance, achieving a remarkable accuracy rate of 99% and a minimal loss of 0.1112. Moreover, the successful integration of the Chatbot into a web-based application using the Streamlit framework or library has expanded its accessibility, allowing users to access it both online and locally. Designed to deliver information about corrosion, this Chatbot is an educational tool that is beneficial and relevant to a wide range of users. Nevertheless, it is important to recognize the limitations of this study's data. Future research efforts should prioritise testing and training on larger data sets to improve the reliability of the Chatbot and gain a more comprehensive understanding of the effectiveness and efficiency of the algorithms used. This approach ensures the continuous evolution and refinement of the Chatbot's capabilities, contributing to disseminating more accurate and robust corrosion information.

REFERENCES

[1] C. A. P. Sumarjono, M. Akrom, and G. A. Trisnapradika, "Perbandingan Model Machine Learning Terbaik untuk Memprediksi Kemampuan Penghambatan Korosi oleh Senyawa Benzimidazole," *Techno.Com*, vol. 22, no. 4, Art. no. 4, Nov. 2023, doi: 10.33633/tc.v22i4.9201.

[2] M. Akrom, DFT Investigation of Syzygium Aromaticum and Nicotiana Tabacum Extracts as Corrosion Inhibitor, Science Tech: Jurnal Ilmu Pengetahuan dan Teknologi, Volume 8, Issue 1, Pages 42-48 (2022), https://doi.org/10.30738/st.vol8.no1.a11775.

[3] M. Akrom, T. Sutojo, Investigasi Model Machine Learning Berbasis QSPR pada Inhibitor Korosi Pirimidin Investigation of QSPR-Based Machine Learning Models in Pyrimidine Corrosion Inhibitors, Eksergi, 20(2), (2023), https://doi.org/10.31315/e.v20i2.9864.

[4] M. Akrom, "INVESTIGATION OF NATURAL EXTRACTS AS GREEN CORROSION INHIBITORS IN STEEL USING DENSITY FUNCTIONAL THEORY," *Jurnal Teori dan Aplikasi Fisika*, vol. 10, no. 1, Art. no. 1, Jan. 2022.

[5] M. Akrom *et al.*, "Artificial Intelligence Berbasis QSPR Dalam Kajian Inhibitor Korosi," *JoMMiT : Jurnal Multi Media dan IT*, vol. 7, no. 1, pp. 015–020, Jul. 2023, doi: 10.46961/jommit.v7i1.721.

[6] M. Akrom, Green corrosion inhibitors for iron alloys: a comprehensive review of integrating data-driven forecasting, density functional theory simulations, and experimental investigation, Journal of Multiscale Materials

Informatics, Volume 1, Issue 1, Pages 22-37 (2024), https://doi.org/10.62411/jimat.v1i1.10495.

[7] M. Akrom, S. Rustad, H.K. Dipojono, R. Maezono, A comprehensive approach utilizing quantum machine learning in the study of corrosion inhibition on quinoxaline compounds, Artificial Intelligence Chemistry, Volume 2, Issue 2, Pages 100073 (2024), https://doi.org/10.1016/j.aichem.2024.100073.

[8] N. V. Putranto, M. Akrom, and G. A. Trinapradika, "Implementasi Fungsi Polinomial pada Algoritma Gradient Boosting Regressor: Studi Regresi pada Dataset Obat-Obatan Kadaluarsa Sebagai Material Antikorosi," *Jurnal Teknologi dan Manajemen Informatika*, vol. 9, no. 2, Art. no. 2, Dec. 2023, doi: 10.26905/jtmi.v9i2.11192.

[9] S. Marzorati, L. Verotta, and S. Trasatti, "Green Corrosion Inhibitors from Natural Sources and Biomass Wastes," *Molecules*, vol. 24, no. 1, p. 48, Dec. 2018, doi: 10.3390/molecules24010048.

[10] Z. M. Hanif, "Pengembangan Aplikasi Whatsapp Chatbot Untuk Pelayanan Akademik Di Perguruan Tinggi," Dec. 2021, Accessed: Jan. 09, 2024. [Online]. Available: https://dspace.uii.ac.id/handle/123456789/37445

[11] R. A. Yunmar and I. W. W. Wisesa, "Pengembangan Mobile based Question Answering System dengan Basis Pengetahuan Ontologi," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 4, Art. no. 4, Aug. 2020, doi: 10.25126/jtiik.2020742255.

[12] H. A. F. Muhyidin and L. Venica, "Pengembangan Chatbot untuk Meningkatkan Pengetahuan dan Kesadaran Keamanan Siber Menggunakan Long Short-Term Memory," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 5, no. 2, pp. 152–161, Oct. 2023, doi: 10.36499/jinrpl.v5i2.8818.

[13] E. L. Amalia and D. W. Wibowo, "Rancang Bangun Chatbot Untuk Meningkatkan Performa Bisnis," *jitika*, vol. 13, no. 2, p. 137, Oct. 2019, doi: 10.32815/jitika.v13i2.410.

[14] S. Rustad, M. Akrom, T. Sutojo, H.K. Dipojono, A feature restoration for machine learning on anti-corrosion materials, Case Studies in Chemical and Environmental Engineering, Volume 10, Pages 100902 (2024), https://doi.org/10.1016/j.cscee.2024.100902.

[15] K. Nimavat and T. Champaneria, "Chatbots: An overview. Types, Architecture, Tools and Future Possibilities," Oct. 2017.

[16] A. Elcholiqi and A. Musdholifah, "Chatbot in Bahasa Indonesia using NLP to Provide Banking Information," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 1, Art. no. 1, Jan. 2020, doi: 10.22146/ijccs.41289.

[17] R. Mahendra and M. Kamayani, "Menerapkan Algoritma Neural Network Pada Chatbot Mengenai Pariwisata Di Provinsi Bangka Belitung," *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 7, no. 2, Art. no. 2, Sep. 2023, doi: 10.30645/j-sakti.v7i2.678.

[18] P. Anki, A. Bustamam, H. S. Al-Ash, and D. Sarwinda, "Intelligent Chatbot Adapted from Question and Answer System Using RNN-LSTM Model," *J. Phys.: Conf. Ser.*, vol. 1844, no. 1, p. 012001, Mar. 2021, doi: 10.1088/1742-6596/1844/1/012001.

[19] V. R. Prasetyo, N. Benarkah, and V. J. Chrisintha, "Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya," *Teknika*, vol. 10, no. 2, Art. no. 2, Jul. 2021, doi: 10.34148/teknika.v10i2.370.

[20] P. B. Wintoro, H. Hermawan, M. A. Muda, and Y. Mulyani, "Implementasi Long Short-Term Memory pada Chatbot Informasi Akademik Teknik Informatika Unila," *Expert J. Manaj. Sist. Inf. dan Teknol.*, vol. 12, no. 1, p. 68, Jun. 2022, doi: 10.36448/expert.v12i1.2593.

[21] 1815061012 Hilmi Hermawan, "IMPLEMENTASI LONG SHORT-TERM MEMORY (LSTM) PADA CHATBOT INFORMASI AKADEMIK DI PROGRAM STUDI TEKNIK INFORMATIKA UNIVERSITAS LAMPUNG." Accessed: Jan. 09, 2024. [Online]. Available: https://digilib.unila.ac.id/65316/

[22] A. Silvanie and R. Subekti, "APLIKASI CHATBOT UNTUK FAQ AKADEMIK DI IBI-K57 DENGAN LSTM DAN PENYEMATAN KATA," *JIKO (Jurnal Informatika dan Komputer)*, vol. 5, no. 1, Art. no. 1, Apr. 2022, doi: 10.33387/jiko.v5i1.3703.

[23] K. A. Nugraha and D. Sebastian, "Chatbot Layanan Akademik Menggunakan K-Nearest Neighbor," *JSI*, vol. 7, no. 1, pp. 11–19, Mar. 2021, doi: 10.34128/jsi.v7i1.285.

[24] Y. Denny, H. L. H. Spits Warnars, W. Budiharto, A. I. Kistijantoro, Y. Heryadi, and L. Lukas, "Lstm And Simple Rnn Comparison In The Problem Of Sequence To Sequence On Conversation Data Using Bahasa Indonesia," Sep. 2018, pp. 51–56. doi: 10.1109/INAPR.2018.8627029.

[25] F. Zakariya, J. Zeniarja, and S. Winarno, "Pengembangan Chatbot Kesehatan Mental Menggunakan Algoritma Long Short-Term Memory," vol. 8, 2024.

ISSN 2355-0082

Beyond Traditional Methods: The Power of Bi-LSTM in Transforming Customer Review Sentiment Analysis

Casey Tjiptadjaja[1] and Moeljono Widjaja[2]
[1,2]Department of Informatics, Universitas Multimedia Nusantara, Tangerang 15810, Indonesia
[1]casey.tjiptadjaja@student.umn.ac.id, [2]moeljono.widjaja@umn.ac.id

Accepted 11 September 2024
Approved 14 March 2025

***Abstract*— In the current generation, many large and small companies compete fiercely to create things better than those on the market, such as smartphones, TVs, and many other things. One way they can do this is by guaranteeing the quality of services or goods that are better than others. The provider must investigate the feedback of their users or customers to improve the quality of service of the goods or services offered. Most medium and small companies, such as Micro, Small, and Medium enterprises (MSMEs), online stores, and so on, conduct research on customer feedback manually by looking at one-by-one feedback from customers, which is very ineffective and inefficient if a lot of customer feedback is obtained. Therefore, this research is conducted with the intention and purpose of helping medium and small companies analyze their customer sentiment, as well as trends over a certain period. This research will apply the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm to perform sentiment analysis on customer feedback. This research also compares other deep learning methods with the proposed method, namely the Uni-LSTM, GRU, CNN, and Simple-RNN algorithms. After testing, the accuracy results of the Uni-LSTM, Bi-LSTM, GRU, CNN, and Simple-RNN algorithms are 52.2%, 92.4%, 52.2%, 90.9%, and 49.7%, respectively. Then it was found that the Bi-LSTM algorithm with 64 units, five (5) training epochs, and a training and testing ratio of 80:20, as well as a model that implements 'relu' activation (rectified linear unit), obtained the maximum results, that is, got 92. 4% accuracy, which is the most superior accuracy compared to accuracy using other models and algorithms.**

***Index Terms*— Bi-LSTM; Customer Review; Deep Learning; Sentiment Analysis**

I. INTRODUCTION

In the 21st century, the urgency of community service is increasing. Society has progressed in various aspects of life, especially with the rapid development of technology and information. This development has opened access to information at all levels of society (information society), which makes it possible for people to compare different services between countries, companies, and individuals. Therefore, the demand for perfect service quality is inevitable. This must be balanced with better customer service because the company may integrate with the global society. In this context, it must be understood that society demands are growing, especially the demands for the quality of services provided by individuals or companies.

To improve the quality of service of the goods or services offered, the provider must research feedback from their users or customers. Most large companies already have a division that will research this customer feedback. For medium and small companies, such as Micro, Small, and Medium Enterprises (MSMEs), online stores (Instagram, Youtube, Tiktok, websites), most of them do this research manually by looking at customer feedback one by one. There are several reasons why this manual method is still used, such as the small number of customers and the cost factor. The manual method is still possible if they only have a small number of customers. However, if their customers are already in large numbers, such as hundreds and thousands, this method is no longer effective and efficient. Therefore, this research is done with the intention and purpose of helping medium and small companies analyze their customer sentiment and trends in a certain period, which in this research is in the context of months.

In the current era of globalization, transactions in the internet world are defined as e-Commerce [1]. In addition to promoting products that each seller owns, the e-Commerce department also handles reviews from E-Commerce users [2]. Over time, e-commerce users continue to grow and develop throughout the year, so the number of customer reviews is also increasing [3]. The development of information and communication technology makes it easier for humans to connect and share information through social networks [4]. Customers can write and post reviews about products and services on social media. Some of the most widely used social networks are Instagram, Twitter, TikTok, and YouTube.

Sentiment analysis, also known as opinion mining, is a technique used in natural language processing (NLP) to analyze and extract subjective information

from text, such as positive and negative emotions [5], [6]. This technique has gained widespread attention in research fields such as NLP and text analysis [7]. Not only among researchers, this technique is also widely used in businesses, governments, and organizations [8], especially to understand the public sentiment expressed in online comments and reviews. Sentiment analysis brings together various research fields, such as Natural Language Processing (NLP), data mining, and text mining, and is quickly becoming of great importance to organizations as they seek to integrate computational intelligence methods into their existing operations and seek to explain and improve their products and services [9]. There are different ways to perform sentiment analysis, ranging from rule-based methods that use a list of positive and negative terms as labeled data to train machine learning algorithms to build a classification [10].

Natural Language Processing (NLP) is a subset of artificial intelligence and linguistics that is devoted to helping computers understand statements or words written in human language. NLP was created to ease the user's work and satisfy the desire to communicate with computers in natural language [11]. In other words, Natural Language Processing aims to accommodate one or more specialization algorithms or systems. NLP assessment metrics in algorithmic systems enable the integration of language understanding and generation. Many things can be applied with NLP, such as detecting spam, becoming virtual agents or chatbots, and sentiment analysis [11].

Deep learning is a branch of machine learning [12]. The algorithm attempts to use high-level data abstraction using multiple processing layers consisting of complex structures or multiple non-linear transformations. Deep learning differs slightly from shallow learning in areas such as support vector machine and logistic regression. These shallow learning models have only one layer or no hidden layer nodes. Deep learning is based on multiple hidden layer nodes [13]. The core of deep learning is a multilayered network of nodes. Deep learning uses the input of the previous layer as the output of the next layer to learn highly abstract data features. So, machine learning algorithms (shallow learning) [14] are faster, but their performance or accuracy could be better. Meanwhile, deep learning methods perform better than the machine learning method for textual sentiment classification [15], [16].

Long-Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that is robust and suitable for handling sequential data with long-term dependencies. LSTM is a type of RNN designed to overcome the missing gradient problem, which is a common problem (limitation) in RNNs [17]–[19]. LSTM has a unique architecture called memory cells that allows it to learn long-term dependencies in data sequences [20]. LSTM algorithm is one of the deep learning methods that can be applied in Natural Language Processing (NLP), including speech recognition, text translation, text summarization, and sentiment analysis. There are two types of LSTM, namely UniLSTM (unidirectional LSTM) and Bi-LSTM (bidirectional LSTM) [21].

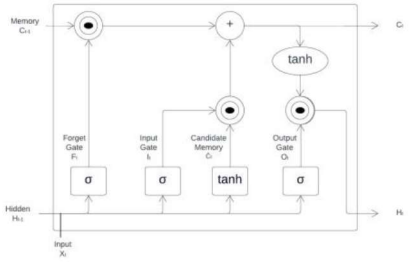


Fig. 1. Unidirectional LSTM Architecture
Source: Geeks For Geeks

Figure 1 shows the architecture of LSTM which consists of input gate, forget gate and output gate. Forget gate is where the information selection process occurs in the cell state using equation 1. Information will be discarded from the cell state if forget gate is 0, otherwise information will be stored in the cell state if forget gate is 1. Then at the input gate, information will pass through 2 layers, namely sigmoid and tanh. The cell state is generated from the output values of the two layers that have been combined. Then the process of updating the cell state value is carried out. Next, it will pass through the output gate, in which the output value of the cell state will be determined. The sigmoid layer is used to select the output based on the existing cell state, and will then be forwarded to the tanh layer.

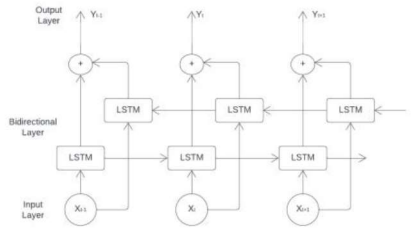


Fig. 2. Bidirectional LSTM Architecture
Source: Geeks For Geeks

The main difference between these two types of LSTM is that Uni-LSTM has the disadvantage [22] of only being able to process the input sequence in one direction (from beginning to end). In contrast, Bi-LSTM processes the input sequence in two directions (from beginning to end and from end to beginning) like shown in Figure 2 [23]. Uni-LSTM can only utilize information from past states. In contrast, Bi-LSTM can use information from past and future states to capture longterm dependencies and context in input sequences, making it suitable for tasks involving Natural Language Processing and data analysis [24]. The LSTM method is used in the research in [25]–[27] to

perform sentiment analysis and is one of the best algorithms to perform sentiment analysis compared to conventional methods (Support Vector, Na¨ıve Bayes, etc.).

Research in [25] conducted sentiment analysis comparing the Long Short-Term Memory (LSTM) and Na¨ıve Bayes (NB) algorithms, which found that LSTM has higher accuracy when compared to the NB algorithm; the LSTM algorithm has an accuracy rate of 72.85%, while the NB algorithm has an accuracy rate of 67.88%. In the study [26], a comparison of several algorithms was carried out to perform sentiment analysis, namely Simple Neural Network, Long Short-Term Memory, and Convolutional Neural Network, where the results show that the LSTM algorithm has an accuracy rate of 87%, which is the highest accuracy compared to the Simple Neural Network algorithm and CNN which has an algorithm of 81% and 82%. Gondhi [27] did a similar thing by comparing the LSTM, Na¨ıve Bayes, and Support Classification Algorithm algorithms to perform sentiment analysis, which the results show that the LSTM algorithm leads in terms of accuracy compared to the NB and SCA algorithms; the accuracy rate is 89%, 57.9%, and 67.7%, respectively.

Based on the problems described and some of the research presented, this research will use the Bi-LSTM method that uses Sigmoid activation, Relu hidden activation, Adam optimizer, and Binary cross-entropy loss function to perform sentiment analysis on customer reviews (feedback) using data on Tokopedia application review products from Mendeley Data [28]. This research also compares other deep learning methods with the proposed method, namely the Uni-LSTM, GRU, CNN, and Simple RNN algorithms. In conclusion, the contribution of this work presents a new algorithm (method) that significantly improves the accuracy of analyzing customer's sentiment and provides a comprehensive comparison of this method with other algorithm (method), highlighting its superior predictive capabilities. These advancements offer a more efficient approach for analyzing customer's sentiment, which could have wide applications in both MSMEs and companies.

II. METHODS

In this research, several stages are carried out to implement the Bi-LSTM algorithm for sentiment analysis of customer reviews (feedback). Figure 3 shows some of the steps taken in this research.

A. Data Collecting

The data used in this study are data obtained from the Mendeley Data website. The dataset was selected because it has balanced label statistics. Ideally, the data set should be balanced because a highly unbalanced dataset will make modeling difficult and lead to less precise accuracy. In the data used in this study, there are 42.75% of data with positive labels and 52.24% of data with negative labels. The total data used are 5400 data.

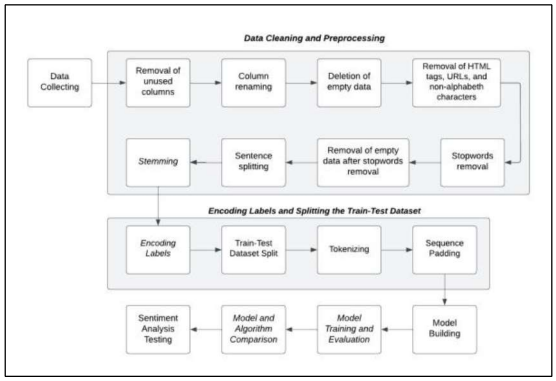


Fig. 3. The stages carried out in the research

B. Data Cleaning and Preprocessing

Data preprocessing is carried out at this stage, such as deleting some elements of the data set that have no value or will interfere with the analysis process. The following will explain in more detail the data cleaning and pre-processing process in this research, starting from the process at that stage until the results.

1) Deleting some unused columns using the drop function.
2) Rename the column using the rename function. The column previously "Customer Review" becomes "review," and the column "Sentiment" becomes "sentiment."
3) Deleting empty data using the dropna function. There is no empty data in the data used.
   Removal of HTML tags, URLs, and non-alphabetic characters. This is done with the help of functions from the Regex library. Table I shows the example of using the function that removes HTML tags, URLs, and nonalphabetic characters. In addition, all words are converted to lowercase using the 'lower' function. Table II shows an example of using the 'lower' function that converts all words to lowercase.
4) Remove stopwords from the corpus, such as 'which,' 'in,' and so on. This is done to not give special meaning to a sentence. This stage uses the library of an NLP, NLTK. Table III shows the process of removing stopwords.
5) After stopwords, there is a possibility that there are empty review data. Therefore, empty data are deleted after removing stopwords using the drop function.
6) Split the sentence into words using a function from the NLTK library, WhitespaceTokenizer(). This is done to facilitate the next stage of the stemming process. This is also used to see the ranking of words often mentioned. Table IV shows an example of breaking sentences into words using WhitespaceTokenizer().

TABLE I
THE RESULT OF REMOVING HTML TAGS, URLS, AND NON-ALPHABETIC CHARACTERS

Deletion Type	Before	After
HTML Tags	Penjual Ramah Melayani Pembeli Dengan Sabar dan Memberikan Berbagai Saran Yang Pembeli Tidak Tahu Mantap Lah	Penjual Ramah Melayani Pembeli Dengan Sabar dan Memberikan Berbagai Saran Yang Pembeli Tidak Tahu Mantap Lah
URL	mantap kipasnya kenceng https://www.tokopedia.com/, barangnya berkualitas sesuai sama harga	mantap kipasnya kenceng, barangnya berkualitas sesuai sama harga
Karakter yang bukan alfabet	Kualitas barang bagus, harga relatif murah, kiriman super cepat, response penjual baik, recommended seller, bintang 5, terima kasih yah, semoga SUKSES usaha-nya, aamiin ??????	Kualitas barang bagus harga relatif murah kiriman super cepat response penjual baik recommended seller bintang terima kasih yah semoga SUKSES usahanya aamiin

TABLE II
THE RESULT OF CONVERTING ALL WORDS INTO LOWERCASE LETTERS

Before	After
Penjual Ramah Melayani Pembeli Dengan Sabar dan Memberikan Berbagai Saran Yang Pembeli Tidak Tahu Mantap Lah	penjual ramah melayani pembeli dengan sabar dan memberikan berbagai saran yang pembeli tidak tahu mantap lah

TABLE III
RESULTS OF STOPWORDS REMOVAL

Before	After
barangnya sangat bagus	barangnya bagus
saya sangat kecewa dengan barang ini	kecewa barang

1) The stemming process is a technique that serves to retrieve the root of a word, commonly referred to as a lexicon. This is done to help reduce unnecessary computation in deciphering the whole word, where separate lemmas express most words' meaning well. This stage uses the library from Sastrawi.Stemmer.StemmerFactory. Table V shows an example of the stemming process.

TABLE IV
RESULT OF SENTENCE BREAKDOWN INTO WORDS

Before	After
barangnya sangat bagus	{barangnya}, {sangat}, {bagus}
saya sangat kecewa dengan barang ini	{saya}, {sangat}, {kecewa}, {dengan}, {barang}, {ini}

TABLE V
THE RESULT OF STEMMING PROCESS

Before	After
barang bagus berfungsi seler ramah pengiriman cepat	barang bagus fungsi seler ramah kirim cepat
barang mengecewakan	barang kecewa

C. *Encoding Labels and Splitting the Train-Test Dataset*

After data cleaning and preprocessing, the next stage is the encoding of labels and the division of the train-test dataset. Then, it will continue with the tokenizing process and sequence padding. Encoding is a method of converting one value into another, mostly from strings to numbers. This opportunity stage is done by converting the labels 'positive' and 'negative' into the numbers '1' and '0'. This is done with the LabelEncoder() function from the sklearn.preprocessing library.

After that, the data set is divided into train and test parts of 80% and 20% [29]. This is done using the train_test_split function from the sklearn.model_selection library. After dividing the train and the test, tokenizing and sequence padding is performed.

There are several parameters used in this research, namely vocab_size, oov_tok, embedding_dim, max_length, padding_type, and trunc_type. The parameter vocab_size is a parameter that specifies the number of dictionaries to be created. It is based on the total unique words when tokenized. The parameter oov_tok, or what stands for Out-of-Vocabulary, represents words that are not found in the vocabulary and are replaced during tokenization. The embedding_dim parameter represents the dimension of the embedding space of the word. It determines the size of the vector representation for words. The parameter max_length specifies the maximum length of a sentence. The padding_type parameter is a parameter that determines the padding in a sentence; in this study, the 'post' padding type is used, where padding is added after the end of the sentence. However, the trunc_type parameter reduces the number of words in a sentence if the sentence exceeds the number of max_length parameters. In this research, the *trunc* type 'post' is

used, where the word to be truncated is at the end of the sentence.

After these parameters are determined, the word is tokenized and converted into a sequence of integers, which uses the parameters vocab_size and oov_tok. This tokenizer builds a vocabulary based on the frequency of words in the training data ('training_sentences'). After that, the text_to_sequences function is used, which converts the sentences in the training and testing data into a sequence of integers using a tokenizer. Then, the pad_sequences function is also used, adding padding to adjust each sentence's length in the training and testing data.

1) Tokenizing: Tokenizing is a method of dividing a sentence into words and creating a dictionary of all the unique words found, and all the words are also assigned unique integers. Each sentence is converted into an array of integers representing all the words in it. In this research, the tokenizer API from the Keras library is used. Table VI shows an example of the tokenizing process.

TABLE VI
TOKENIZING PROCESS

Review	Tokenizing Result
barangnya bagus	kamus[0]: {barangnya}, kamus[1]: {bagus}
kecewa barang	kamus[2]: {kecewa}, kamus[3]: {barang}

2) Sequence Padding: At this stage, each array representing each sentence in the data set is filled with zeros from the end of the sentence to make the array size 200 and make all sentences the same length. Table VII shows an example of the sequence padding process with a maximum of 200 arrays and the 'post' method.

TABLE VII
SEQUENCE PADDING PROCESS

Review	Sequence Padding Result
barangnya bagus	[{barangnya}, {bagus}, {0}, {0}, ..., {0}]
kecewa barang	[{kecewa}, {barang}, {0}, {0}, ..., {0}]

D. Model Building

At this stage, like shown in Figure 4, the model is created with the Keras library and uses the Bi-LSTM method, which will output the probability of a positive sentence if the label is '1'. This model will be compiled using binary crossentropy loss and Adam's optimizer because the data contain binary classification. Adam's optimizer uses stochastic gradient descent to train the deep learning model and compare the probabilities of each predicted label class (0 or 1). Accuracy is used as the primary metric. Figure 5 shows the code snippet of the model implemented in this study. The created model uses the algorithm *Bidirectional* LSTM algorithm with 64 units.

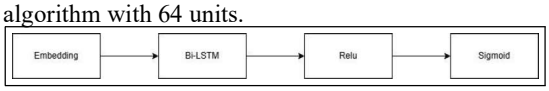


Fig. 4. Model Building

```
#INISIALISASI MODEL
model = keras.Sequential([
    keras.layers.Embedding(vocab_size, embedding_dim, input_length
    =max_length),
    keras.layers.Bidirectional(keras.layers.LSTM(64)),
    keras.layers.Dense(24, activation='relu'),
    keras.layers.Dense(1, activation='sigmoid')
])
#KOMPILASI MODEL
model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy'])
#BENTUKAN MODEL
model.summary()
```

Fig. 5. Model Building Code

E. Model Training and Evaluation

The model in this study was trained using five (5) epochs. After that, the model is evaluated by calculating its accuracy. The accuracy is calculated by dividing the number of correct predictions by the total number of predictions.

F. Model and Algorithm Comparison

In this study, several models are tried, both models that use training and testing data with a division of 70 and 30, models that use epochs of six (6), ten (10), and fifteen (15), models that use LSTM units of 32 and 128, and models that do not use relu activation (Rectified Linear Unit). This study also compares the Bi-LSTM algorithm with the Uni-LSTM, GRU, CNN, and Simple-RNN algorithms, most of which are algorithms used for comparison in previous studies, namely research in [25]–[27].

G. Sentiment analysis Testing

This test uses the same model as the one previously created with Bi-LSTM. Here, users can input strings in the form of sentences and can input files as well. The output for users who input sentences is the sentiment prediction result ('Positive' or 'Negative'), while for users who input files, the output will be in the form of a diagram that illustrates the percentage of total sentiment predictions that are 'Positive' and 'Negative,' as well as sentiment analysis trends in a certain period, namely per month. Users who input files will receive the prediction result file and the date. The test data used for string input is manual input, while the test data used for file input is restaurant review data from Kaggle which is manually dated.

III. RESULTS AND DISCUSSION

After several data cleaning and preprocessing stages, the results will be used and entered into the model made in this study.

A. Data Results Before and After Data Cleaning and Preprocessing

Data must be cleaned and preprocessed before being entered into the model that has been created to eliminate irrelevant data that can interfere with the training and testing process.

B. Model Implementation

The model implemented in this study uses the Bidirectional LSTM algorithm with 64 units. After various experiments, both with other algorithms such as Unidirectional LSTM, GRU, and CNN and with the number of units of 32 and 128, it was found that even the Bidirectional LSTM algorithm with 64 units produced the best accuracy.

Table VIII shows the accuracy results of the Uni-LSTM, BiLSTM, GRU, CNN, and Simple-RNN algorithms are 52.2%, 92.4%, 52.2%, 90.9%, and 49.7% respectively. Table IX shows the accuracy results of using Bi-LSTM with LSTM units of 32 units, 64 units, and 128 units are 90.7%, 92.4%, and 91.1%, respectively.

It can be seen that the model used starts from the embedding layer first, then the Bi-LSTM layer, and then the dense layer.

TABLE VIII
COMPARISON RESULTS OF ALGORITHMS

Algorithms	Accuracy Percentage
Uni-LSTM	52.2%
Bi-LSTM	92.4%
GRU	52.2%
CNN	90.9%
Simple-RNN	49.7%

TABLE IX
LSTM UNIT COMPARISON RESULT

Total LSTM Units	Accuracy Percentage
32 unit	90.7%
64 unit	92.4%
128 unit	91.1%

The embedding layer is used to convert integer indices into fixed-size dense vectors. Table X shows the process performed in the embedding layer. After that, the input goes into the BiLSTM layer, which processes data from both directions.

TABLE X
100-DIMENSIONAL LAYER EMBEDDING PROCESS

Review	Embedding result
barangnya bagus	{barangnya}: {01001.....1010}, {bagus}: {10101.....1001}
kecewa barang	{kecewa}: {11010.....0101}, {barang}: {00101.....0110}

Then, two dense layers are used in this research; the first is 'relu' (Rectified Linear Unit), and the second is 'sigmoid.' 'Relu' introduces non-linearity to the model, thus allowing the model to learn complex mappings between input and output features. This minimizes loss and helps to effectively propagate the gradient through the network, allowing for efficient training. The 'sigmoid' serves to output 0 or 1 (binary class). In this research, 'relu' activation is used because a comparison of models that use 'relu' and not is made. Table XI shows the accuracy comparison results of models that use 'relu' activation and not. The model that uses 'relu' has an accuracy rate of 92.4%, while the model that does not use 'relu' has an accuracy rate of 90.4%.

TABLE XI
COMPARISON RESULTS FOR THE USE OF 'RELU' ACTIVATION

'relu' Activation	Accuracy Percentage
Using 'relu' activation	92.4%
Not using 'relu' activation	90.4%

The loss function used in this study is 'binary_crossentropy,' a general loss function used for binary classification problems. Then, the 'Adam' optimizer updates the neural network weights during training. Adam is an adaptive optimization algorithm that performs well in various deep-learning tasks. Then, the evaluation metric used in this research is 'accuracy,' which is used to monitor the model's performance during training. Finally, the model architecture is notified. This is done to make it easier to check the model structure and help debug and optimize the model architecture.

C. Training and Testing Process

As explained earlier, in this study, the training and testing data division was carried out with a ratio of 80:20. This is because after experimenting with the 70:30 ratio, the accuracy results still need to be improved compared to the 80:20 ratio. Table XII shows the accuracy results for models with a training and testing ratio of 70:30 and 80:20. The accuracy of the model using a training and testing ratio of 80:20 is 92.4%, while the accuracy of the model using a training and testing ratio of 70:30 is 89.9%.

TABLE XII
COMPARISON RESULTS OF THE USE OF TRAINING AND TESTING RATIOS

Training dan Testing Ratio	Accuracy Percentage
Training 70% dan testing 30%	89.9%
Training 80% dan testing 20%	92.4%

D. Training Process

After the data is implemented into the model, the data will be trained first. In this research, training is done with an epoch of five (5). This was determined after experimenting with epochs six (6), ten (10), and

fifteen (15), the results of which are shown in Table XIII. The accuracy results obtained with epochs five (5), six (6), ten (10), and fifteen (15) are 92.4%, 89.9%, 90.1%, and 89.9%, respectively. Figure 6 and Figure 7 show the result plots of loss and accuracy values from training and validation. The orange line shows the result of data accuracy from the 'Validation' variable, which is the data used for validation. In contrast, the blue line shows the result of data accuracy from the 'Training' variable, which is the data used for training.

TABLE XIII
DIFFERENT EPOCH COMPARISON RESULTS

Total epoch	Accuracy Percentage
5 *epoch*	92.4%
6 *epoch*	89.9%
10 *epoch*	90.1%
15 *epoch*	89.9%

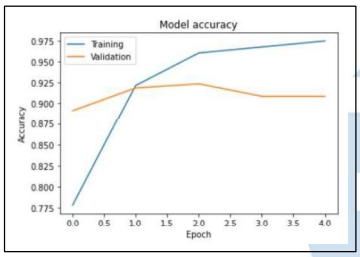


Fig. 6. Plot of training and validation accuracy results

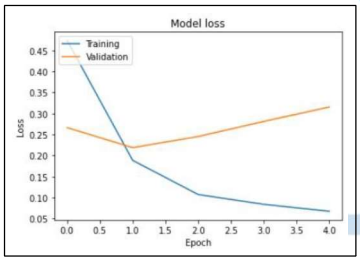


Fig. 7. Plot of training and validation loss values

E. Testing Process

After completing the training data, testing data will be carried out using the previously trained model. This is to determine the actual accuracy results, which with a training and testing ratio of 80:20, LSTM units of 64 units, and epochs of 5 produce very optimal accuracy results, which are 92.4%, this can be seen in Figure 8. Figure 8 also shows that the results of sentiment predictions more than 0.5 will go to the 'Positive' label, while prediction results less than 0.5 will go to the 'Negative' label.

```
43/43 [==============================] - 5s 82ms/step
Accuracy of prediction on test set :  0.9244444444444444
```

Fig. 8. Testing data accuracy results

In Figure 9, the results of the classification of confusion metrics in this study can also be seen. In Figure 9, we can see the results of the 'precision,' 'recall,' 'f1-score', and 'support' values. These values are the results obtained from the model that has been trained and *tested*. Figure 10 shows the results of the confusion metrics obtained in this study. The following section will explain the calculation formula for the data. Please note that TP (True Positive) is data that is in the lower right confusion matrix (positive, positive), FP (False Positive) is data that is in the upper right (positive, negative), FN (False Negative) is data that is in the lower left (negative, positive), and TN (True Negative) is data that is in the upper left (negative, negative). From the result shown, notice that the numbers of FN and FP are still quite large, which is a loss 10% after classification. This could be because many factors, such as insufficient training data, there is some of the data that uses another language, or some data have some typos that make the data invalid.

```
from sklearn import metrics
print(metrics.classification_report(test_labels,pred_labels))
```

	precision	recall	f1-score	support
0	0.95	0.91	0.93	705
1	0.90	0.94	0.92	645
accuracy			0.92	1350
macro avg	0.92	0.93	0.92	1350
weighted avg	0.93	0.92	0.92	1350

Fig. 9. Results from confusion metrics classification

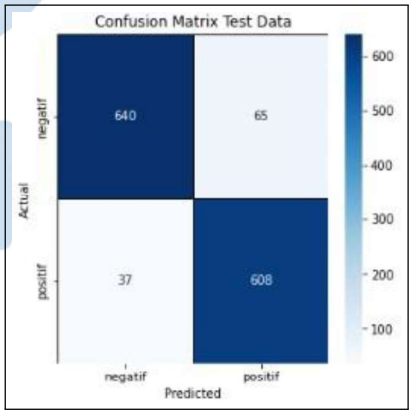


Fig. 10. Confusion metrics result

F. Usage

In this study, there are several types of use for the models that have been previously trained. There are sentiment analysis with manual sentence input, sentiment analysis with files without dates, and sentiment analysis and monthly sentiment trends with files that have dates. To get review files from customers, users can visit pages or download software

on the internet that provides data scrapping services, such as https://id.exportcomments.com, https://scrapy.org, and https: //webscrapper.io. After that, users can upload the data to platforms such as Jupyter Notebook, Google Colab, Kaggle, and others.

1) Manual sentence input: Here, the user can do sentiment analysis by inputting as many sentences as possible into the 'sentence' array to find the sentiment of these sentences. Figure 11 shows the test results performed by manually entering sentences. It can be seen that the results obtained are very accurate with the customer reviews, where positive reviews are predicted to have positive labels, and negative reviews are predicted to have negative labels.

```
1/1 [==============================] - 0s 161ms/step
Alhamdulillah barang sudah sampai, bisa dipakai semoga berfungsi dengan baik dan awet, kurirnya baik
smoga rezekinya lancar terus
Predicted sentiment :  Positive
Sangat kecewa. Baru 4 bulan scroll sudah rusak.
Predicted sentiment :  Negative
seller luar biasa.. packing baik, pengiriman cepat, penting nya barang sesuai deskripsi dan berfungs
i. terima kasih..
Predicted sentiment :  Positive
Penjualnya Ramah Melayani Pembeli Dengan Sabar dan Memberikan Berbagai Saran Yang Pembeli Tidak Tahu
<br /><br /> ... Mantap Lah ??????<br /><br />
Predicted sentiment :  Positive
Barangnya sangat bagus
Predicted sentiment :  Positive
Saya sangat kecewa dengan barang ini
Predicted sentiment :  Negative
```

Fig. 11. Results from testing by manually inputting sentences

2) Files without date: Here are the steps that users must take to do sentiment analysis using files without a date:

1) Perform data extraction using the previously mentioned pages or software regarding customer reviews on platforms such as Instagram, TikTok, etc.
2) After the data is obtained, the user can upload it to the platform used.
3) After a successful upload, the user can change the file name in the code according to the file name that was previously uploaded.
4) Then, if desired, the user can change the column index you want to remove or is unnecessary in this sentiment analysis. This can be done with the 'drop' function. In this research and testing, the columns used contain customer reviews and columns with sentiment labels only.
5) To make it easier for the user, the user can change the column's name containing the review data to 'review.' This can be done with the 'rename' function. In this test, we will change the name of the data column to be used, from the previous 'Customer Review' to 'review' and from the previous 'Sentiment' to 'sentiment'.
6) After that, the user only needs to run the entire customized code. The code has been modified to predict sentiment from customer reviews based on files uploaded by users before. The steps taken are the same as those taken in this research previously, mentioned in Chapter 3, from pre-processing data to tokenizing and padding data.
7) The user can see the sentiment results in a file called 'output.csv.' On the result, the pre-processed customer review data is entered into the column named 'Lemmatized Sentence' and the results of sentiment analysis (prediction) into the column called 'Predicted Sentiment.' The data is entered into a file with the format 'csv'.
8) Users can also see the positive and negative data ratio from the sentiment results obtained. Figure 12 shows the ratio of data before being processed into the model (the user cannot see this since the data uploaded by the user does not have sentiment labels). In contrast, Figure 13 shows the data ratio after input into the model. It can be seen in both figures that the difference in the label ratio is very little, which is about 1%.

```
Average length of each review :  16.095
Percentage of reviews with positive sentiment is 47.75925925925926%
Percentage of reviews with negative sentiment is 52.24074074074074%
```

Fig. 12. Before being entered into the model

```
Average length of each review :  11.273703703703704
Percentage of reviews with positive sentiment is 48.962962962962962%
Percentage of reviews with negative sentiment is 51.03703703703704%
```

Fig. 13. After being entered into the model

9) Users can also see the frequently mentioned words in each positive and negative label. Figure 14 shows 10 (ten) frequently mentioned words in the 'Positive' label and the total number of times they were mentioned. In contrast, Figure 15 shows 10 (ten) frequently mentioned words in the 'Negative' label and the total number of times they were mentioned.

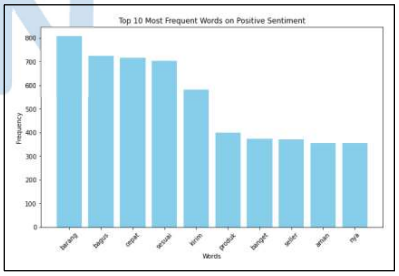


Fig. 14. Top 10 words on 'Positive' label

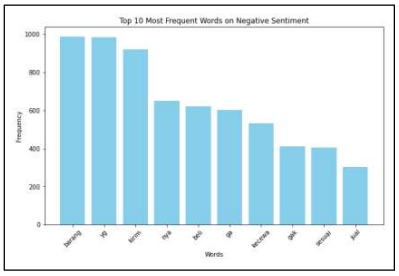


Fig. 15. Top 10 words on 'Negative' label

3) Files with date: Here are the steps that users must do to do sentiment analysis using files with dates:

1) Follow steps 1 - 6 from the usage steps of the file without a date.
2) The user can see the results of sentiment in the file 'output_date.csv' as shown in Figure 16. In the code in Figure 16, the pre-processed customer review data is entered into the column named 'Lemmatized_Sentence' and the results of sentiment analysis (prediction) into the column called 'Predicted Sentiment.' The data is entered into a file with the format 'csv'.

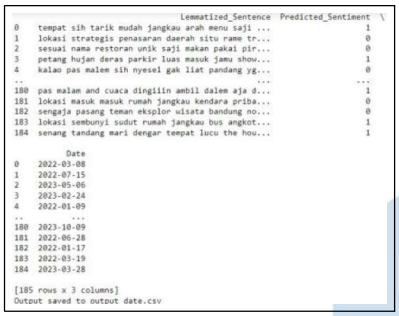


Fig. 16. The result of the processed data

3) Users can see the monthly trend results from the sentiment results obtained previously. Figure 17 shows the results of the trends obtained by the user. The orange plot illustrates the number of customer reviews with the label 'Negative,' while the blue plot illustrates the number of customer reviews with the label 'Positive.'

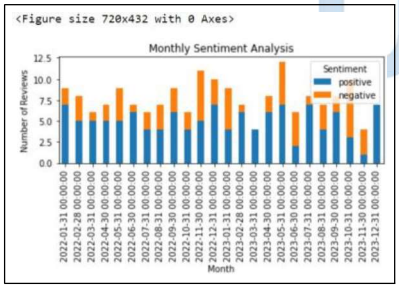


Fig. 17. Results of monthly sentiment trends

4) Users can also see the positive and negative data ratio from the sentiment results obtained. Figure 18 shows the data ratio before being processed into the model (the user cannot see this since the data uploaded by the user does not have sentiment labels). In contrast, Figure 19 shows the data ratio after inputting into the model. It can be seen in both figures that the difference in the label ratio is very little, which is about 1%.

```
Average length of each review :  38.78378378378378
Percentage of reviews with positive sentiment is 64.86486486486487%
Percentage of reviews with negative sentiment is 35.13513513513514%
```

Fig. 18. Before being entered into the model

```
Average length of each review :  22.335135135135136
Percentage of reviews with positive sentiment is 63.78378378378379%
Percentage of reviews with negative sentiment is 36.21621621621622%
```

Fig. 19. After being entered into the model

5) Users can also see the frequently mentioned words in each positive and negative label. Figure 20 shows 10 (ten) frequently mentioned words in the 'Positive' label and the total number of times they were mentioned. In contrast, Figure 21 shows 10 (ten) frequently mentioned words in the 'Negative' label and the total number of times they were mentioned.

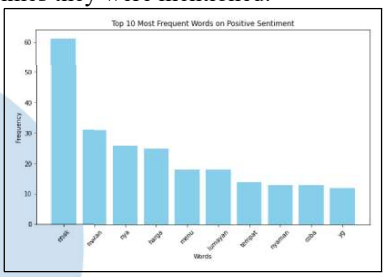


Fig. 20. Top 10 words on the 'Positive' label

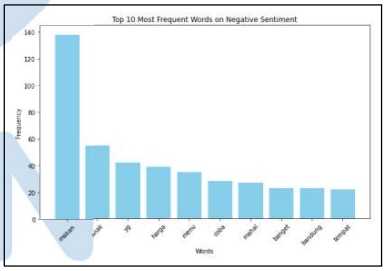


Fig. 21. Top 10 words on the 'Negative' label

IV. CONCLUSIONS

This research used 5400 customer review data samples to test sentiment analysis. This research implemented several deep learning algorithms, such as Bi-LSTM, Uni-LSTM, GRU, CNN, and Simple RNN. It was found that the BiLSTM algorithm with 64 units, five (5) training epochs, and a training and testing ratio of 80:20, as well as a model that implements 'relu' activation (Rectified Linear Unit), got the most maximum results, namely getting an accuracy of 92.4%, which is the most superior accuracy when compared to accuracy using other deep learning algorithms.

To obtain a better model performance, the user can add a stemming dictionary for unstandardized words and informal abbreviations, use a bilingual language

(Indonesian and English), use other algorithms than the algorithm that has been used in this research, because there are still many algorithms that are likely to get better model performance than the BiLSTM algorithm, and last but not least design (integrating) the website on the code or model that has been created so that testing for users can be more accessible, more effective, and efficient.

ACKNOWLEDGEMENTS

This research is fully funded by Universitas Multimedia Nusantara as part of the undergraduate thesis work.

REFERENCES

[1] K. Kasmi and A. N. Candra, "Penerapan e-commerce berbasis business to consumers untuk meningkatan penjualan produk makanan ringan khas pringsewu," *Jurnal AKTUAL*, vol. 15, no. 2, p. 109–116, Dec. 2017. [Online]. Available: http://dx.doi.org/10.47232/aktual.v15i2.27

[2] R. Nainggolan, F. Tobing, and E. Harianja, "Sentiment; clustering; k-means analysis sentiment in bukalapak comments with k-means clustering method," *IJNMT (International Journal of New Media Technology)*, vol. 9, no. 2, pp. 87–92, Jan. 2023. [Online]. Available: https://ejournals.umn.ac.id/index.php/IJNMT/article/view/2914

[3] P. H. Christian and R. I. Desanti, "The comparison of sentiment analysis algorithm for fake review detection of the leading online stores in indonesia," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, 2022, pp. 01–04.

[4] D. A. Kristiyanti, R. Aulianita, D. A. Putri, L. A. Utami, F. Agustini, and Z. I. Alfianti, "Sentiment classification twitter of lrt, mrt, and transjakarta transportation using support vector machine," in *2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*. IEEE, Nov. 2022. [Online]. Available: http://dx.doi.org/10.1109/ICSINTESA56431.2022.10041651

[5] B. Liu, *Sentiment Analysis and Opinion Mining*. Springer International Publishing, 2012. [Online]. Available: http://dx.doi. org/10.1007/978-3-031-02145-9

[6] S. Saad and B. Saberi, "Sentiment analysis or opinion mining: A review," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 5, p. 1660, Oct. 2017. [Online]. Available: http://dx.doi.org/10.18517/ijaseit.7.4.2137

[7] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, p. 5731–5780, Feb. 2022. [Online]. Available: http://dx.doi.org/10.1007/s10462-022-10144-1

[8] J. F. Sanchez-Rada and C. A. Iglesias, "Social context in sentiment´ analysis: Formal definition, overview of current trends and framework for comparison," *Information Fusion*, vol. 52, p. 344–356, Dec. 2019. [Online]. Available: http://dx.doi.org/10.1016/j.inffus.2019.05.003

[9] M. Farhadloo and E. Rolland, *Fundamentals of Sentiment Analysis and Its Applications*. Springer International Publishing, 2016, p. 1–24. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-30319-2 1

[10] F. Aftab, S. U. Bazai, S. Marjan, L. Baloch, S. Aslam, A. Amphawan, and T. K. Neo, "A comprehensive survey on sentiment analysis techniques," *International Journal of Technology*, vol. 14, no. 6, p. 1288, Oct. 2023. [Online]. Available: http://dx.doi.org/10.14716/ijtech. v14i6.6632

[11] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimedia Tools and Applications*, vol. 82, no. 3, p. 3713–3744, Jul. 2022. [Online]. Available: http://dx.doi.org/10.1007/s11042-022-13428-4

[12] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, p. 27–48, Apr. 2016. [Online]. Available: http://dx.doi.org/10. 1016/j.neucom.2015.09.116

[13] Z. Hao, "Deep learning review and discussion of its future development," *MATEC Web of Conferences*, vol. 277, p. 02035, 2019. [Online]. Available: http://dx.doi.org/10.1051/matecconf/201927702035

[14] K. Gulati, S. Saravana Kumar, R. Sarath Kumar Boddu, K. Sarvakar, D. Kumar Sharma, and M. Nomani, "Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to covid-19 pandemic," *Materials Today: Proceedings*, vol. 51, p. 38–41, 2022. [Online]. Available: http://dx.doi.org/10.1016/j.matpr.2021.04.364

[15] S. K. Bharti, R. K. Gupta, P. K. Shukla, W. A. Hatamleh, H. Tarazi, and S. J. Nuagah, "Multimodal sarcasm detection: A deep learning approach," *Wireless Communications and Mobile Computing*, vol. 2022, p. 1–10, May 2022. [Online]. Available: http://dx.doi.org/10.1155/2022/1653696

[16] P. D. Mahendhiran and S. Kannimuthu, "Deep learning techniques for polarity classification in multimodal sentiment analysis," *International Journal of Information Technology & Decision Making*, vol. 17, no. 03, p. 883–910, May 2018. [Online]. Available: http://dx.doi.org/10.1142/S0219622018500128

[17] B. Ghojogh and A. Ghodsi, "Recurrent neural networks and long shortterm memory networks: Tutorial and survey," 2023. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2023arXiv230411461G/abstract

[18] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13. JMLR.org, 2013, p. III–1310–III–1318. [Online]. Available: https://dl.acm.org/doi/10.5555/3042817.3043083

[19] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 06, no. 02, p. 107–116, Apr. 1998. [Online]. Available: http://dx.doi.org/10.1142/S0218488598000094

[20] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, p. 2222–2232, Oct. 2017. [Online]. Available: http://dx.doi.org/10.1109/TNNLS.2016. 2582924

[21] G. for Geeks, "Deep learning — introduction to long short term memory," 12 2023. [Online]. Available: https://www.geeksforgeeks.org/ deep-learning-introduction-to-long-short-term-memory/

[22] H. Elfaik and E. H. Nfaoui, "Deep bidirectional lstm network learning-based sentiment analysis for arabic text," *Journal of Intelligent Systems*, vol. 30, no. 1, p. 395–412, Dec. 2020. [Online]. Available: http://dx.doi.org/10.1515/jisys-2020-0021

[23] G. for Geeks, "Bidirectional lstm in nlp," 12 2023. [Online]. Available: https://www.geeksforgeeks.org/bidirectional-lstm-in-nlp/

[24] A. Agarwal, "Sentiment analysis using bi-directional lstm," 5 2020. [Online]. Available: https://www.linkedin.com/pulse/sentiment-analysis-using-bi-directional-lstm-ankit-agarwal

[25] M. A. Nurrohmat and A. SN, "Sentiment analysis of novel review using long short-term memory method," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*,

vol. 13, no. 3, p. 209, Jul. 2019. [Online]. Available: http://dx.doi.org/10.22146/ijccs.41236

[26] A. C. M. V. Srinivas, C. Satyanarayana, C. Divakar, and K. P. Sirisha, "Sentiment analysis using neural network and lstm," *IOP Conference Series: Materials Science and Engineering*, vol. 1074, no. 1, p. 012007, Feb. 2021. [Online]. Available: http://dx.doi.org/10.1088/1757-899X/ 1074/1/012007

[27] N. K. Gondhi, Chaahat, E. Sharma, A. H. Alharbi, R. Verma, and M. A. Shah, "Efficient long short-term memory-based sentiment analysis of e-commerce reviews," *Computational Intelligence and Neuroscience*, vol. 2022, p. 1–9, Jun. 2022. [Online]. Available: http://dx.doi.org/10.1155/2022/3464524

[28] said achmad, "Product reviews dataset for emotions classification tasks - indonesian (prdect-id) dataset," 2022. [Online]. Available: https://data.mendeley.com/datasets/574v66hf2v/1

[29] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation," 2 2018. [Online]. Available: https://scholarworks.utep.edu/cs techrep/ 1209/

ISSN 2355-0082

Mapping User Dissatisfaction in Mobile Banking Applications Using Ensemble Clustering LDA and LSA

Kartikadyota Kusumaningtyas[1], Alfirna Rizqi Lahitani[2], Irmma Dwijayanti[3], Muhammad Habibi[4]
[1,4]Informatics, Universitas Jenderal Achmad Yani Yogyakarta, Yogyakarta, Indonesia
[2] Information Technology, Universitas Jenderal Achmad Yani Yogyakarta, Yogyakarta, Indonesia
[3] Information System, Universitas Jenderal Achmad Yani Yogyakarta, Yogyakarta, Indonesia
[1]kartikadyota@gmail.com, [2]alfirnarizqi@gmail.com, [3]irmmadwijayanti@gmail.com,
[4]muhammadhabibi17@gmail.com

Accepted 30 September 2024
Approved 18 June 2025

***Abstract*— The large number of user complaints about mobile banking applications in Indonesia is a significant concern, considering the role of these applications in supporting people's financial activities. Negative reviews across the Google Play Store reflect user dissatisfaction with the application's features, performance, or services. This background encourages research to analyze negative user reviews in depth to identify the main topics of dissatisfaction. This study aims to map the topics contained in negative reviews using an ensemble clustering approach that combines the Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) methods. The research stages are divided into three main stages: Data Collection, Sentiment Analysis, and Topic Modeling. The built SVM model obtained evaluation results of 90% accuracy, 92% precision, 95% recall, and 94% F1-score. Based on the analysis of user review topics for mobile banking applications using the ensemble clustering LDA and LSA methods, it was found that each application has a different topic focus. The BCA mobile application is more associated with user interaction, ease of login process, features, and ease of use. Then BRImo complained about application service fees and speed, transaction security, transaction convenience, transaction balance, customer features, BRImo login, and BRImo application menu improvements.**

***Index Terms*— Topic Modeling; LDA; LSA; SVM; Ensemble Clustering.**

I. INTRODUCTION

The digital era has significantly changed various aspects of life, including financial services. Now, many banks provide online services to make transactions easier for their customers. Online banking services in Indonesia have experienced rapid growth in recent years. Based on data from Bank Indonesia, the value of digital banking transactions in August 2023 reached IDR 5.1 trillion or increased by 11.9% compared to 2022 [1]. This growth was driven by several factors, such as increasing internet penetration in Indonesia[2], government policies through the Gerakan Nasional Non-Tunai (GNTT) [3], and innovation from banks [4], [5]. Among the online services available, mobile banking is one of the most popular services because of the benefits felt by customers, such as ease of access, flexibility, and the ability to carry out various types of transactions, such as fund transfers, bill payments, and real-time balance monitoring [6], [7]. Amid the various conveniences and benefits offered, users of mobile banking applications also have an essential role in building a better online banking service ecosystem. One way is to share experiences in using mobile banking applications. The Google Play Store is an Android application store platform that allows users to provide ratings and comments on their experiences using the application. Positive reviews usually contain good experiences that indicate user satisfaction, while negative reviews usually contain bad experiences that indicate user complaints and dissatisfaction.

Currently, Google Play Store can automatically group user reviews into positive and negative. However, Google Play Store does not yet have a feature to map the main topics in positive and negative reviews automatically. This problem can make it difficult for service providers to understand the root of the problem and take appropriate corrective steps, especially for negative reviews. Negative reviews need to be handled more quickly. Slow handling of negative reviews can reduce customer reputation and loyalty [8].

This study aims to identify topics of user dissatisfaction based on negative reviews of popular mobile banking applications in Indonesia using ensemble clustering LDA and LSA. According to a survey conducted by Top Brand Award, two popular mobile banking applications in Indonesia are BCA Mobile and BRImo [9]. The application of the LDA method in this study was used to find the topics underlying the reviews, while the LSA method was used to reduce dimensions and extract semantic patterns from the review text. Previous research in digital banking has focused on understanding service quality and its impact on customer satisfaction and loyalty. Commonly used methods include multiple regression, Kendall-Tau correlation analysis, and SEM, which evaluate variables of responsiveness, device

compatibility, ease of use, risk, and service features that affect user satisfaction. Meanwhile, user satisfaction affects customer loyalty [10], [11], [12].

Furthermore, sentiment analysis on application reviews has become popular to measure user perception. This analysis utilizes machine learning techniques such as Support Vector Machine and Naive Bayes to analyze positive and negative sentiments contained in user reviews, providing an overview of factors that affect user experience [13], [14].

The main topics in mobile banking user review data were identified using the LDA method [15]. The data used were overall user reviews on the BNI, BCA, and BRI applications. Based on the data analysis, differences and similarities in service quality between the three applications were seen. These findings provide important insights into user preferences and experiences in using mobile banking applications. Given that complaint handling affects customer loyalty, mapping topics to user dissatisfaction is essential to encourage improvements in mobile banking services. This study introduces an ensemble approach that combines LDA and LSA to identify and analyze key topics that cause dissatisfaction in mobile banking user reviews [15], [16].

II. METHOD

This study consists of three main stages: data collection, sentiment analysis, and topic modeling (as shown in Fig. 1). It utilizes user reviews from the Google Play Store as the primary data source.

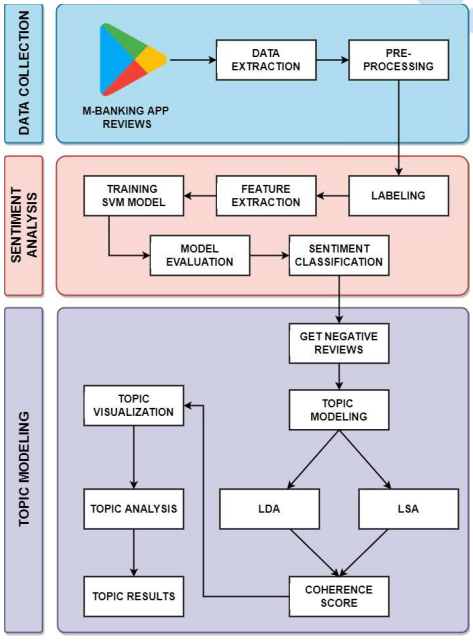


Fig. 1. Research phase

A. Data Collection

Data collection is the process of collecting mobile banking app review data. The input includes user reviews, ratings, and other Google Play Store platform metadata. At the same time, the output is a raw dataset in a structured format that is ready for further analysis.

The first step is to collect review data from mobile banking applications. This extraction data stage focuses on four mobile applications: BCA Mobile and BRImo. This study used a scraper to gather the data.

The next step is to clean the raw data through the preprocessing stages:

- Tokenization: The process of splitting sentences into tokens.
- Stopword removal: Removing common words that do not contribute significantly to the sentence's meaning, such as "and," "the," or "or."
- Stemming: Converting words to their base form.

B. Sentiment Analysis

The sentiment analysis process starts with the crucial labeling step, which involves categorizing text data based on sentiment to distinguish between positive and negative reviews. This labeling is essential for providing the SVM model with the necessary data to learn and make predictions.

Following labeling, the next step is feature extraction using the Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF is a method for assessing the importance of a word in a document relative to a set of documents. This process will change each word as a token that has gone through a preprocessing stage into a vector representing the existing word. Term Frequency (TF) measures how often a word appears in a document. Inverse Document Frequency (IDF) measures how unique or rare the word appears across all documents in the corpus. The equation for calculating TF-IDF is shown in (1).

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

Where:
ft,d= number of occurrences of word t in document d
N = total number of documents in the corpus
nt = number of documents containing word t

Support Vector Machine (SVM) is a supervised machine learning algorithm based on vectors [17]. SVM works by finding the best hyperplane (line or dividing plane) that separates data into two categories in the feature space [18]. The hyperplane is between two classes with a distance d at the closest point of each class. This distance *d* is called the margin and the points on this margin are called support vectors.

Once the feature extraction is complete, the SVM model is trained using data from the BCA Mobile application. The data is split into 70% for training and 30% for testing, ensuring that the model learns from most of the data while its performance is evaluated on the test data. The training process helps the model

understand the patterns in the data that correlate with positive or negative sentiment.

After training, the model's performance is assessed using a confusion matrix. A confusion matrix is a table used to evaluate the performance of a classification algorithm by comparing the actual labels with the predicted ones (as shown in Fig. 2). It summarizes the results into four categories: True Positives (correctly predicted positives), True Negatives (correctly predicted negatives), False Positives (incorrectly predicted positives), and False Negatives (incorrectly predicted negatives) [20].

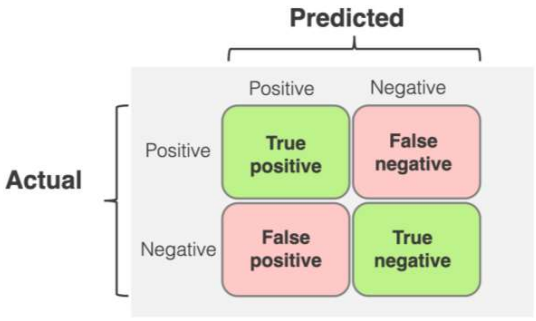


Fig. 2. Confusion matrix (Source: [21])

The confusion matrix provides the foundation for calculating key evaluation metrics in classification tasks. Accuracy measures the proportion of correct predictions (2), while precision evaluates the correctness of optimistic predictions by comparing true positives to all predicted positives (3). Recall (or sensitivity) measures the model's ability to correctly identify actual positive cases (4). The F1-score is the harmonic mean of precision and recall, offering a balanced metric when the class distribution is uneven (5).

$$Accuracy = \frac{TP}{TP+FP}\ x\ 100\% \quad (2)$$

$$Precision = \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (3)$$

$$Recall = \frac{TP}{TP+FN}\ x\ 100\% \quad (4)$$

$$F1-score = 2\ x\ \frac{Precision\ x\ Recall}{Precision+Recall}\ x\ 100\% \quad (5)$$

In the final step, sentiment classification is conducted using the trained SVM model, which classifies the reviews from four mobile banking applications into positive and negative categories.

C. Topic Modeling

Ensemble clustering is a approach to solve this problem. It combines two topic modeling methods to group data based on identified topics.

Latent Dirichlet Allocation (LDA) is an algorithm that distributes the words in the comments into several topics, where one comment can have one or more topics [22]. Latent Semantic Analysis (LSA) algorithm uses matrix decomposition to identify hidden relationships between words and documents, providing an overview of recurring topics in the comments [23].

Choosing the correct number of topics is a crucial step in topic modeling. A commonly used metric for this purpose is the coherence score, which evaluates how semantically consistent or interpretable a set of top words in a topic is. A higher coherence score suggests that the topic contains words that frequently appear together in the corpus and are more likely to be interpreted as a meaningful theme. In practice, multiple models with different topic numbers (k) are trained, and the model with the highest coherence score is selected as optimal.

Once the LDA or LSA model is trained, topic visualization becomes important to help users interpret and explore the results. Visualization tools are commonly used to display the inter-topic distance map and the relevance of terms to each topic. The visual interface allows users to explore how distinct or overlapping topics are and which words are most representative of each topic.

In topic analysis, the top keywords from each topic and their associated documents are reviewed to assign descriptive labels to the topics. This involves human judgment and is crucial for deriving actionable insights from the model output.

Each method (LDA and LSA) assigns a topic label to each comment. In the next stage, the results of LDA and LSA will be combined using the voting method. The voting method selects the most frequently occurring label as the final label (majority voting) [24].

III. RESULT AND DISCUSSION

A. Data Collection

The data collection phase is a critical process of gathering data from various sources. This study uses user reviews from BCA Mobile and BRImo mobile banking applications .

1. First, data is extracted from the Google Play Store site to get the raw user review comments. The result from this stage is text data.
   "Aplikasinya sangat membantu, transfer cepat dan mudah digunakan*!" (BCA)*
2. Preprocessing stage in this study involves tokenization, stopword removal and stemming. Result from this stage is clean data.
   "Aplikasi, sangat, bantu, transfer, cepat, mudah, guna"
3. Next, feature extraction using TF-IDF produce the weighted terms.

B. Sentiment Analysis

The BCA mobile application review dataset consisting of 1990 reviews was divided into 70% (1393 reviews) for training and 30% (597 reviews) for testing. The evaluation results of the SVM model are shown in Fig. 3.

```
Accuracy of the SVM model: 0.90
Precision: 0.92
Recall: 0.95
F1-Score: 0.94

Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.79      0.82        81
           1       0.92      0.95      0.94       213

    accuracy                           0.90       294
   macro avg       0.89      0.87      0.88       294
weighted avg       0.90      0.90      0.90       294
```

Fig. 3. Evaluation results

Based on the evaluation results shown in Figure 3, the SVM model performs well, with an accuracy of 95%, precision of 92%, recall of 95%, and F1-score of 94%. Furthermore, the model is used to classify all review data on each m-banking application. The distribution of negative and positive reviews on each m-banking application is shown in Fig. 4 and Fig. 5.

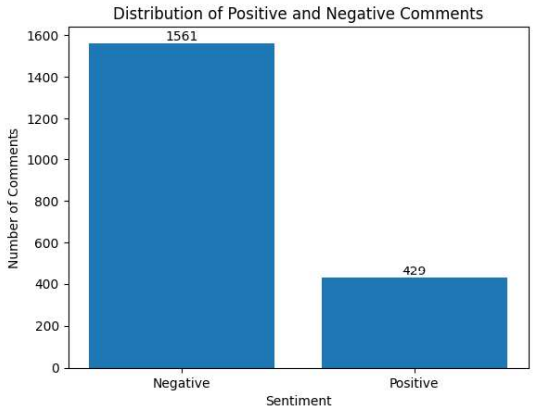


Fig. 4. Sentiment analysis of BCA Mobile reviews

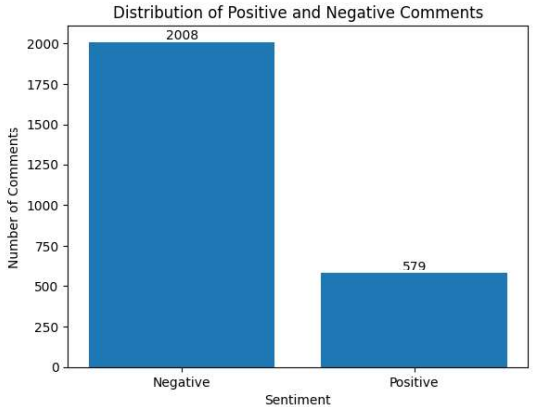


Fig. 5. Sentiment analysis of BRImo reviews

Based on the distribution of sentiment analysis from Fig. 4 and Fig. 5, the results show that negative sentiment data has a larger proportion than positive sentiment data.

C. *Topic Modeling*

In this experiment, we tried to generate 1 to 15 *k*-topics. Fig. 6 and Fig. 7 show the coherence score graphs of LDA and LSA on the BCA mobile and BRImo.

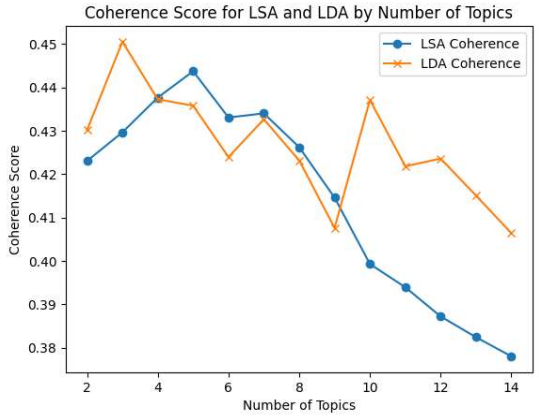


Fig. 6. Coherence score for LDA and LSA of BCA Mobile

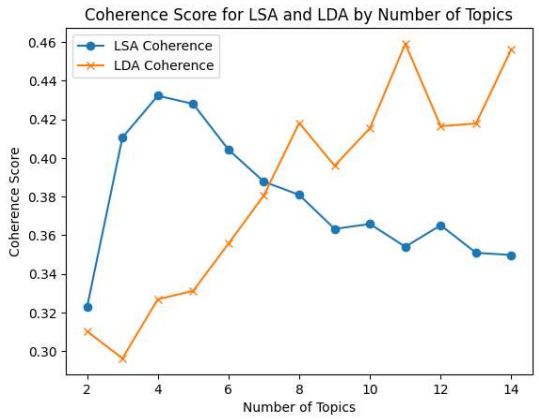


Fig. 7. Coherence score for LDA and LSA of BRImo

The higher the coherence score, the better the quality of the topic. Table 1 summarizes the highest coherence score and the number of topics for each m-banking application.

TABLE I. NUMBER K-TOPICS FOR LDA AND LSA

App	LDA		LSA	
	Coherence Score	Number of *k*-topics	Coherence Score	Number of *k*-topics
BCA mobile	0.450540	3	0.443739	5
BRImo	0.459125	11	0.432213	4

Fig. 8 and Fig. 9 show the results of the distance map between topics generated by the LDA method.

Meanwhile Fig. 10 and Fig. 11 show the results of t-SNE generated by the LSA method.

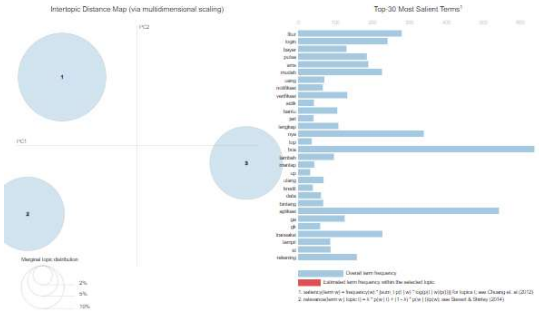


Fig. 8. Distance map generated by the LDA for BCA Mobile

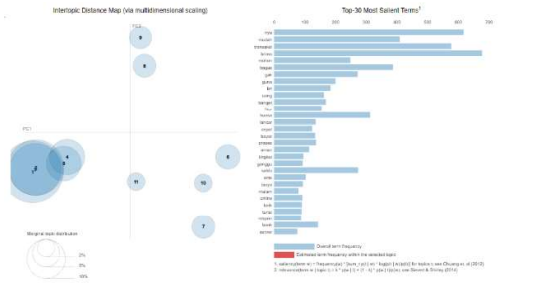


Fig. 9. Distance map generated by the LDA for BRImo

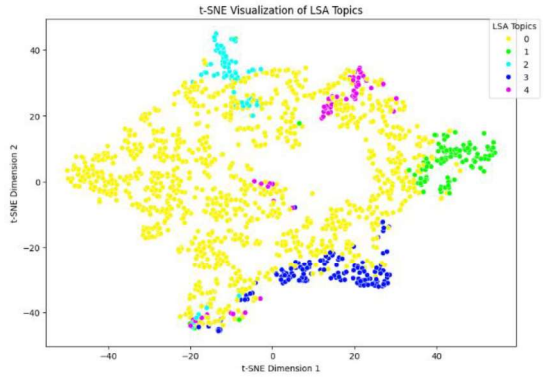


Fig. 10. t-SNE generated by the LSA for BCA Mobile

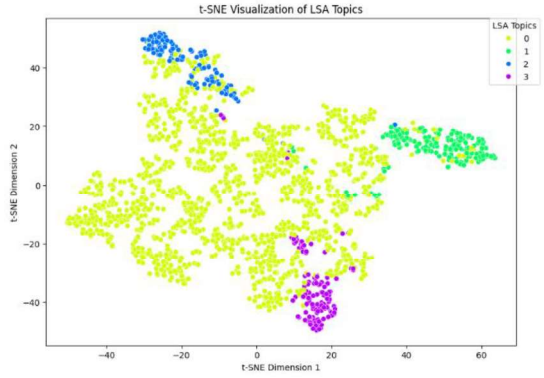


Fig. 11. t-SNE generated by the LSA for BRImo

In Fig. 8, there is no overlapping LDA topic, while in Fig. 9, there are several overlapping LDA topics. The overlapping topics indicate that similar words build different topics. Based on the LDA visualization, the relationship between topics can be identified by grouping overlapping topics. Tables 2 and 3 display the topic analysis of LDA for BCA Mobile and BRImo.

TABLE II. TOPIC ANALYSIS OF LDA FOR BCA MOBILE

Topic Category	Related Topics	Description	Number of reviews
A-1	-	User interaction	568
A-2	-	Ease of login process	405
A-3	-	Features and ease of use of the application	582

TABLE III. TOPIC ANALYSIS OF LDA FOR BRIMO

Topic Category	Related Topics	Description	Number of reviews
B-1	1, 2, 3, 4, 5	Application service fees and speed	827
B-2	6	Transaction security	35
B-3	7	Transaction convenience	66
B-4	8	Transaction balance	157
B-5	9	Customer features	12
B-6	10	BRImo login	400
B-7	11	BRImo application menu improvements	511

Meanwhile, the LSA visualization in Fig. 10 and Fig. 11 show that 5 LSA topics were formed on the BCA mobile application and 4 LSA topics were formed on the BRImo application. Topic 0 on BCA Mobile and BRImo dominates and is widespread, indicating that this topic has a more general scope. While the other topics on BCA Mobile and BRImo are grouped more specifically in certain areas, indicating that this topic is homogeneous. Table 4 and Table 5 show the analysis results for each LSA topic of the BRImo application.

TABLE IV. TOPIC ANALYSIS LSA FOR BCA MOBILE

Topic Category	Description	Number of reviews
C-1	Application usage	1164
C-2	Convenience, speed, and efficiency	116
C-3	Technical constraints	81
C-4	Payment features, notifications, and financial transactions	133
C-5	BCA services	61

TABLE V. TOPIC ANALYSIS LSA FOR BRIMO

Topic Category	Description	Number of reviews
D-1	Use of the BRImo application	1576
D-2	Ease of transactions	181
D-3	Authentication, login, password, and account creation	118
D-4	Management of balances, notifications, fees, and transfers	133

D. Ensemble Clustering: Voting Method

A voting-based ensemble clustering begins by collecting the topic or cluster labels assigned to each

document by LDA and LSA models. For each document, the system gathers the labels produced by these models and counts how often each label appears. The label with the highest frequency, or the majority vote, is then chosen as the final ensemble label for that document. This method helps to stabilize clustering outcomes by combining the perspectives of different models. However, if a tie occurs—meaning two or more labels appear with the same frequency—, the system must apply a tie-breaking strategy. Common approaches include selecting the label from the model with the higher coherence score, which indicates better topic quality. Table 6 shows an example of the implementation of the voting method.

TABLE VI. EXAMPLE OF VOTING METHOD OF BCA MOBILE

Reviews	LDA Topic	LSA Topic	Voting Result
koneksi internet stabil bisa muter loading doang …	2	0	2 (Tie)
servernya berat intip saldo buka tampil rekening ...	0	0	0
moga tambah fitur bca mobile bayar via qris topup flazz	1	3	1 (Tie)
bagus keluh adain via chat gak musti telpon hallo bca	2	2	2

As we know, Table 1 shows that the LDA topic coherence score is higher than the LSA coherence score on BCA Mobile. Therefore, if a tie occurs, the LDA topic is chosen.

After conducting the voting method, the final label is obtained for each document. Table 7 shows the topics on the BCA Mobile and BRImo applications after conducting the voting method on the results of the LDA topic and the LSA topic.

TABLE VII. FINAL TOPIC FOR BCA MOBILE AND BRIMO

Application	Topic	Description
BCA Mobile	A-1	User interaction
	A-2	Ease of login process
	A-3	Features and ease of use of the application
BRImo	B-1	Application service fees and speed
	B-2	Transaction security
	B-3	Transaction convenience
	B-4	Transaction balance
	B-5	Customer features
	B-6	BRImo login
	B-7	BRImo application menu improvements

IV. CONCLUSION

This process provides a clear overview of the negative issues frequently complained about by users of mobile banking applications, as well as the appreciated features. The following topics were successfully identified based on ensemble clustering using LDA and LSA for BCA Mobile: user interaction, ease of login process, and the applications' features and ease of use. Then, for BRImo there are: application service fees and speed, transaction security, transaction convenience, transaction balance, customer features, BRImo login, and BRImo application menu improvements. This study allows application developers to take more appropriate actions to improve the application.

ACKNOWLEDGMENT

This acknowledgment is addressed to the Directorate of Research, Technology, and Community Service (DRTPM) of the Ministry of Education, Culture, Research, and Technology (Kemdikbudristek) of the Republic of Indonesia, which has provided the research team with the opportunity to conduct research in 2024 under the New Lecturer Research scheme (PDP).

REFERENCES

[1] A. Ahdiat, "Transaksi Digital Banking di Indonesia Tumbuh 158% dalam 5 Tahun Terakhir," Katadata Media Network, 2023. Available: https://databoks.katadata.co.id/datapublish/2023/07/05/transaksi-digital-banking-di-indonesia-tumbuh-158-dalam-5-tahun-terakhir. [Accessed: Mar. 15, 2024]

[2] A. Ahdiat, "Penetrasi Internet Indonesia Capai 78% pada 2023, Rekor Tertinggi Baru," 2024. Available: https://databoks.katadata.co.id/datapublish/2024/01/30/penetrasi-internet-indonesia-capai-78-pada-2023-rekor-tertinggi-baru. [Accessed: Mar. 21, 2024]

[3] I. Ulfi, "Tantangan dan Peluang Kebijakan Non-Tunai: Sebuah Studi Literatur," Jurnal Ilmiah Ekonomi Bisnis, vol. 25, no. 1, pp. 55–65, 2020, doi: 10.35760/eb.2020.v25i1.2379

[4] Y. B. Adji, W. A. Muhammad, A. N. L. Akrabi, and N. Noerlina, "Perkembangan Inovasi Fintech di Indonesia," Business Economic, Communication, and Social Sciences Journal (BECOSS), vol. 5, no. 1, pp. 47–58, Jan. 2023, doi: 10.21512/becossjournal.v5i1.8675

[5] E. B. Yusuf, M. I. Fasa, and Suharto, "Inovasi Layanan Perbankan Syariah Berbasis Teknologi sebagai Wujud Penerapan Green Banking," Istithmar : Jurnal Studi Ekonomi Syariah, vol. 7, no. 1, pp. 34–41, Jun. 2022, doi: 10.30762/istithmar.v6i1.33

[6] Dirwan, "Keputusan Nasabah Menggunakan Mobile Banking dari Sisi Kemudahan, Manfaat dan Kenyamanan," SEIKO: Jurnal of Management and Business, vol. 5, no. 1, pp. 323–332, 2022.

[7] A. F. Iriani, "Minat Nasabah dalam Penggunaan Mobile Banking pada Nasabah Bank Syariah Mandiri Kota Palopo," DINAMIS-Journal of Islamic Management and Bussines, vol. 2, no. 2, pp. 99–111, 2019.

[8] M. Ade Kurnia Harahap, Normansyah, S. Nurhayati, I. M. Nawangwulan, and S. P. Anantadjaya, "Pengaruh Penanganan Keluhan Dan Komunikasi Pemasaran Terhadap Loyalitas," COSTING: Journal of Economics, Business and Accounting, vol. 7, no. 2, 2024.

[9] Mutia Annur C. Katadata Media Network. 2022 [cited 2024 Mar 15]. Aplikasi Mobile Banking Terpopuler di Indonesia, Siapa Juaranya? Available from: https://databoks.katadata.co.id/datapublish/2022/06/22/aplikasi-mobile-banking-terpopuler-di-indonesia-siapa-juaranya

[10] Hariansyah FA, Wardani NH, Herlambang AD. Analisis Pengaruh Kualitas Layanan Mobile Banking terhadap Kepuasan dan Loyalitas Nasabah pada Pengguna Layanan BRI Mobile Bank Rakyat Indonesia di Kantor Cabang Cirebon. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer [Internet]. 2019 Apr;3(5):4267–75. Available

from: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5177

[11] Makmuriyah AN, Vanni KM. Analisis Faktor-faktor yang Mempengaruhi Kepuasan Nasabah dalam Menggunakan Layanan Mobile Banking (Studi Kasus Pada Nasabah Bank Syariah Mandiri di Kota Semarang). Eduka: Jurnal Pendidikan, Hukum, dan Bisnis . 2020;5(1):37–44.

[12] Parera NO, Susanti E. Loyalitas Nasabah dari Kemudahan Penggunaan Mobile Banking. International Journal of Digital Entrepreneurship and Business. 2021;2(1):39–48.

[13] Ivana Ruslim K, Pandu Adikara P, Indriati. Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Support Vector Machine dan Lexicon Based Features. 2019 [cited 2024 Mar 15];3(7):6694–702. Available from: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5792/2748

[14] Nadira A, Yudi Setiawan N, Purnomo W. Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Naive Bayes dengan Kamus InSet. INDEXIA : Informatic and Computational Intelligent Journal [Internet]. 2023 [cited 2024 Mar 15];5(1):35–47. Available from: https://journal.umg.ac.id/index.php/indexia/article/view/5138/3113

[15] Tondang BA, Muhammad Rizqan Fadhil, Muhammad Nugraha Perdana, Akhmad Fauzi, Ugra Syahda Janitra. Analisis pemodelan topik ulasan aplikasi BNI, BCA, dan BRI menggunakan latent dirichlet allocation. INFOTECH : Jurnal Informatika & Teknologi. 2023 Jun 17;4(1):114–27.

[16] Fiqri M, Setya Perdana R. Klasifikasi Data Twitter pada Masa Transisi Pandemi menuju Endemi menggunakan Latent Semantic Analysis (LSA) [Internet]. Vol. 7. 2023. Available from: http://j-ptiik.ub.ac.id

[17] Malihatin S U, Findawati Y, Indahyanti U. TOPIC MODELING IN COVID-19 VACCINATION REFUSAL CASES USING LATENT DIRICHLET ALLOCATION AND LATENT SEMANTIC ANALYSIS. Jurnal Teknik Informatika (Jutif). 2023 Oct 3;4(5):1063–74.

[18] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.

[19] J. A. Costales, E. M. Lorico, and C. M. de Los Santos, "A Comparative Sentiment Analysis about HIV and AIDS on Twitter Tweets Using Supervised Machine Learning Approach," in 2023 5th International Conference on Computer Communication and the Internet (ICCCI), 2023, pp. 27–32. doi: 10.1109/ICCCI59363.2023.10210162.

[20] scikit learn developer, "Confusion Matrix." Accessed: Mar. 29, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html

[21] Evidently AI, "How to interpret a confusion matrix for a machine learning model." Accessed: Mar. 29, 2025. [Online]. Available: https://www.evidentlyai.com/classification-metrics/confusion-matrix

[22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[23] T. Hofmann, H. Schütze, and CD. Manning, *Introduction to Information Retrieval (Updated Edition)*. Cambridge University Press, 2019. Accessed: Mar. 29, 2025. [Online]. Available: https://nlp.stanford.edu/IR-book/

[24] A. Göker and G. Demir, "Topic Ensemble: Combining Topic Models Using Clustering Ensembles," *Inf Process Manag*, vol. 58, no. 6, 2021.

ISSN 2355-0082

Web Design and Development of 'Peduli PMI' System for Managing Complaints and Protection of Indonesian Migrant Workers

Murie Dwiyaniti[1], Rika Novita Wardhani[2], Nuha Nadhiroh[3], Asri Wulandari [4], Viving Frendiana[5], Farhan Yuswa Biyanto[6]
[1,3]Program Study of Industrial Electrical Automation Engineering, Department of Electrical Engineering, Jakarta State Polytechnic, Jakarta, Indonesia
[2]Program Study of Industrial Instrumentation and Control, Department of Electrical Engineering, Jakarta State Polytechnic, Jakarta, Indonesia
[4,5,6]Program Study of Broadband Multimedia, Department of Electrical Engineering, Jakarta State Polytechnic, Jakarta, Indonesia
[1]murie.dwiyaniti@elektro.pnj.ac.id, [2]rika.novitawardhani@elektro.pnj.ac.id, [3]nuha.nadhiroh@elektro.pnj.ac.id , [4]asri.wulandari@elektro.pnj.ac.id, [5]viving.frendiana@elektro.pnj.ac.id, [6]farhan.yuswabiyanto.te20@mhsw.pnj.ac.id

Accepted 2 November 2024
Approved 18 June 2025

***Abstract*— Indonesian Migrant Workers (PMI) face high occupational risks, making it essential to provide them with appropriate attention to prevent and reduce unwanted incidents. A computerized system in the form of a web application can assist administrators in managing PMI complaints and protection more effectively. Through this system, administrators can handle PMI data, manage content such as news and announcements, monitor statistics related to users, complaints, and protection cases, and manage complaint and protection data submitted by PMI to ensure their safety. This study aims to design and develop a Complaint and Protection Management System Website for PMI in Malaysia. The system includes four main features: secure admin access with login, user registration, password recovery, and administrative capabilities. Administrators can manage user data, handle complaints, update report status, respond to complaints, view other administrators' lists, and modify personal data and password settings. Additionally, they can manage front-end content for the mobile application, including news and announcements. This system is designed to be efficient and responsive to the needs of PMI. The study conducted four tests—Functionality Testing, API Testing, Performance Testing, and Security Testing. Performance testing using GTMetrix showed optimal results, with scores reaching 96% in several test server locations. "Peduli PMI" has met the ISO/IEC 25010 standards in terms of functionality, performance, security, and integration, making it a ready and reliable system.**

***Index Terms*— Indonesian Migrant Worker (PMI); Admin; Website Management System, Web Design, Laravel.**

I. INTRODUCTION

Based on Law No. 18 of 2017 on the Protection of Indonesian Migrant Workers, Indonesian Migrant Workers (PMI) are individuals employed overseas under fixed-term agreements [1], [2]. Despite this legal recognition, PMIs face high occupational risks such as mistreatment, neglect, and unfair treatment. Issues like human rights violations, inadequate protection, and high unemployment rates or arbitrary dismissals create significant challenges for these workers [3], [4]. For example, the tragic abuse case of Adelina, a PMI from East Nusa Tenggara who suffered violence in Malaysia in 2018, reveals the vulnerabilities migrant workers face abroad and highlights the weaknesses in the protection system available to them [5]. These vulnerabilities are further exacerbated by various challenges in the complaint system, including the manual creation of complaint forms, inefficient data recapitulation, inconsistent data formats, data duplication, outdated data, and obstacles in accessing information to update ongoing case developments [4].

To address these issues, a structured and reliable system for managing PMI complaints and protection is essential. This solution is realized through the development of a web-based system designed to support the comprehensive administration of PMI data. This web application aims to simplify administrative tasks by enabling administrators to monitor complaint statistics, manage content on partnerships, news, and other important information. Additionally, the system facilitates real-time management of complaint and protection status reports, allowing administrators to respond promptly to incoming reports from PMIs working abroad.

Through the use of this complaint and protection management system—developed with the Laravel framework and MySQL database—this platform is expected to be an effective tool for enhancing PMI security and comfort. Following ISO/IEC 25010 testing standards for functionality, performance, and security, alongside API testing with ThunderClient, this system

aims to not only be reliable but also provide PMIs with a sense of security. The hope is that this system will serve as a dependable resource for PMIs, easing concerns over work environments far from home and allowing them to lead more secure and peaceful lives in their host countries.

II. LITERATURE REVIEW

A. PMI (Indonesian Migrant Workers)

Law No. 18/2017 on the Protection of Indonesian Migrant Workers defines Indonesian workers, now referred to as Indonesian Migrant Workers (PMI), as individuals who migrate abroad to fulfill work contracts for a specified period [1], [2].

Full attention should be given to Indonesian Migrant Workers to reduce and prevent unwanted incidents, as they face a high level of occupational risk, including mistreatment or violence by employers. Violence is a deliberate act intended to harm someone, affecting people indiscriminately, although women are particularly vulnerable. In addition to legal protection, a system is needed to address human rights violations against migrant workers, one of which is a monitoring system. In this regard, the government must prioritize addressing these issues [3], [4].

In Malaysia, around 2.5 million low-wage migrant workers are at risk of being laid off without pay, and sadly, 400,000 workers have been evicted from rented homes due to inability to pay rent. Concerns are growing about the economic well-being of families back home, as these workers have been unable to send money for several months. The case of abuse involving Adelina, an Indonesian migrant worker from East Nusa Tenggara, in Malaysia on February 10, 2018, became a tragic event that highlighted the vulnerability and insecurity of migrant workers abroad [5].

B. Management System

A Management System is a platform used to oversee various aspects of an organization, including risk management, employee productivity, and performance evaluation. This management system website enables users to manage processes, reduce risks, enhance employee productivity, and conduct evaluations. The system can store, retrieve, modify, process, and delete information or data received from other information systems or equipment [6].

C. Website

The World Wide Web (WWW), commonly known as a website, is a key facility on the vast internet, serving as an information medium and a promotional tool. According to Abdullah, a website is defined as a collection of pages containing digital information in the form of text, images, animations, sound, video, or a combination thereof, accessible via an internet connection to viewers worldwide. Web pages are created using a standard language, HTML, which is translated by a web browser to display information in a readable format for users [7].

D. Complaints and Protection

According to the Big Indonesian Dictionary (KBBI), a complaint is an expression of dissatisfaction regarding matters that may seem minor but still require attention, while public complaints reflect the community's aspiration to actively participate in monitoring an agency's performance. Presidential Regulation Number 76 of 2013 defines a complaint as a submission by the complainant to the public service complaint manager concerning services that do not meet established standards, neglect of duties, and/or violations by public service providers. Meanwhile, KBBI defines protection as the act of providing safety, and Article 1, Point 25 of the Criminal Procedure Code states that a complaint is a notification accompanied by a request from an interested party to an authorized official to take legal action against a person who has committed a criminal offense causing harm to the complainant [8].

E. Laravel

Laravel is an open-source, free PHP-based web framework created by Taylor Otwell, designed for developing web applications that use the MVC pattern [9]. The MVC structure in Laravel is slightly different from the traditional MVC pattern. In Laravel, routing bridges requests between the user and the controller, so the controller does not receive requests directly. Additionally, Laravel stands out for its expressive syntax, enabling developers to write code more easily and efficiently. The Blade templating engine, used to manage views, provides a clean and powerful way to generate dynamic views. The framework also includes various integrated features, such as an intuitive routing system, user authentication management, session handling, security, and ready-to-use cache management [10].

F. API

An Application Programming Interface, commonly known as an API, is a component of a software system that consists of a collection of functions, commands, and protocols that enable computer systems to interact with one another [11]. The purpose of using an API is to accelerate the process of building interactions between software by utilizing pre-existing or separately developed functions [12].

G. Functionality Testing

In the aspect of functionality testing, a research instrument in the form of test cases is used with a Guttman scale. The Guttman scale is employed to obtain definitive answers to the problems being addressed [13]. This type of measurement scale provides clear responses, namely “Yes” or “No,” where “Yes” is assigned a value of 1 and “No” a value of 0 for each item [13].

TABLE I. GUTTMAN SCALE CONVERSION [13]

Result	Score
Yes	1
No	0

The calculation is performed using the success percentage and the feasibility percentage table as follows:

$$percentage\ of\ success = \frac{1}{r} x100\% \quad (1)$$

Where:
i = number of functional requirements successfully implemented
r = total number of functional requirements

H. Perfomance Testing

Performance testing assesses the relative performance level of a system's resources under specific conditions [14] . It measures the time required to access various functionalities, the capacity needed during system access, and the resources utilized by the system. This stage of testing is conducted with the help of web tools. Among the performance tools available for search engines is GTMetrix. GTMetrix is used to determine performance scores, which include metrics such as page speed, fully loaded time, total page size, and the number of requests on the analyzed system. These scores help evaluate the system's performance efficiency [15]. For interpreting the GTMetrix score and the metrics LCP (Largest Contentful Paint), TBT (Total Blocking Time), and CLS (Cumulative Layout Shift), please refer to Tables 2 and 3.

TABLE II. GTMETRIX SCORE INTERPRETATION [15]

Score	Color Code	Adjective Rating
0-49	*Red*	*Poor*
0-89	*Orange*	*Needs Improvement*
90-100	*Green*	*Good*

TABLE III. INTERPRETATION OF LCP, TBT, AND CLS IN GTMETRIX [15]

Adjective Rating	CLS	TBT	LCP
Good	<0.1	<150	<1200
OK	0.1 – 0.15	150 – 224	1200 – 1666
Longer	0.15 –0.25	224 – 350	1666 – 2400
Much Longer	>0.25	>350	>2400

I. Security Testing

Security refers to the ability of a system to protect information and data from access by unauthorized parties, ensuring control over data access based on user permission levels [16]. Website testing concerning security utilizes Sucuri Online Web Vulnerability Scanner software. To evaluate the security of the "Peduli PMI" Management System Website, which serves as a platform for complaints and protection for Indonesian Migrant Workers in Malaysia, Sucuri software is employed.

J. API Testing using ThunderClient

API testing using ThunderClient aims to ensure that all requests and responses between the website (acting as a server) and the "Peduli PMI" mobile application (acting as a client) function correctly and efficiently. This test will evaluate the API integrated with the "Peduli PMI" application to observe how data is sent, received, and processed. ThunderClient provides the necessary tools to conduct thorough API testing.

III. METHODOLOGY

This study aims to design and develop a Complaint and Protection Management System Website for Indonesian Migrant Workers (PMI) in Malaysia. The primary objectives are to enhance the management of PMI complaints and protection cases, ensure secure access for administrators, and provide effective data handling.

A. System Design and Development

The development process involves several stages:

- Requirements Analysis: Identify the specific needs of PMI and administrators regarding complaint management and protection services.
- System Design: Create a detailed design of the web application, including user interface (UI) mockups and database schema.
- Development: Use appropriate programming languages and frameworks to build the system, ensuring it includes the following key features:
  - Secure admin access with login functionality.
  - User registration and password recovery options.
  - Administrative capabilities for managing user data, complaints, and report statuses.
  - Content management for news and announcements relevant to PMI.

B. Testing Methods

Four types of testing will be conducted to ensure the system meets required standards:

- Functionality Testing: Verify that all features work as intended, including user registration, complaint submission, and administrative functions.
- Performance Testing: Assess the system's response times and load handling capabilities to ensure it can support multiple users simultaneously.
- API Testing: Use ThunderClient to ensure correct and efficient communication between the web application and the "Peduli PMI" mobile application.
- Security Testing: Implement Sucuri Online Web Vulnerability Scanner to identify potential security risks and ensure data protection for PMI.

C. *Use Case Diagram*

To visualize the interaction between users and the system, a use case diagram is needed to describe the functionality. The following is a use case diagram for the "Peduli PMI" Management System Website, as shown in Figure 1.

Based on Figure 1, users can perform the following activities using the "Peduli PMI" Management System Website:

1. Reset their password if they forget it.
2. Register through the registration form.
3. Log in to access the website.
4. Manage announcements and news for the "Peduli PMI" Mobile App.
5. Manage all data and activities related to Registrant Data.
6. Manage all data and activities related to User Data.
7. Manage all data and activities related to Complaint Data reported by Indonesian Migrant Workers, either by themselves or others.
8. Manage all data and activities related to Protection Data reported by Indonesian Migrant Workers, both self-reported and reported by others.
9. Change the admin's personal data, including email, name, NIK, mobile number, and password as needed.

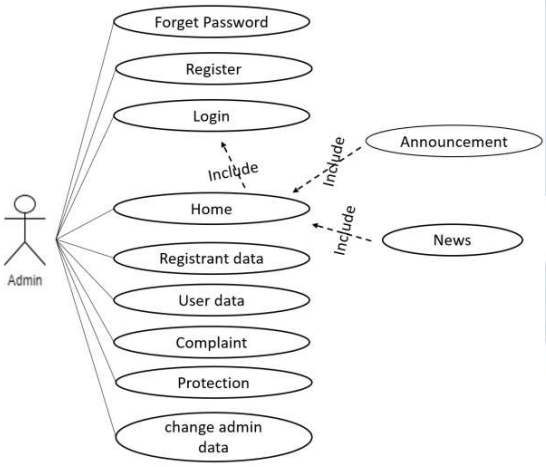


Fig. 1. Use case diagram

D. *Architecture Diagram*

The following is PMI care system architecture diagram, on the "Peduli PMI" Management System Website as follows in Figure 2.

The "Peduli PMI" Management System website serves as a server that interacts with the application through the Hypertext Transfer Protocol (HTTP) to facilitate communication between the client and the server. This website is responsible for managing all existing REST APIs, which include News, Announcements, Registration Forms, Complaint Data, and Protection in the "Peduli PMI" Mobile application. The REST API data will be processed and forwarded to the "Peduli PMI" Mobile application using a MySQL database. This database stores data that is input and received on the user side. On the client side, the application will send and receive the REST APIs that have been created on the "Peduli PMI" Management System Website. The application receives news and announcement data through the existing API. The process of sending the REST API occurs when applying for protection or submitting data, which will be stored in the database via the existing API POST method.

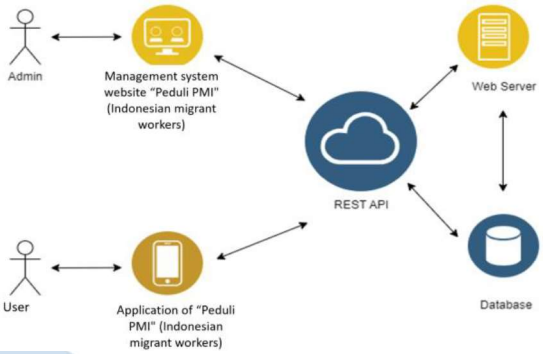


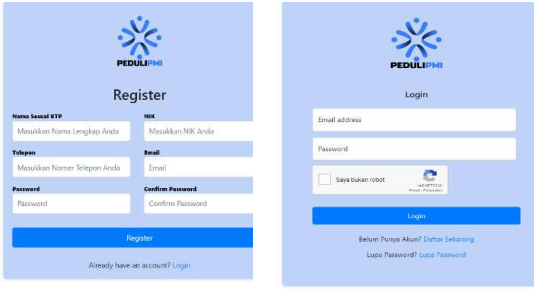
Fig. 2. PMI care system architecture diagram

The admin workflow in the management system begins with selecting a menu based on needs, such as "Complaint Data" to add, view, and manage complaint statuses, "News" and "Announcements" to manage related content, as well as "Registrant Data" and "User Data" to manage registrant and user information. Admins can also view admin and writer data, update their personal profiles, and change passwords. Once all tasks are completed, admins can log out to end the session, ensuring efficiency and order in the administrative process.

IV. Results and Discussion

A. *Website Realization*

The "Peduli PMI" Management System website was realized using the PHP programming language using the Laravel framework, and using bootsrap 5 styling. The following are some views of the website.



(A) (B)

Fig 3. (A) Register Page and (B) Login Page

Figures 3 depict the Register and Login pages of the "Peduli PMI" Management System Website; Figure 3a depict the Register page allows admins without an account to create one, while Figure 3b the Login page enables admins to access the system by entering their email, password, and Google Captcha confirmation to prevent database attacks.

Fig 4. Forgot Password Page

In Figure 4 is the Forgot Password page, users can change their password by sending an email address on the forgot password page. Then a link will be sent to reset the password via email. Creating a forgotpassword() function, this function functions for the process when resetting the password to the website, the process of resetting the password by entering the admin email in the tb_admin table data. The system will send an email to reset the password.

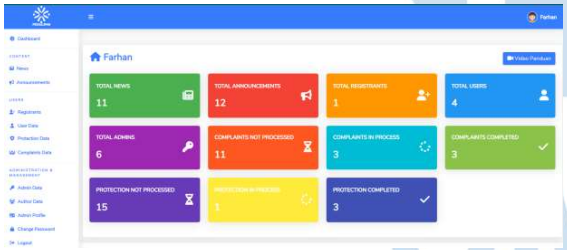


Fig 5. Dashboard Page

In Figure 5 is the dashboard page, it will display all the features contained on the website, and display the total count of data in that feature.

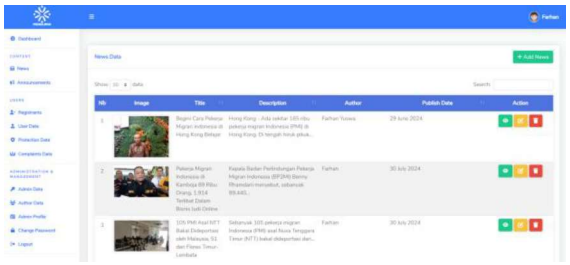


Fig 6. News And Annoucement Page

In Figure 6 is the News and Announcement page, these two pages have the same features, there is a CRUD feature for news and announcements that functions to manage all news and announcement content that will be displayed on the mobile application.

Fig 7. Registrant Data Page

In Figure 7 is Registrant Data Page, for this page serves to manage all data on Prospective Indonesian Migrant Workers who have registered on the Mobile Applicatio. Admin can view the complete data of Prospective Indonesian Migrant Workers (CPMI), can approve if the complete data matches the requirements or reject if the data does not match the requirements.

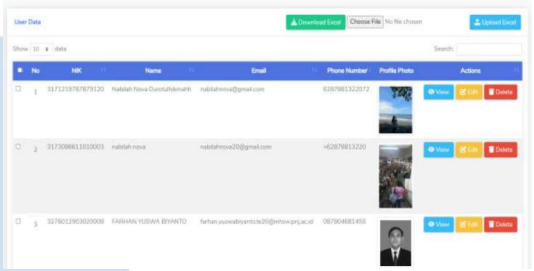


Fig 8. User Data Page

In Figure 8 is the User Data page, this page serves to manage all active user data that has passed the requirements. In this feature the admin can view the complete data of Indonesian Migrant Workers (PMI), can edit user data if there is inappropriate data, there is also a delete user feature, namely if the user is no longer a PMI, and in this feature there is also an export of all user data and import all user data into excel format (xls).

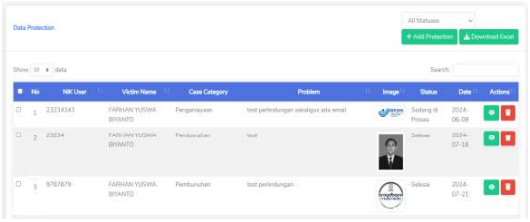


Fig 9. Protection Data Page

Figure 9 shows the Protection Data page, which functions to manage all protection data reported by PMI. This page allows for reporting of personal cases or cases on behalf of others if there are any law violations or activities posing a threat to Indonesian Migrant Workers. It contains a comprehensive collection of cases and issues that endanger their lives. In this feature, detailed protection data is displayed: if you report a case for yourself, it will show the victim's name, NIK, phone number, email, incident location,

date of incident, issue, photo evidence, and assistance request. If you report a case on behalf of someone else, it will display the reporter's and victim's details, including names, NIK, phone numbers, emails, incident location, date, issue, photo and audio evidence, and assistance request. Additionally, there is an option to export the data to PDF, as well as features for adding responses and updating the protection status. If the admin changes the status, the email associated with the report will receive an update. Lastly, there is an option to add new Protection cases, allowing PMI to directly contact the admin if they cannot make a report through the mobile application.

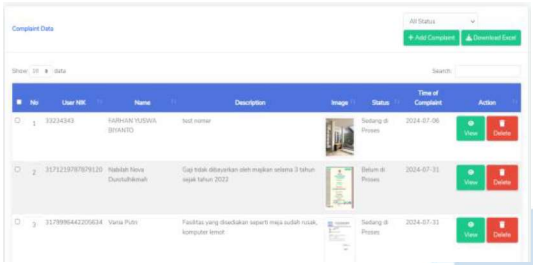


Fig 10. Complaint Data Page

Figure 10 displays the Complaint Data page, which manages all complaint data reported by PMI. This page allows users to report complaints for themselves or on behalf of others regarding issues encountered at work or concerning infrastructure for Indonesian Migrant Workers, such as unpaid wages, excessive work hours, inadequate workplace facilities, or garbage buildup. This feature presents detailed complaint data: if you report a complaint for yourself, it will show the victim's name, NIK, phone number, email, complaint location, date, description, and both photo and audio evidence. If reporting on behalf of someone else, it will display both the complainant's and victim's details, including names, NIK, phone numbers, emails, complaint location, date, description, and evidence. Additionally, the feature includes options to export data to PDF, add responses, and update the complaint status. If the admin updates the status, the email associated with the complaint will receive a notification. There is also an option to add a complaint, allowing PMI to contact the admin directly if they are unable to submit a report through the mobile application.

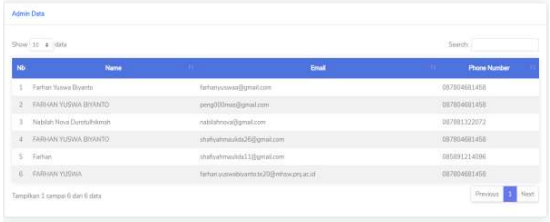


Fig 11. Admin Data Page

In Figure 11 is the Admin Data Page, for this page serves to display all admins who have registered.

Fig 12. Author Data Page

In Figure 12 is the Author Data page, for the author data page functions to add authors and emails that will be displayed on news and announcements.

Fig 13. Admin Profile Page

In Figure 13 is the Admin Profile Page, on this page it functions to manage the admin profile that is currently logged in. In this feature, the admin can update his/her own data such as name, NIK, mobile number and email.

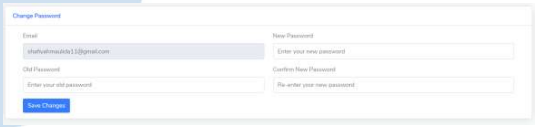


Fig 14. Change Password Page

In Figure 14 is the Change Password Page, on this page it functions to reset the admin password, by entering the old password, new password, and confirming the new password. To realize the process of changing the password page using the resetPassword() function on the AdminController. Then create a view for the change Password page in the peduli-pmi/resources/views/backend directory with the file name change-password.blade.php. Furthermore, creating a route to display and process the controller so that all functions can be called, is in the peduli-pmi/routes/web.php directory.

B. Website Testing

Website testing of the "Peduli PMI" Management System, which serves as a platform for complaints and protection for Indonesian Migrant Workers (PMI) in Malaysia, follows the ISO/IEC 25010 testing standard. The types of testing conducted include functional testing, performance testing, security testing, and API testing using Thunder Client.

B.1 Functionality Testing

The purpose of this functional testing is to ensure that all features of the "Peduli PMI" Management System Website, which serves as a platform for complaints and protection for Indonesian Migrant Workers in Malaysia, operate correctly and align with the intended design and implementation. Table 4 shows the test results from testing the “Peduli PMI” Management System Website as a platform for complaints and protection for Indonesian Migrant Workers in Malaysia.

The testing results in Tabel 4 indicate that the "Peduli PMI" Management System Website operates effectively in all scenarios tested. Each feature functions as intended, with the system properly handling both successful operations and errors. The successful registration and verification processes, as well as the functionality for adding and managing news data, reflect a well-designed user experience. The system also demonstrates its ability to guide users in case of input errors, ensuring reliability and usability.

Overall, these results affirm that the website meets the expected functionality criteria and can serve its intended purpose of providing a platform for complaints and protection for Indonesian Migrant Workers in Malaysia. Further testing could focus on more complex scenarios or edge cases to ensure robustness under varied conditions

TABLE IV. RESULT OF FUNCTIONALITY TESTING

Case	Testing Scenario	Expected Result	Test Result	Conclusion
Successfully registered an admin account	Admin completes registration by filling in the form fields according to the required format, including Name as per ID, NIK, Phone Number, Email, Password, and Confirm Password	Admin successfully registers an account	Success	Valid
Successfully received an email for account verification	Admin receives an email containing the account verification process to enable login	Admin successfully receives the verification email	Success	Valid
Failed to register an admin account	Admin attempts registration without following the required form format or leaves fields incomplete	Admin fails to register an account	Success	Valid
Click "Add News"	Admin clicks "Add News" and is directed to the Add News page	Add News page opens successfully	Success	Valid
Successfully filtered news data and sorted news data	Admin filters news data according to preferences of 10, 25, 50, and 100 entries, and can sort announcement data as desired	Admin successfully filters and sorts news data as intended	Success	Valid

B.2 Performance Testing

Performance testing is done to test the time required when accessing a condition on the system, the capacity required when the system is accessed, and the resources used by the system. At this stage, testing is done using the help of a web-tool, namely GTMetrix. From the performance testing conducted using GTmetrix software, the test results are shown in Table 5.

TABLE V. RESULT OF PERFORMANCE TESTING

Testing Server Location	Result				
	Performance (%)	Structure (%)	LCP (s)	TBT (ms)	CCS
Sao Paulo, Brazil	74	97	2.8	0	0,04
Hongkong, China	96	97	1.1	0	0
Sydney, Australia	96	97	1.2	76	0,04
Vancouver, Canada	88	97	1,8	0	0,04

The server performance test results based on location show variations in performance across different regions. In São Paulo, Brazil, the performance scored 74% with a structure rating of 97%, a Largest Contentful Paint (LCP) of 2.8 seconds, a Total Blocking Time (TBT) of 0 ms, and a Cumulative Layout Shift (CLS) of 0.04. In Hong Kong, China, the performance achieved its highest score of 96% with a structure rating of 97%, an LCP of 1.1 seconds, TBT of 0 ms, and CLS of 0. Similarly, Sydney, Australia, also recorded a performance score of 96% with a structure rating of 97%, an LCP of 1.2 seconds, TBT of 76 ms, and CLS of 0.04. Finally, in Vancouver, Canada, the performance reached 88% with a structure rating of 97%, an LCP of 1.8 seconds, TBT of 0 ms, and CLS of 0.04. These results indicate consistent structural integrity across locations, though performance and LCP vary slightly depending on the testing region.

B.3 Security Testing

To perform website security testing using Sucuri SiteCheck, visit the Sucuri SiteCheck homepage at https://sitecheck.sucuri.net/. Enter the complete URL of the website you wish to test in the provided text box, including "http://" or "https://". Click "Scan Website" to initiate the scanning process, which may take a few seconds to minutes depending on the site’s size and complexity. Once the scan is complete, review the results, which will indicate any detected malware, the site's blacklist status, outdated software, and security recommendations. If issues are found, follow Sucuri's guidelines for remediation or consult a web security professional for further assistance. Finally, you can save or print the scan results for future reference or to

share with your technical team. Figure 15 shows the result of security testing.

Fig 15. Result of Security Testing

Based on the test results from the Sucuri website (Figure 15), no malware was found, and this website is not blacklisted by security services such as Google Safe Browsing. This means there is no indication that the website is dangerous to users or likely to spread malware. Additionally, there is no evidence of SEO spam, such as the use of hidden keywords or links typically used for browser manipulation.

B.4 API Testing using Thunder Client

To conduct API testing with ThunderClient, start by downloading the ThunderClient extension for Visual Studio Code. Open the extension and click on "New Request." In the "Define API Call" section, enter the API call name, select the desired method type, and input the API URL. Next, fill in the required parameters in the "Body" tab and create any necessary variables in the "Variables" tab. Once everything is configured, click "Send" to view the API response. A status of 401 indicates a failure, while a status of 200 signifies success, with the JSON response depending on the API input.

TABLE VI. RESULT OF API TESTING USING THUNDER CLIENT

API Name	API URL	Method	Result
PengaduanStore	https://foruminiujian.my.id/api/v9/396d6585-16ae-4d04-9549-c499e52b75ea/pengaduan/store	POST	Success
Perlindungan Store	https://foruminiujian.my.id/api/v9/396d6585-16ae-4d04-9549-c499e52b75ea/perlindungan/store	POST	Success
GetAllDataNews	https://foruminiujian.my.id/api/v3/98765432-1abc-0fed-cba9-87643210fed/news	GET	Success
GetAllDataPerlindungn	https://foruminiujian.my.id/api/v9/396d6585-16ae-4d04-9549-c499e52b75ea/perlindungan	GET	Success

The successful results for all APIs suggest robust functionality and reliable integration within the "Peduli PMI" Management System, ensuring that users can effectively access and manage complaint and protection data.

V. CONCLUSION

This article evaluates the development of a web-based system, "Peduli PMI," aimed at enhancing protection for Indonesian Migrant Workers (PMI) in Malaysia. With key features such as secure administrator access, user registration, password recovery, and content management, the system enables PMIs to report complaints and receive assistance more safely and quickly. Functionality testing shows that each feature operates as expected, while security testing confirms that the system is free from malware and is not blacklisted. These results support the system's success in providing a safe and efficient platform for PMI protection. Quantitatively based on table 5, performance testing using GTMetrix demonstrated optimal results, with scores reaching up to 96% in testing server locations in Hongkong and Sydney, although there was a slight drop in performance in certain locations. API testing also confirmed that all APIs for managing complaint and protection data functioned smoothly without errors. Based on these results, "Peduli PMI" meets ISO/IEC 25010 standards in terms of functionality, performance, security, and integration, making it a ready and reliable system to support the safety and comfort of PMIs in their host country.

ACKNOWLEDGMENT

The authors thank P3M Jakarta State Polytechnic for funding this Community Service program, supported by the 2024 DIPA budget per Decree No. 966/PL3/PT.00.00/2024 (April 24, 2024).

REFERENCES

[1] L. P. Hasugian, R. Sidik, Y. H. Putra, Y. Y. Kerlooza, and D. A. Wahab, "Analisis Kebutuhan Sistem Informasi Pemantauan Pekerja Migran Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2, pp. 216–226, 2019.

[2] H. Khalid and A. Savirah, "Legal protection of Indonesian migrant workers," *Golden Ratio of Law and Social Policy Review*, vol. 1, no. 2, pp. 61–69, 2022.

[3] M. Aswindo, M. Hanita, and A. J. SIMON, "Kerentanan dan Ketahanan Pekerja Migran Indonesia di Malaysia pada Masa Pandemi Covid-19," *Jurnal Lemhannas RI*, vol. 9, no. 1, pp. 1–10, 2021.

[4] J. F. Haumahu, C. H. J. De Fretes, and T. R. Simanjuntak, "Diplomatic Protection Efforts of the Consulate General of the Republic of Indonesia in Johor Bahru for Indonesian Migrant Workers as the Domestic Helpers (PLRT) in 2018-2019," *ARRUS Journal of Social Sciences and Humanities*, vol. 3, no. 3, pp. 335–347, 2023.

[5] G. D. T. Wahyudi, D. G. S. Mangku, and N. P. R. Yuliartini, "Perlindungan Hukum Tenaga Kerja Indonesia Ditinjau Dari Perspektif Hukum Internasional (Studi Kasus Penganiayaan Adelina TKW Asal NTT Di Malaysia)," *Jurnal Komunitas Yustisia*, vol. 2, no. 1, pp. 55–65, 2019.

[6] G. V. Putri, "Konsep Dasar Sistem Informasi Manajemen Dan Implementasi Sistem Informasi Manajemen Di Sekolah," 2019.

[7] T. Sulistiati, F. Yuliansyah, M. Romzi, and R. Aryani, "Membangun website toko online pempek nthree menggunakan PHP dan MYSQL," *JTIM: Jurnal Teknik Informatika Mahakarya*, vol. 3, no. 1, pp. 35–44, 2020.

[8] A. I. Amilia and A. Y. S. Rahayu, "Pusat Pelayanan Informasi dan Pengaduan (Pindu) Kabupaten Pinrang Dalam Perspektif Best-Practice Manajemen Pengaduan," *Kolaborasi: Jurnal Administrasi Publik*, vol. 6, no. 3, pp. 330–350, 2020.

[9] O. W. Purbo, "A systematic analysis: Website development using Codeigniter and Laravel framework," *Enrichment: Journal of Management*, vol. 12, no. 1, pp. 1008–1014, 2021.

[10] N. Yadav, D. S. Rajpoot, and S. K. Dhakad, "LARAVEL: a PHP framework for e-commerce website," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, IEEE, 2019, pp. 503–508.

[11] N. Kiesler and D. Schiffner, "What is a Good API? A Survey on the Use and Design of Application Programming Interfaces," in *International Conference on Internet of Everything*, Springer, 2023, pp. 45–55.

[12] M. Boyd, L. Vaccari, M. Posada, and D. Gattwinkel, "An Application Programming Interface (API) framework for digital government," *European Commission, Joint Research Centre*, 2020.

[13] H. Abdi, "Guttman scaling," *Encyclopedia of research design*, vol. 2, pp. 1–5, 2010.

[14] R. D. R. Dako and W. Ridwan, "Pengujian karakteristik Functional Suitability dan Performance Efficiency tesadaptif. net," *Jambura Journal of Electrical and Electronics Engineering*, vol. 3, no. 2, pp. 66–71, 2021.

[15] G. Tyas, D. Purnamasari, and A. Suroso, "Analisis Kualitas Aplikasi E-Exam Menggunakan Standar ISO 25010," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 2, pp. 126–132, 2021.

[16] K. Deniswara, E. M. Gunawan, A. N. Mulyawan, and Y. Lisanti, "Exploration of software implementation on cloud accounting and security system towards accounting practices case study from a private company in Indonesia," in *2021 International Conference on Information Management and Technology (ICIMTech)*, IEEE, 2021, pp. 706–711.

ISSN 2355-0082

Analysis of Student Performance Differences in Computer Network Courses with Learning Modes in Multimedia Nusantara University

Livia Junike[1], Karen Yapranata[2], Fransiscus Ati Halim[3]
[1, 2, 3] Department of Informatics, Multimedia Nusantara University, Tangerang, Indonesia
[1]livia.junike@student.umn.ac.id, [2]karen.yapranata@student.umn.ac.id, [3]fransiscus.ati@lecturer.umn.ac.id

Accepted 24 December 2024
Approved 26 June 2025

***Abstract*— Global academic environments have been significantly impacted by the change in educational delivery techniques brought about by the COVID-19 pandemic. This study examines the variations in student performance in Multimedia Nusantara University's Computer Network course across on-site, hybrid, and online learning modes. 526 students data (2019–2022) were evaluated using a variety of statistical techniques, such as the Kruskal-Wallis test, tests for homoscedasticity and normality, pairwise comparison, and post-hoc Dunn's analysis. By taking into consideration the various circumstances and difficulties presented by each learning mode, these techniques guaranteed a thorough assessment of performance variances. The findings indicate that online learning yielded the highest average scores (80.1), demonstrating better performance consistency compared to on-site (75.8) and hybrid modes (72.2), the latter of which showed the widest score dispersion. Statistical evidence revealed significant performance differences between online and other learning modes, whereas no notable differences were observed between hybrid and on-site modalities. These results highlight the effectiveness of online learning in delivering technical courses like Computer Networks, particularly when supported by reliable infrastructure and engaging content design. Nevertheless, improvements in hybrid learning are crucial to reduce performance variability and maximize its potential as a balanced approach to education. This research advocates for future research to explore additional influencing factors, such as teaching strategies, learner engagement, and the role of emotional aspects, in optimizing educational outcomes across various delivery methods.**

***Index Terms*— academic performance; computer network course; online; hybrid; on-site**

I. INTRODUCTION

The introduction of technology has had a profound impact on education around the world, particularly during the COVID-19 epidemic, which forced a switch from offline to online schooling. Since research shows that online and offline learning modes differ in content understanding and academic outcomes, this shift had an impact on academic performance [4][19]. In contrast to regulated offline learning, Zimmerman emphasizes the importance of self-regulation in online learning, where a lack of it can impair performance. But online education also gives students the flexibility to use technology and manage their time. In assessing the efficacy of online learning [20], stress the significance of technological infrastructure, lecturer proficiency, and course quality [13][2].

Computer networks play a pivotal role in enabling online learning. A computer network is a system of interconnected devices facilitating data sharing. Stable internet access allows seamless participation in online learning, while technical issues like unstable connections hinder performance [16][7]. Research suggests that computer-based technology enhances student engagement and academic outcomes [14]. Bernard et al. argue that blended learning and technology use in education significantly boost academic performance [3], aligning with Mayer's multimedia learning theory, which posits that multimedia resources improve comprehension and retention [10].

The COVID-19 pandemic expedited the integration of computer networks in education, ensuring learning continuity despite restrictions [5]. This study examines differences in academic performance in online, hybrid, and on-site learning modes among students from three majors Computer Engineering, Informatics Engineering, and Information Systems at Multimedia Nusantara University. Using a statistical test, it seeks to determine the impact of these modes on academic outcomes during and after the pandemic.

Several previous research have addressed topics similar to this research. For instance, Akpen et al. [1] conducted a systematic review analysing the effects of online learning on student performance and engagement. Their study revealed that online learning offers flexibility and accessibility, which can enhance academic performance. However, it also highlighted challenges such as reduced engagement, feelings of isolation, and diminished interaction with lecturers and peers. Trask et al. [17] explored performance predictions in online courses using heterogeneous knowledge graphs. Their research developed a model to

identify at- risk students, achieving 70–90% accuracy by analysing factors like consumed content, the institution, and learning modes.

Additionally, Bowers and Kumar found that technology integration in education significantly influences learning outcomes, emphasizing the necessity for robust digital tools [4]. Bernard et al. discussed the impact of blended learning in higher education, concluding that combining online and traditional methods enhances academic performance [3]. Mayer provided insights into multimedia learning, underscoring how technology aids in knowledge retention and comprehension [10]. Meanwhile, Prabowo et al. [12] identified that course quality, lecturer competence, and infrastructure are pivotal for successful online education. These research collectively highlight the opportunities and challenges of online learning and underscore the importance of technological and pedagogic improvements.

This research aims to identify and analyze whether there are differences in the academic performance of Informatics students in the Computer Network course across full online, hybrid, and full on-site learning modes at Multimedia Nusantara University. The findings are expected to provide insights into significant variations in exam results under these different learning conditions.

H_0: There is no significant difference in student performance across the three learning modes (online, hybrid, and on-site) for the Computer Network course.

H_1: There is a significant difference in student performance across at least two of the learning modes (online, hybrid, and on-site) for the Computer Network course.

II. Theoretical Basis

A. Final Exam Scores as a Measure of Performance

Final exam scores serve as an objective measure of students' academic achievements, reflecting their understanding of theoretical concepts and practical skills. Cognitive processes such as remembering, understanding, and applying are critical in assessing student performance In Computer Network courses, final exams evaluate students' mastery of networking theory and technical problem-solving[18].

B. The Impact of Learning Modes on Student Performance

The choice of learning mode online, hybrid, or on-site significantly influences student outcomes, particularly in practice-intensive courses like Computer Network.

1. Online Learning

Software simulations like Cisco Packet Tracer are among the digital tools used in online learning to replace in-person encounters. Nonetheless, studies indicate that experiential learning using actual hardware is more successful in promoting profound comprehension and skill recall [11].

2. Hybrid Learning

Balanced approach to theory and practice is provided by hybrid learning, which blends in-person practical sessions with virtual lectures. Research shows that by combining the advantages of both approaches, it can improve learning results and student engagement [8].

3. On-site Learning

In a completely immersive setting, on-site learning allows students to interact with real hardware, like switches and routers. According to experiential learning theories, the real-world learning opportunities provided by this method have been demonstrated to improve situational awareness and practical abilities[9].

Several research have explored the impact of online learning on student performance. For instance, Akpen et al. [1] conducted a systematic review highlighting both the benefits and challenges of online learning. While flexibility and accessibility can enhance academic performance, issues like reduced engagement and isolation remain significant concerns. Similarly, Trask et al. [17] analyzed student performance predictions in online courses using heterogeneous knowledge graphs, achieving a 70-90% accuracy rate in identifying at-risk students based on factors like learning mode and consumed content.

III. Methodology

A. Research Participants

The participants in this study were undergraduate Informatics students from Multimedia Nusantara University (2019–2022) who had completed at least four semesters, including the Computer Network course. Data were collected from lecturers teaching this course across online, hybrid, and on-site modalities during and after the COVID-19 pandemic. A total of 526 data points were analyzed, comprising 221 from online, 197 from hybrid, and 108 from on-site modes. This selection ensured academic homogeneity while enabling an effective comparison of learning outcomes across different teaching methods.

B. Research Procedure

Prior to analyzing the collected data, researchers must perform residual tests using normality test, homoscedasticity test, and autocorrelation tests. The normality test uses the Shapiro-Wilk Test, the homoscedasticity test uses the Levene Test, and the autocorrelation test uses the Durbin-Watson Test. All three residual tests will be done in R using the related functions. The following hypotheses are set for the residual test:

Normality test:

H_0: Data is distributed normally.

H_1: Data is not distributed normally.

Homoscedasticity test:

H_0: Data is homoscedastic.

H_1: Data is not homoscedastic.

Autocorrelation test:

H_0 : Final score and mode are not autocorrelated.

H_1 : Final score and mode are autocorrelated.

The computer network score index is calculated using the following equations:

Online mode with lab components:

$$0.67 \times ((Mid\ exam\ theory \times 0.3) + (Final\ exam\ theory \times 0.4) + (Theory\ Activities \times 0.3)) + 0.33 * ((Mid\ exam\ practicum * 0.3) + (Final\ exam\ practicum \times 0.4) + (Practicum\ Activities \times 0.3))$$

Hybrid & On-site mode:

$$= ((Mid\ exam \times 0.3) + (Final\ exam \times 0.4) + (Activities \times 0.3))$$

This equation is in accordance with the scoring guidelines of UMN.

C. Data collection

Data was obtained directly from lecturers who taught the Computer Network course in different year of study. The lecturer provided performance records and academic assessments of students across the three learning modalities. This data is considered primary data because it was sourced directly from individuals with firsthand knowledge of the participants' academic performance. Such direct access enhances data validity, as the information originates from trusted academic professionals. The approach aligns with the methodologies discussed by Creswell (2014) for collecting valid, primary data in educational research.

D. Data Analysis and Equations

In this research, the mean is used to provide a central measure of UMN Informatics students performance in the Computer Network course, offering an average score that reflects the typical achievement across online, hybrid, and on-site learning modes. The median serves as a robust central tendency measure, particularly valuable for handling skewed distributions or outliers, while the mode highlights the most frequently occurring scores. Variance and standard deviation are analysed to understand the spread and consistency of scores across the different modalities[6].

Skewness assesses the asymmetry of the score distribution, offering insights into biases and whether performance leans toward higher or lower scores. A distribution table and normality tests, such as the Shapiro-Wilk test, verify the data's conformity to a normal distribution, ensuring the validity of subsequent analyses. The Levene test evaluates homogeneity of variances across the learning modes, while the Durbin-Watson test examines potential autocorrelation in residuals, preserving the regression model's assumptions[12].

The Kruskal-Wallis test is used to detect significant differences in median performance across the three learning modes. When significant differences are found, the Dunn test is applied for pairwise comparisons, revealing specific differences between learning modes. Additionally, a pairwise Wilcoxon test is conducted to further confirm and explore the pairwise differences, providing more robust insights into the performance variability among the groups. This combination of tests ensures a thorough analysis of performance trends and the factors influencing academic achievement in the Computer Network course across online, hybrid, and on- site settings at Multimedia Nusantara University[15].

IV. Results and Analysis

A. Central Tendency and spread analysis

```
> descriptive_stats <- data %>%
+     group_by(Mode) %>%
+     summarise(
+         Mean = mean(Final_Score),
+         Median = median(Final_Score),
+         Mode = names(sort(table(Final_Score), decreasing = TRUE))[1],
+         Variance = var(Final_Score))
> print(descriptive_stats)
# A tibble: 3 × 4
  Mode      Mean Median Variance
  <chr>    <dbl>  <dbl>    <dbl>
1 80.7      72.2   75.5     269.
2 4.09959   80.1   82.5     200.
3 0         75.8   76.7     186.
```

Fig 1 Results of mean, median, mode, and variance of hybrid, online, and on-site modes

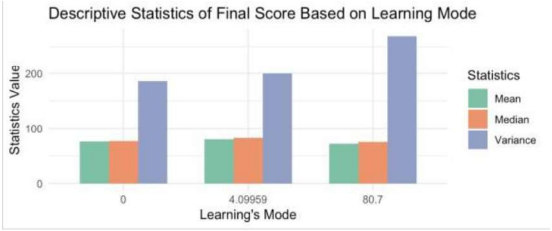


Fig 2 Histogram of central tendency

Descriptive statistics summarize the central tendency and variability in student performance across the three learning modes (Hybrid, Online, and On-site). Hybrid learning has the lowest mean score (72.2), followed by On-site (75.8), and Online with the highest mean (80.1), indicating that students in Online mode tend to perform better. However, variability differs significantly across modes, with Hybrid showing the highest variance (269), indicating greater score

dispersion, followed by Online (200) and On-site with the lowest variance (186). The mode for Hybrid is 80.7, appearing five times, suggesting optimal performance, while Online and On-site show no repeated scores. The medians for all modes slightly exceed the means, reflecting left-skewed (negatively skewed) distributions.

B. Shape measure Analysis

```
> skew_online <- skewness(data$Final_Score[data$Mode == "Online"], na.rm = TRUE)
> skew_hybrid <- skewness(data$Final_Score[data$Mode == "Hybrid"], na.rm = TRUE)
> skew_onsite <- skewness(data$Final_Score[data$Mode == "Onsite"], na.rm = TRUE)
> skew_online
[1] -3.051532
> skew_hybrid
[1] -2.37345
> skew_onsite
[1] -2.233519
```

Fig 3 Shape measure result

The skewness analysis of the final score distributions across the three learning modes Online, Hybrid, and On-site indicates that all distributions are left-skewed, as evidenced by negative skewness values (-3.072 for Online, -2.391 for Hybrid, and -2.265 for On-site). This left-skewness implies that the majority of students achieved higher scores, with fewer occurrences of extremely low scores. Among the three modes, the Online mode exhibits the strongest left-skewness, suggesting that students in this mode generally performed better, with a tighter concentration of high scores and minimal extreme low values. The Hybrid and On-site modes, while also showing a tendency for higher scores, display slightly less pronounced skewness, indicating a relatively more balanced score distribution compared to the Online mode. These patterns may reflect differences in the learning environment's influence on student performance, with the Online mode potentially fostering conditions conducive to achieving higher overall scores.

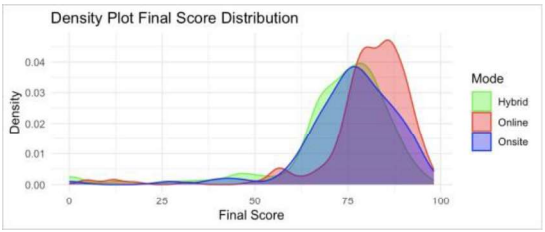


Fig 4 Density plot for skewness distribution of final score

The distribution of Final Scores among UMN students in the Computer Network course reveals that the On-site mode demonstrates the most stable and high performance, with scores concentrated in the 70–90 range and a density peak around 75–80, indicating consistent achievement. The Hybrid mode shows a wider spread, with a density peak around 65–75, reflecting greater variability. Meanwhile, the Online mode exhibits a more asymmetric distribution, with most scores falling in the 70–80 range. Overall, the On-site mode excels in score stability compared to Hybrid and Online modes.

C. Residual Test

Normality Test

```
> shapiro_results <- data %>%
+     group_by(Mode) %>%
+     summarise(p_value = shapiro.test(Final_Score)$p.value)
> print(shapiro_results)
# A tibble: 3 × 2
  Mode    p_value
  <chr>     <dbl>
1 Hybrid 2.33e-16
2 Online 1.68e-19
3 Onsite 7.55e-10
```

Fig 5 Normality test result

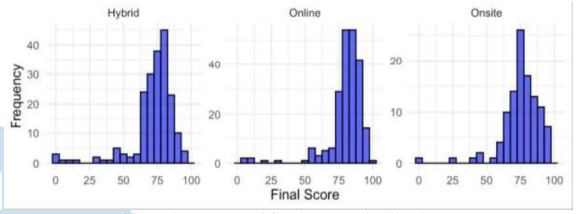


Fig 6 Histogram of final score distribution

The Shapiro-Wilk test was conducted to evaluate the normality of scores across learning modalities. The null hypothesis (H_0) assumes a normal distribution, while the alternative (H_1) indicates non-normality. The p-values for Online ($1.68e^{\wedge} - 19$), Hybrid ($2.33e^{\wedge} - 16$), and On-site ($7.55e^{\wedge} - 10$) were all significantly below 0.05, leading to the rejection of H_0. These results confirm that none of the distributions are normal.

Homoscedasticity Test

```
> leveneTest(Final_Score ~ Mode, data = data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2   1.196 0.3032
      523
```

Fig 7 Homoscedasticity Test result

The Levene's test was conducted to assess the equality of variances across the three learning modes (online, hybrid, and full onsite), a critical assumption for tests such as the Kruskal-Wallis test. The null hypothesis (H_0) posits that the variances are equal across the groups, while the alternative hypothesis (H_1) suggests that the variances differ. The test produced a p-value of 0.3032, which is greater than the significance threshold of 0.05. Consequently, the null hypothesis cannot be rejected, indicating that there is no statistically significant evidence to suggest unequal variances among the three groups. This result confirms that the variances are homogeneous, supporting the assumption of equal variances and ensuring the validity of subsequent statistical analyses.

Autocorrelation Test

```
> model <- lm(Final_Score ~ Mode, data = data)
> dwtest(model)

        Durbin-Watson test

data:  model
DW = 1.7295, p-value = 0.0006993
alternative hypothesis: true autocorrelation is greater than 0
```

Fig 8 Autocorrelation test results

The Durbin-Watson test was conducted to assess the presence of autocorrelation in the residuals of the regression model. The test produced a Durbin-Watson (DW) statistic of 1.7295 and a p-value of 0.0006993. Since the p-value is significantly smaller than the conventional threshold of 0.05, we reject the null hypothesis, which assumes that there is no autocorrelation in the residuals of the model. The results indicate that there is positive autocorrelation present. This suggests that the residuals of the model are not independent and exhibit a pattern where successive residuals are correlated. Positive autocorrelation can impact the validity of standard statistical inferences, such as confidence intervals and hypothesis tests, necessitating further investigation or adjustments to the model to address this issue.

D. Regression Model

```
> model <- lm(Final_Score ~ Mode, data = data)
> summary(model)

Call:
lm(formula = Final_Score ~ Mode, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-75.956  -3.950   2.536   8.530  23.430

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   72.170      1.063  67.893  < 2e-16 ***
ModeOnline     7.886      1.462   5.395 1.04e-07 ***
ModeOnsite     3.602      1.786   2.017   0.0443 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.92 on 523 degrees of freedom
Multiple R-squared:  0.05292,   Adjusted R-squared:  0.0493
F-statistic: 14.61 on 2 and 523 DF,  p-value: 6.677e-07
```

Fig 9 Multiple linear regression model summary

The linear regression model analyses the effect of learning modes (Hybrid, Online, and On-site) on final scores in the Computer Network course. The intercept, representing the Hybrid mode as the baseline, indicates an average score of 72.17 for this mode. Compared to Hybrid, the Online mode significantly improves scores by an average of 7.287 points (p-value < 0.0001), suggesting a strong positive effect on student performance. Similarly, the On-site mode adds an average of 3.602 points (p-value = 0.0443), but its impact is smaller and less statistically significant than the Online mode. These findings indicate that both Online and On- site modalities positively influence performance, with Online showing the strongest effect.

Despite these significant results, the model's adjusted R-squared value of 0.0493 reveals that only 4.93% of the variability in final scores is explained by the learning mode. This indicates that other factors, such as individual effort, instructor quality, or course design, likely play a significant role in performance. The F-statistic (14.61, p-value < 0.0001) confirms the model's overall significance, validating that learning mode impacts scores. However, the residual standard error (14.92) reflects considerable unexplained variability, highlighting the need for further research into additional determinants of academic achievement.

E. Kruskal-Wallis Test

```
> kruskal.test(Final_Score ~ Mode, data = data)

        Kruskal-Wallis rank sum test

data:  Final_Score by Mode
Kruskal-Wallis chi-squared = 57.24, df = 2, p-value = 3.719e-13
```

Fig 10 Statistical test result using Kruskal-Wallis

The Kruskal-Wallis test was employed to examine whether there were statistically significant differences in student performance among the three learning modalities (online, hybrid, and full onsite). This non-parametric test is particularly suitable for comparing groups when the assumption of normality may not be met. The null hypothesis (H_0) of the Kruskal-Wallis test posits that there are no significant differences in performance between the groups, meaning the distributions of scores are similar across the three modalities. Conversely, the alternative hypothesis (H_1) suggests that at least one group differs significantly from the others. The test produced a chi-square statistic, with degrees of freedom equal to 2, and a p-value of 3.719×10^{-13}. Since the p- value is far smaller than the standard significance threshold of 0.05, the null hypothesis is decisively rejected. This result provides strong evidence that there are statistically significant differences in student performance between at least two of the learning modalities. These findings highlight variations in how different teaching modes impact student outcomes, warranting further investigation into which specific modalities contribute to these differences.

F. Pairwise Comparison Test

```
> pairwise.wilcox.test(data$Final_Score, data$Mode, p.adjust.method = "bonferroni")

        Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data:  data$Final_Score and data$Mode

       Hybrid  Online
Online 3.4e-13 -
Onsite 0.13    8.0e-05

P value adjustment method: bonferroni
```

Fig 11 Pairwise comparisons result

The results of the pairwise Wilcoxon rank sum test show comparisons of Final Score across different Mode categories (Hybrid, Online, and Onsite), with the Bonferroni adjustment method used to reduce the risk of errors caused by testing multiple comparisons. This test checks if the distributions of final scores between each pair of modes are similar or not. For the comparison between Online and Hybrid, the p-value is 3.4e-13, which is very small and indicates a significant difference in final scores between these two modes. This means that the mode of delivery has a major effect on the final scores, and Online and Hybrid modes are quite different.

On the other hand, the comparison between Hybrid and Onsite gives a p-value of 0.13, which is not statistically significant at the common threshold of 0.05. This suggests that there is no meaningful difference in final scores between these two modes. Similarly, the comparison between Online and Onsite shows a p-value of 8.0e-05, which is statistically significant, indicating a significant difference between the two modes. The Bonferroni correction helps ensure that the results are accurate by adjusting for the multiple comparisons, reducing the chances of finding false positives.

G. Post-Hoc Test

```
> data$Mode <- as.factor(data$Mode)
> dunnTest(Final_Score ~ Mode, data = data, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

      Comparison         Z      P.unadj        P.adj
1 Hybrid - Online -7.441764 9.934986e-14 2.980496e-13
2 Hybrid - Onsite -1.990254 4.656293e-02 1.396888e-01
3 Online - Onsite  4.181100 2.901024e-05 8.703071e-05
```

Fig 12 Post-Hoc test results

The Dunn test was employed as a post-hoc analysis following the Kruskal-Wallis test to identify specific differences in student performance between the three learning modes: Hybrid, Online, and On-site. This method allows for pairwise comparisons to determine which learning modes have statistically significant differences in final scores. The results indicate that Online learning significantly outperforms Hybrid learning, with a Z-value of -7.441764 and an adjusted p- value of 2.98×10^{-13}. The extremely low p-value (less than 0.05) provides strong evidence of a substantial advantage for students in the Online mode compared to those in the Hybrid mode. Similarly, Online learning also outperforms On-site learning, as demonstrated by a Z- value of 4.181100 and an adjusted p-value of 8.70×10^{-5}. These findings highlight that Online learning yields the highest overall performance among the three modalities. In contrast, the comparison between Hybrid and On-site learning did not reveal a statistically significant difference in final scores. The Z-value of -1.990254 and the adjusted p-value of 0.1397 (greater than 0.05) suggest that the observed performance differences between these two modes are not significant. This implies that while Online learning stands out as the most effective mode in improving student outcomes, Hybrid and On-site modes perform similarly, with neither showing a clear advantage over the other.

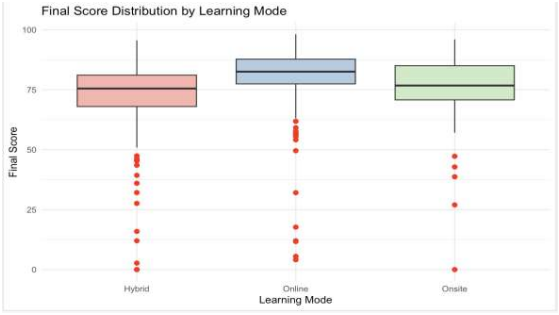


Fig 13 Box plot of final score distribution

The boxplot further supports these findings by visually illustrating the distribution of final scores across the three learning modes. Online learning exhibits a higher median and less variability in scores compared to Hybrid and On-site modes. Both Hybrid and On-site learning display broader score distributions, with overlapping interquartile ranges, which corroborates the statistical analysis indicating no significant difference between these two modes. Collectively, the results emphasize the effectiveness of Online learning in enhancing student performance while suggesting that Hybrid and On-site modes are comparable in their outcomes.

V. CONCLUSION

The research aimed to identify and analyse differences in academic performance among Informatics students in the Computer Network course across Online, Hybrid, and On-site learning conditions. The study revealed significant differences, with Online learning showing the highest mean performance (80.1), followed by On-site (75.8) and Hybrid (72.2). Online learning also demonstrated greater consistency with lower variance, while Hybrid learning exhibited the widest score dispersion, reflecting varied student experiences. Statistical analyses, including the Kruskal-Wallis test, pairwise comparison and post-hoc Dunn's test, confirmed significant performance differences between Online learning and the other modes, though no significant distinction was found between Hybrid and On-site learning.

In technical courses like computer networks, these results demonstrate the relative efficacy of online learning while highlighting areas where hybrid learning needs to be improved to provide more reliable results. Further studies are encouraged to investigate how instructional design, assessment types, and interaction levels contribute to student performance across various learning modes, particularly in practice-intensive courses like Computer Networks. In addition to offering deeper insights into how learning preferences and adaptability vary over time, longitudinal research

that monitor changes over several semesters may also assist educators in creating inclusive and fair teaching methods.

REFERENCES

[1] Akpen, C. N., Asaolu, S., Atobatele, S., Okagbue, H., & Sampson, S. (2024). Impact of Online Learning on Student's Performance and Engagement: A systematic review. Smart Learning Environments, 3(205). https://doi.org/10.1007/s44217-024-00253-0

[2] Aristovnik, A., Karampelas, K., Umek, L., Ravselj, D. (2023). Impact of the COVID-19 pandemic on online learning in higher education: A bibliometric analysis. Frontiers in Education, 8, Article 1225834. https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2023.1225834/full?form=MG0 AV3

[3] Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta- analysis of blended learning and technology use in higher education: from the general to the applied, Journal of Computing in Higher Education, 26(1), 87-122.https://doi.org/10.1007/s12528-013-9077-3

[4] Bowers, J., & Kumar, P. (2015). Students' perceptions of teaching and social presence: A comparative analysis of face-to-face and online learning environments. International Journal of Web-Based Learning and Teaching Technologies, 10(1), 27–44. https://eric.ed.gov/?id=EJ1111153

[5] Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. Journal of Educational Technology Systems, 49(1), 5–22. https://doi.org/10.1177/0047239520934018

[6] Febriani, S. (2022). Analisis deskriptif standar deviasi. Jurnal Pendidikan Tambusai, 6(1), 81–94. https://doi.org/10.31004/jptam.v6i1.8194

[7] Fikri, M., Ananda, M. Z., Faizah, N., Rahmani, R., & Elian, S. A. (2021). Kendala dalam pembelajaran jarak jauh di masa pandemi COVID-19: Sebuah kajian kritis. Jurnal Pendidikan dan Pengajaran, 9(1), 145–154. https://media.neliti.com/media/publications/561978-kendala-dalam-pembelajaran-jarak-jauh-di-08de8022.pdf

[8] Graham, C. R. (2013). Emerging practice and research in blended learning. In M. G. Moore (Ed.), Handbook of distance education (3rd ed., pp. 333–350). Routledge.

[9] Kolb, D. A. (2014). Experiential learning: Experience as the source of learning and development (2nd ed.). Pearson Prentice Hall.

[10] Mayer, R. E. (2014). Multimedia learning. Psychology of Learning and Motivation, 60, 85–139.

[11] Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. Teachers College Record, 115(3), 1–47. https://doi.org/10.1177/016146811311500307

[12] Prabowo, H., Ikhsan, R. B., & Yuniarty, Y. (2022). Student performance in online learning in higher education: A preliminary research. Proceedings of the 2022 International Conference on Information Technology, 7, 1–6. https://www.semanticscholar.org/paper/Student-performance-in-online-learning-higher-A-Prabowo-Ikhsan/4a6802d39378ea493a6d9b2c67fe7594d01e74f8

[13] Schindler, L. A., Burkholder, G. J., Morad, O. A., & Marsh, C. (2017). Computer-based technology and student engagement: A critical review of the literature. International Journal of Educational Technology in Higher Education, 14(1), 1–22. https://doi.org/10.1186/s41239-017-0063-

[14] Sherwani, R. A. K., Shakeel, H., Awan, W. B., Faheem, M., & Aslam, M. (2021). Analysis of COVID-19 data using neutrosophic Kruskal-Wallis H test. BMC Medical Research Methodology, 21, Article 105. https://doi.org/10.1186/s12874-021-01410-x

[15] Stallings, W. (2016). Data and computer communications (10th ed.). Pearson.

[16] Tanenbaum, A. S., & Wetherall, D. J. (2011). Computer networks (5th ed.). Pearson.

[17] Trask, T., Lytle, N., Boyle, M., Joyner, D., & Mubarak, A. (2024). A comparative analysis of student performance predictions in online courses using heterogeneous knowledge graphs. arXiv. https://doi.org/10.48550/arXiv.2407.12153

[18] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9, Article 11. https://doi.org/10.1186/s40561-022-00192-z

[19] Yildirim, Z. (2020). COVID-19 pandemic and distance education: The impacts on students' performance. Journal of Modern Education Review, 1(2), 45–52.

[20] Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. Journal of Educational Psychology, 81(3), 329–339.

ISSN 2355-0082

E-Commerce Product Review Sentiment Analysis: A Comparative Study of Naïve Bayes Classifier and Random Forest Algorithms on Marketplace Platforms

Cian Ramadhona Hassolthine[1], Toto Haryanto[2],
Fenina Adline Twince Tobing[3], Muhammad Ikhwani Saputra[4]
[123]School of Data Science, Mathematics and Informatics, IPB University
hassolthinecian@apps.ipb.ac.id, totoharyanto@apps.ipb.ac.id,
feninatobing@apps.ipb.ac.id, ikhwanisaputra@apps.ipb.ac.id

Accepted 16 June 2025
Approved 14 July 2025

***Abstract*— Achieving customer satisfaction and trust is a major challenge for success in the business world. Entrepreneurs must identify problems that arise from reviews given by customers. However, reading and sorting each review is time-consuming and considered inefficient. In order to overcome this, a study was conducted that aims to analyze sentiment on products sold in the Shopee marketplace using the Naïve Bayes Classifier and Random Forest algorithms. The focus of this study is on product reviews from XYZ Store. The main objective of this study is to determine a more accurate and efficient algorithm in classifying review sentiment, which can help companies in marketing strategies and product development. The results of this study can provide insight for companies about consumer responses to marketed products, so that they can be used as a basis for making strategic decisions to improve the quality of services and products. The results of the Random Forest method classification produce superior predictions compared to the Naïve Bayes Classifier method with an accuracy value of 92.5%, precision of 93%, Recall of 92.5% and F1-Score of 90%.**

***Index Terms*— Marketplace, Naïve Bayes Classifier Algorithm; Product Review; Random Forest Algorithm; Sentiment Analysis**

I. INTRODUCTION

Marketplaces have become the primary platform for consumers to search for and purchase products, making customer reviews an important source of information [1]. Sentiment analysis of these reviews helps understand customer perceptions, which is essential for improving service quality and making better business decisions. Various machine learning algorithms have been applied in sentiment analysis, including Naïve Bayes and Random Forest [2]. However, the effectiveness of these algorithms can vary depending on the characteristics of the data and different e-commerce platforms. Comparative research comparing the performance of these two algorithms on a particular marketplace platform is still needed to identify the best approach to understanding customer sentiment [3]. Sentiment analysis is a discipline of machine learning and Natural Language Processing (NLP) that functions to identify and extract opinions, sentiments, reviews, attitudes, and emotions contained in text, thus providing deeper insight into responses and views in a particular context [4]. The Naïve Bayes Classifier algorithm is a probability-based classification method that predicts an event based on historical data using Bayes' theorem [5]. Its main advantages are ease of implementation, computational speed, and effectiveness on high-dimensional datasets. The Random Forest algorithm was developed as an evolution of the CART (Classification and Regression Trees) method by utilizing the bagging technique (combining random samples) and random feature selection [6].

Research on sentiment analysis of e-commerce product reviews using the Naïve Bayes and Random Forest algorithms has been widely conducted, but the focus and context vary. In research [7] conducted sentiment analysis on Shopee e-commerce using the KNN and Support Vector Machine (SVM) algorithms. Another study by [8] analyzed sentiment reviews on Shopee using Naïve Bayes and SVM. The differences in context, dataset, and algorithms used indicate that there is still room for further research that focuses on comparing Naïve Bayes and Random Forest specifically on different marketplace platforms or with unique data characteristics. This study aims to conduct a comparative study between the Naïve Bayes Classifier and Random Forest algorithms in sentiment analysis of product reviews on e-commerce marketplace platforms. The results of this study are expected to provide guidance for e-commerce practitioners in choosing the most appropriate algorithm for sentiment analysis. In addition, this study also contributes to the literature by providing empirical evidence regarding the performance of both algorithms in the context of sentiment analysis of e-commerce product reviews.

II. METHODOLOGY

The methodology that will be conducted in this research is as follows.

A. Literature Study

Literature studies are carried out by taking and studying information from various literary sources such as journals, books or other scientific sources to support research according to existing theories.

B. Research Flow Diagram

The sentiment analysis system is designed to process and analyze buyer reviews from the Shopee marketplace on XYZ Store. This study begins by collecting a dataset of product reviews to be analyzed.

Consumer review data was taken from a product named "*Meja Belajar Polos A*" sold by the XYZ store on the Shopee Indonesia marketplace. A total of 844 reviews available on the review page were taken from the selected product page on the Shopee marketplace. Figure 1 presents a flow diagram in this research.

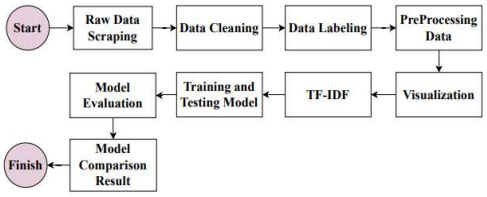


Fig 1. Research Flow Diagram

- Raw Data Scraping : The review data successfully collected from the product review page on the Shopee website amounted to 844 review data. This review data is called raw data which will then go through a data cleaning process.
- *Data Cleaning* : The results of raw data scraping are then continued to the data cleaning process. this stage involves data from csv format to xlsx format, Sort data based on "Transaction Time" by taking data from a time range, Cleaning noise in the data, namely "comment ID", "item ID", "Shop ID", "username", "User Name", "Anonymous", "Region", "Item Name", and "Transaction Time".
- *Data Labeling* : Data labeling on the dataset used is based on 2 categories, positive and negative sentiments. Positive sentiment if the review gets a rating of 4-5, while negative sentiment if the review gets a rating of 1-3. Labeling process is conducted manually and separately using tools of Microsoft Excel application. From the labeling results, a total of 351 positive sentiment records were obtained, while 42 negative sentiment records, a total of 393 records for the dataset used. So that there is an imbalance in the data, therefore the SMOTE (Synthetic Minority Over-sampling Technique) technique is used to understand the distribution of words in the data and identify dominant words. This happens because the dataset used does have a more dominant number of positive sentiments compared to the number of negative sentiments. The SMOTE oversampling method overcomes the problem of data imbalance by generating synthetic data for negative sentiments. Thus, the data distribution becomes more balanced, so that the machine learning model that will be applied in the next stage can learn more effectively and avoid bias that leads to positive sentiments.
- *Pre-processing Data* : The successful pre-processing process is carried out in several stages which can be explained as follows: Text Cleaning, Lowercase Folding, Tokenizing, Slang Word Conversion, Stopword Removing and Stemming [9].

1. Text Cleaning : The process of cleaning noise and removing symbols in the review. This process involves a series of steps to clean and process raw text, remove noise, and format text to make it more structured and relevant. Text cleaning aims to improve data quality so that analysis or modeling performed on text data can produce more accurate and meaningful results [10].
2. Lowercase Folding : Standardize all letters written in the review into lower case letters to have a standard form [11]. Figure 2 present Lowercase Folding.

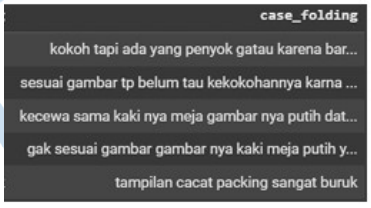


Fig 2. Lowercase Folding

3. Tokenizing : Changing sentences into words that are in accordance with the rules of the Indonesian dictionary so that sentences are more meaningful and are converted into token arrays [12]. The tokenizing stage is carried out using the help of the "nltk" library in the Google Collab application. Figure 3 present tokenizing result.

Fig 3. Tokenizing

4. Slang Word Conversion : perform conversion so that reviews containing slang words can be normalized into a more formal language. This is done to ensure that the text processing model can understand and process text more accurately, especially in the context of sentiment analysis, text classification, or other natural language processing applications [13]. Figure 4 slang word conversion

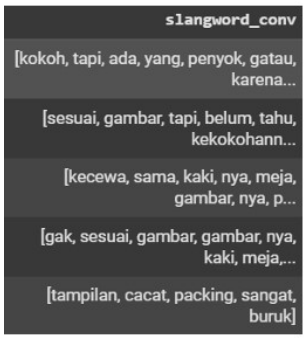


Fig 4. Slang Word Conversion

- *Visualization* : After the data has successfully gone through the stemming process, the next step is to present the visuals in the form of a word cloud which helps to make it easier to show the main words or topics that often appear in reviews that are grouped into positive and negative sentiments [14]. Frequently occurring words or topics will be large, while less common ones will be small. Figure 5 present Word Cloud for positive review.

Fig 5 Word Cloud or positive review

Figure 5 some words that often appear in positive sentiment reviews according to text size include "kokoh", "meja", "bagus", "barang", etc. In addition to positive sentiment reviews, here are the results of word cloud visualization on negative sentiment. Figure 6 present Word Cloud for negative reviews.

Fig 6. Negative Reviews Word Cloud

Figure 6 shows several words that often appear in negative sentiment reviews according to text size, including "kirim", "cacat", "retak", "kecewa", and "patah", etc. This word cloud visualization can make it easier to understand the main topic according to the sentiment in the data.

- *TF-IDF* : The TF-IDF stage begins by taking stemming data. One of the methods used to measure how important a word is in a document in a collection or corpus of documents. This method is often used in text analysis, information retrieval, and document modeling, such as in recommendation systems and text classification. [15]. This stage divides the data into 20% test data: 80% training data, then calculations are performed to find and display the frequency of the 10 most common words based on TF-IDF values. The goal is to understand the distribution of words in the data and identify dominant words.
- *Training and Testing Model* : The modeling process is carried out using the Naïve Bayes Classifier and Random Forest algorithms [16]. In this process, the data used is 393 data. The modeling process begins by dividing the training and test data. The ratio for training and test data is divided into 3 ratios, namely 90:10, 80:20, and 70:30. The purpose of this division is to find the optimal data division proportion for the model to be used, besides that it is also used to prevent overfitting.
- *Model Evaluation* : After the modeling process is complete, a model evaluation can be carried out from the confusion matrix produced by the model [17].

1. Accuracy : A measure of how well an SVM model classifies the given data [18]. Accuracy is calculated by comparing the number of correct predictions to the total number of data points tested

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

2. Precision : a metric used to measure how many positive predictions are actually correct (relevant) out of all the positive predictions made by the model [19].

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

3. Recall : a metric used to measure how many positives the model actually detected out of all the actual positive data [20].

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4. F1-Score : A metric that combines two important metrics in evaluating classification models, namely precision and recall, into one number that provides an overview of the balance between the two[20].

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

- *Model Comparison Result* : The results of an analysis that compares the performance of several machine learning models or algorithms in a particular task, with the aim of determining which model is the most effective and efficient based on relevant evaluation metrics.

III. RESULT AND DISCUSSION

After getting the confusion matrix results according to the data division during modeling, the next step is to calculate the accuracy, precision, recall, and F1 Score values of each algorithm model. The calculation results can be seen in Figure 7.

Test Size	Model	Accuracy	Precision	Recall	F1
90:10	Naive Bayes	0.875	0.915584	0.875	0.889328
90:10	Random Forest	0.925	0.930769	0.925	0.904
80:20	Naive Bayes	0.848101	0.908178	0.848101	0.868883
80:20	Random Forest	0.911392	0.895292	0.911392	0.892761
70:30	Naive Bayes	0.847458	0.915158	0.847458	0.868705
70:30	Random Forest	0.915254	0.904391	0.915254	0.903719

Fig 7. Model Evaluation Report

The figure 7 shows the evaluation results of the Naive Bayes and Random Forest models using various test data sizes for classification. The models used were evaluated based on four main metrics: Accuracy, Precision, Recall, and F1-Score. From the data shown, it can be seen that the Random Forest model consistently gives better results compared to Naive Bayes, especially at larger test sizes. For example, at a test size of 90:10, Random Forest has an accuracy of 0.925, while Naive Bayes only reaches 0.875. In general, Random Forest shows more stable performance across all metrics, with higher precision, recall, and F1 values than Naive Bayes. This shows that Random Forest is more effective in processing data with various training and test set sizes, while Naive Bayes tends to be slightly less optimal in this case. Below is a graphical image of the trend based on the data distribution ratio.

The figure 8 shows the performance trends of two classification models, Naive Bayes and Random Forest, based on four different metrics: Accuracy, Precision, Recall, and F1 Score. The graph compares the two models at different training and testing data split ratios. In general, Random Forest tends to give better results in Precision and Recall, especially at larger data split ratios, but there is a decrease in performance in F1 Score at a 70:30 split ratio. On the other hand, Naive Bayes shows more stable consistency in Accuracy, but with lower values in Precision and F1 Score compared to Random Forest. This trend indicates that although Random Forest is superior in terms of recall and precision, Naive Bayes can be a good choice if consistency in Accuracy is more important. A comparison of the evaluation metrics generated by each algorithm can be seen in the table 1.

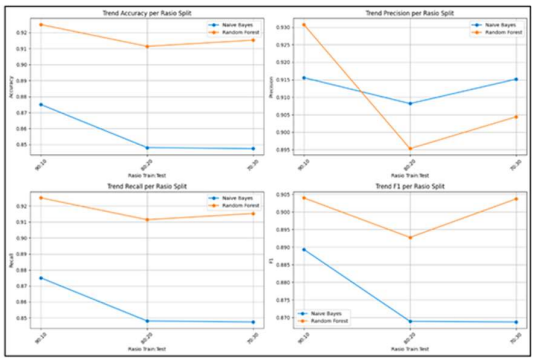


Fig 8. Data Share Ratio Trend Chart

TABLE I. COMPARISON OF OVALUATION METRICS RESULT

Evaluation Metrics	Naïve Bayes Classifier	Random Forest
Accuracy	87.5 %	92.5 %
Precision	91 %	93 %
Recall	87.5 %	92.5 %
F-1 Score	89 %	90 %

According to the table I, it shows that the performance of the Random Forest algorithm is better in predicting the majority class (positive). Based on the results of the model classification performance, it can be seen that the Random Forest algorithm has the highest accuracy value. Accuracy explains the extent to which the testing model can classify data correctly. When viewed from other evaluation metrics, the results also explain that the Random Forest algorithm is still superior. The confusion matrix for the Naive Bayes Classifier algorithm can be seen in Figure 9.

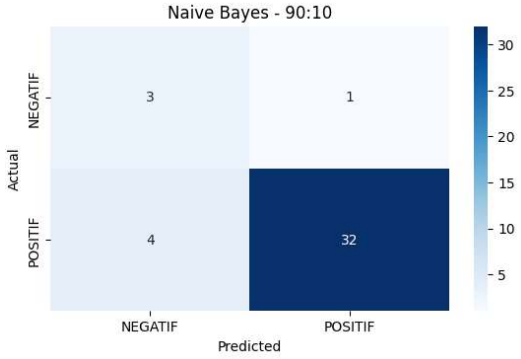


Fig 9. Confusion Matrix Naïve Bayes Classifier

The figure 9 shows the confusion matrix of the classification results using the Naive Bayes algorithm with a 90:10 data sharing scheme. From the matrix, it can be seen that the model successfully classified 3 negative data correctly (true negative) and 32 positive data correctly (true positive). However, there was 1 negative data that was incorrectly classified as positive (false positive) and 4 positive data that were incorrectly classified as negative (false negative). The model shows quite good performance in recognizing positive data, as seen from the high number of true positives. However, the model still has weaknesses in detecting negative data, because the number of false negatives and false positives still exists. This could be an indication that positive data is more dominant or the model is more sensitive to the positive class, so further evaluation is needed, for example by analyzing precision, recall, and F1-score to get a more comprehensive picture of model performance.The confusion matrix for the Random Forest algorithm can be seen in Figure 10.

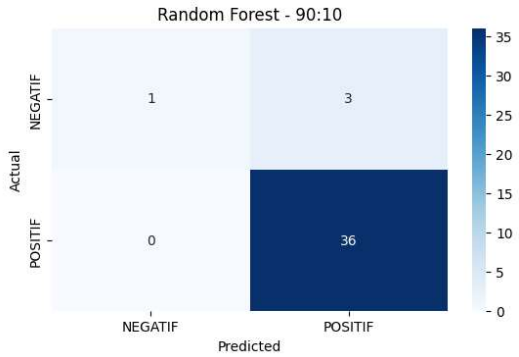


Fig 10. Confusion Matrix Random Forest

The confusion matrix in Figure 10 shows the classification results using the Random Forest algorithm with a 90:10 data sharing scheme. From the matrix, the model successfully classified 36 positive data correctly (true positive) and 1 negative data correctly (true negative). However, there were 3 negative data that were incorrectly classified as positive (false positive), while no positive data were incorrectly classified as negative (false negative). The Random Forest model is very good at recognizing positive data, as seen from the absence of false negatives and the high value of true positives. However, the model is still less than optimal in detecting negative data, as evidenced by the higher number of false positives compared to true negatives. This can be a consideration for adjusting the threshold or balancing the data so that the model's performance in the negative class can be improved.

Precision measures how many of the predicted positive cases are actually correct. In this case, Random Forest has a precision of 93%, while Naïve Bayes has 91%. This suggests that Random Forest has a slightly better ability to avoid false positives compared to Naïve Bayes. Recall measures how many actual positive cases were correctly identified by the model. Both models have the same recall value of 87.5% for Naïve Bayes and 92.5% for Random Forest, showing that Random Forest is better at identifying actual positive cases, leading to fewer false negatives. he F-1 Score balances precision and recall. Random Forest has a slightly better F-1 Score (90%) than Naïve Bayes (89%), reinforcing that it is more balanced in terms of both precision and recall.

IV. CONCLUSION

This research compares the Naïve Bayes Classifier and Random Forest algorithms in analyzing sentiment of e-commerce product reviews in the marketplace. The stages carried out include pre-processing, classification, and testing with data division at ratios of 90:10, 80:20, and 70:30. The best results were obtained at a ratio of 90:10. At the model evaluation stage, a confusion matrix was used to measure model performance, followed by the calculation of evaluation metrics. Random Forest produced an accuracy of 92.5%, higher than the Naïve Bayes Classifier which only reached 88%. In addition, Random Forest also excels in precision metrics (93%), recall (92.5%), and F1-Score (90%) in the positive class, indicating that this algorithm is more effective in predicting positive reviews. The results show that Random Forest is better than Naïve Bayes Classifier in classifying sentiment of e-commerce product reviews, especially in predicting positive reviews consistently. This research can be used by e-commerce players to improve marketing strategies, identify potential product problems, and optimize product development based on customer feedback. In addition, the use of sentiment analysis can help e-commerce automate the process of assessing and monitoring product reviews, thereby increasing operational efficiency and customer satisfaction.

REFERENCES

[1] A. Sharma, S. Kumar Mishra, dan V. Kant Srivastav, "The Evolution And Impact Of E-Commerce," 2023.

[2] S. Rashiq Nazar dan T. Bhattasali, "SENTIMENT ANALYSIS OF CUSTOMER REVIEWS," *Azerbaijan Journal of High Performance Computing*, vol. 4, no. 1, hlm. 113–125, Jun 2021, doi: 10.32010/26166127.2021.4.1.113.125.

[3] L. R. Krosuri dan R. S. Aravapalli, "Feature level fine grained sentiment analysis using boosted long short-term memory with improvised local search whale optimization," *PeerJ Comput Sci*, vol. 9, 2023, doi: 10.7717/PEERJ-CS.1336.

[4] K. Alemerien, A. Al-Ghareeb, dan M. Z. Alksasbeh, "Sentiment Analysis of Online Reviews: A Machine Learning Based Approach with TF-IDF Vectorization," *Journal of Mobile Multimedia*, vol. 20, no. 5, hlm. 1089–1116, 2024, doi: 10.13052/jmm1550-4646.2055.

[5] S. Ruan, H. Li, C. Li, dan K. Song, "Class-specific deep feature weighting for naïve bayes text classifiers," *IEEE Access*, vol. 8, hlm. 20151–20159, 2020, doi: 10.1109/ACCESS.2020.2968984.

[6] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu, dan L. Zhu, "A missing power data filling method based on improved

random forest algorithm," *Chinese Journal of Electrical Engineering*, vol. 5, no. 4, hlm. 33–39, Des 2019, doi: 10.23919/CJEE.2019.000025.

[7] A. Nurian, M. S. Ma'arif, I. N. Amalia, dan C. Rozikin, "ANALISIS SENTIMEN PENGGUNA APLIKASI SHOPEE PADA SITUS GOOGLE PLAY MENGGUNAKAN NAIVE BAYES CLASSIFIER," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, Jan 2024, doi: 10.23960/jitet.v12i1.3631.

[8] Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, dan Anis Fitri Nur Masruriyah, "Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine," *INFOTECH : Jurnal Informatika & Teknologi*, vol. 4, no. 1, hlm. 16–26, Jun 2023, doi: 10.37373/infotech.v4i1.422.

[9] H. He, G. Zhou, dan S. Zhao, "Exploring E-Commerce Product Experience Based on Fusion Sentiment Analysis Method," *IEEE Access*, vol. 10, hlm. 110248–110260, 2022, doi: 10.1109/ACCESS.2022.3214752.

[10] H. Huang, A. A. Zavareh, dan M. B. Mustafa, "Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions," 2023, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2023.3307308.

[11] S. Zhang dan H. Zhong, "Mining Users Trust From E-Commerce Reviews Based on Sentiment Similarity Analysis," *IEEE Access*, vol. 7, hlm. 13523–13535, 2019, doi: 10.1109/ACCESS.2019.2893601.

[12] S. A. A. Shah, M. Ali Masood, dan A. Yasin, "Dark Web: E-Commerce Information Extraction Based on Name Entity Recognition Using Bidirectional-LSTM," *IEEE Access*, vol. 10, hlm. 99633–99645, 2022, doi: 10.1109/ACCESS.2022.3206539.

[13] B. M. Shoja dan N. Tabrizi, "Customer Reviews Analysis with Deep Neural Networks for E-Commerce Recommender Systems," *IEEE Access*, vol. 7, hlm. 119121–119130, 2019, doi: 10.1109/ACCESS.2019.2937518.

[14] D. Chehal, P. Gupta, dan P. Gulati, "Evaluating Annotated Dataset of Customer Reviews for Aspect Based Sentiment Analysis," 2022, *River Publishers*. doi: 10.13052/jwe1540-9589.2122.

[15] L. Huang, Z. Dou, Y. Hu, dan R. Huang, "Textual analysis for online reviews: A polymerization topic sentiment model," 2019, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2019.2920091.

[16] H. Tufail, M. U. Ashraf, K. Alsubhi, dan H. M. Aljahdali, "The Effect of Fake Reviews on e-Commerce during and after Covid-19 Pandemic: SKL-Based Fake Reviews Detection," 2022, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2022.3152806.

[17] A. Qazi, N. Hasan, R. Mao, M. E. M. Abo, S. K. Dey, dan G. Hardaker, "Machine Learning-Based Opinion Spam Detection: A Systematic Literature Review," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3399264.

[18] A. Mardjo dan C. Choksuchat, "HyVADRF: Hybrid VADER-Random Forest and GWO for Bitcoin Tweet Sentiment Analysis," *IEEE Access*, vol. 10, hlm. 101889–101897, 2022, doi: 10.1109/ACCESS.2022.3209662.

[19] S. Gupta, B. Kishan, dan P. Gulia, "Comparative Analysis of Predictive Algorithms for Performance Measurement," *IEEE Access*, vol. 12, hlm. 33949–33958, 2024, doi: 10.1109/ACCESS.2024.3372082.

[20] Y. Huang, R. Wang, B. Huang, B. Wei, S. L. Zheng, dan M. Chen, "Sentiment Classification of Crowdsourcing Participants' Reviews Text Based on LDA Topic Model," *IEEE Access*, vol. 9, hlm. 108131–108143, 2021, doi: 10.1109/ACCESS.2021.3101565.

ISSN 2355-0082

Personalized Restaurant Recommendations: A Hybrid Filtering Approach for Mobile Applications

Christopher Matthew Marvelio[1], Alexander Waworuntu[2]
[1,2] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]christopher.marvelio@stutent.umn.ac.id, [2]alex.wawo@umn.ac.id

Accepted 16 June 2025
Approved 15 July 2025

***Abstract*— Selecting a restaurant that suits your taste can be a major challenge for consumers, especially given the vast array of online dining options. Traditional recommendation systems or simple filtering methods often fail to handle this complexity well. To address these limitations, we developed a mobile app-based restaurant recommendation platform that combines content-based filtering and collaborative filtering methods in a hybrid approach. The application was built using Expo, React Native, Express, and Flask technologies. The evaluation was conducted using the End-User Computing Satisfaction (EUCS) framework, and the results showed a very high user satisfaction rate of 93.9%. This result demonstrates that the recommendation system we developed is effective in providing relevant suggestions and is well-received by users.**

***Index Terms*— Collaborative Filtering; Content-Based Filtering; Hybrid Filtering; Mobile Application; Restaurant Recommendation**

I. INTRODUCTION

The rapid development of digital technology has brought about major changes in various aspects of human life. People now communicate, search for information, and organize daily activities more easily and efficiently [1] [2]. This progress not only makes life more practical but also opens up new opportunities in the fields of education, entertainment, and daily lifestyles. One real form of this progress is the existence of mobile applications that have now become an important part of modern society [3]. In the past, applications only functioned as a simple tool. However, now, applications have developed into smart solutions that support work, maintain health, and facilitate access to consumption services, such as online food delivery. These services are very helpful, especially for people who are busy, have no time to cook, or have difficulty leaving the house [4]. However, these conveniences also raise a new challenge. The many menu choices often make it difficult for users to determine foods that suit their taste. The results of a survey from 1,000 users of online food delivery services in Indonesia (age 18–45 years) show many users felt frustrated when they had to choose food because too many choices didn't match their personal preference [5].

To overcome the problem of confusion in choosing food, a restaurant recommendation system can be a very helpful solution. This system is designed to provide suggestions for places to eat or menus that suit the user's taste [6]. One method that is quite effective in this system is Collaborative Filtering (CF). The CF method works by analyzing the assessment patterns and habits of users in choosing food. Based on this data, the system looks for similarities between one user and another. Thus, the system can recommend food that is liked by people who have similar tastes [7]. However, although quite effective, the CF method also has several weaknesses. One of them is the cold start problem, which is when the amount of user data is still small, so the system has difficulty providing accurate recommendations. In addition, CF is also less effective in handling menus that are rarely rated, and it requires a large computational process if the data is very large. To overcome these weaknesses, some previous studies have tried to use the user-based CF approach, which relies on user assessment and demographic data like age or gender to find users who have similarities. In fact, there is also research that combines CF with the matrix factorization method so the food recommendation system can work more optimally.

Besides the CF method, the Content-Based Filtering (CBF) method is also often used in recommendation systems. CBF works by analyzing the characteristics of items, such as food types, and matching them with the preferences of each user [8]. Unlike CF, which relies on the opinions of many users, CBF focuses more on the details of the item itself, so it can provide more personal and specific suggestions. However, CBF also has several weaknesses. This method requires a complete description of the item and tends to only suggest items that are similar to those previously liked, making the recommendations given less varied. In the context of food recommendations, this can cause users to receive suggestions that are too limited. A number of studies on CBF have tried to develop food recommendation systems not only based on user tastes but also adjusted

to healthy eating patterns based on the food's nutritional value. In addition, various recent studies have developed more sophisticated mobile application-based recommendation systems by combining techniques such as matrix factorization, feature extraction, and user location information. With the addition of these techniques, the system is expected to provide more accurate and contextual recommendations, according to user conditions in real-time.

With the various limitations of each method, it's important to continue developing how recommendation systems work so they can provide more accurate suggestions tailored to user needs. One promising solution is a hybrid recommendation system, which combines Collaborative Filtering (CF) and Content-Based Filtering (CBF) methods [9]. Research shows that hybrid recommendation systems generally perform better than using only one method [10]. By combining the advantages of CF—which analyzes patterns from many users—and CBF—which focuses on item characteristics—this system is able to provide more relevant, diverse, and personalized suggestions. There are various ways to combine these two methods in a hybrid recommendation system. One is weighted combination, where the results of CF and CBF are combined by assigning specific weights to each result, allowing the system to adjust how much influence each method has. Additionally, a switching strategy approach dynamically selects the most appropriate method based on user data conditions—for example, using CBF when user data is still limited and switching to CF when user data is sufficient. Another method is feature combination, which integrates information from user and item characteristics into a more comprehensive recommendation model. Finally, there's a multilevel structure approach, where the recommendation process is carried out in stages; for instance, the initial stage filters items using one method, and the next stage refines the results with another. These four approaches enable hybrid recommendation systems to work more flexibly and effectively in generating relevant and diverse suggestions for users. One study applied a hybrid model to a music playlist recommendation system, combining collaborative data and song characteristic information, thus overcoming the problem of limited data. Another study developed a hybrid system specifically for food and restaurant recommendations, and the results show this approach has great potential to improve the quality and accuracy of recommendations [11].

II. Theoretical Framework

A. Machine Learning

Machine Learning is a field of science that focuses on how to make computer systems learn by themselves from data, without having to be specifically programmed for every situation. So, instead of writing code for every possibility, developers simply provide data, and the machine will learn from there. In this way, Machine Learning algorithms can carry out tasks automatically after learning and processing the available information. This ability to keep learning and adapting is the main strength of Machine Learning and is the reason why this technology is widely used in various aspects of life. In recommendation systems, Machine Learning plays an important role. This technology helps the system recognize user tastes and understand the characteristics of existing items. As a result, the system can provide more appropriate suggestions that are in accordance with user needs [12].

B. Recommendation System

Recommendation systems are intelligent platforms designed to provide suggestions tailored to each user's taste and needs. They are used in many fields and help improve the user experience of the services or products they use. Their main goal is to suggest items or content that users are more likely to find interesting or useful [13]. To do this, recommendation systems use special models and algorithms that analyze data. These algorithms study data from previous user activity, such as search history or ratings, to predict what users might like in the future. In this way, the system helps users make more informed choices that are in line with their interests [14].

C. Collaborative Filtering (CF)

Collaboration Filtering (CF) is a common and proven approach in recommendation systems. CF works by looking for patterns of similarity between users or items to generate suggestions [15]. Its main goal is to predict how a user will rate or respond to an item they haven't seen or used before.

CF has two main methods: user-based CF and item-based CF. In user-based CF, the system searches for other users who have similar preferences to the target user. After that, the system recommends items that are liked by similar users. In contrast, item-based CF focuses on finding items that are similar to items that have been liked by the target user. This similarity is calculated based on rating patterns from many users. As a result, the system suggests similar items to the users [16].

One of the main advantages of CF is that it does not require detailed item descriptions or complete user profiles. However, CF also has several challenges. One of them is the data scarcity problem, which happens when the initial data is still small, so the system has difficulty providing accurate recommendations. There is also the cold-start problem, which is the difficulty of giving suggestions for new users or new items that don't have any history of usage. In addition, if the number of users and items increases, CF systems can require large computing power to process all this information [17].

To measure similarity in CF, various techniques are employed. For item-based collaborative filtering, a set of items rated by a user is utilized to identify the most similar items based on their ratings. The Mean Squared Differences (MSD) method is used to quantify the similarity between items by evaluating the accuracy of predicting one item as a sole recommender for another [18].

The prediction of a rating for an item a by user u (Pu,a) using only one neighboring item b is calculated as shown in Equation (1):

$$P_{u,a} = \bar{r}_a + \left(r_{u,a} - \bar{r}_b\right) \quad (1)$$

In Equation (1), $r_{u,a}$ represents the rating of item a by user u. $\bar{r}_a$ and $\bar{r}_b$ are the average ratings for item a and item b, respectively.

The MSD similarity ($\text{iSim}_{a,b}^{MSD}$) is defined as shown in Equation (2):

$$i\text{Sim}_{a,b}^{MSD} = 1 - \left(\frac{\sum_{u=1}^{|U_{a\sim b}|}\left(P_{u,a} - r_{u,a}\right)^2}{|U_{a\cap b}|}\right) \quad (2)$$

This formula measures similarity based on the differences between predicted and actual ratings of common users. The scores are normalized using max-min normalization to ensure values fall between 0 and 1.

Furthermore, the Ochiai similarity measure is used to consider the percentage of common users who have rated both items ($U_{a\cap b}$) relative to the total number of users who have rated each item individually (U_a and U_b), as shown in Equation (3):

$$\text{iSim}_{a,b}^{\text{Ochiai}} = \frac{|U_{a\cap b}|}{\sqrt{|U_a| \times |U_b|}} \quad (3)$$

The overall CF similarity ($\text{iSim}_{a,b}^{CF}$) is then computed by combining the MSD and Ochiai similarities, as shown in Equation (4):

$$i\text{Sim}_{a,b}^{GF} = i\text{Sim}_{a,b}^{MSD} \times i\text{Sim}_{a,b}^{Och} \quad (4)$$

D. Content-Based Filtering (CBF)

Content-Based Filtering (CBF) is an algorithm used in recommendation systems that operates by recommending content similar to a user's past interactions. This algorithm analyzes the characteristics of the content a user has interacted with, such as keywords, topics, or genres, to predict their preference for similar content in the future. This approach helps users to discover content that aligns with their historical taste and preferences [19].

A key advantage of CBF is its ability to provides recommendation for new user or new items without needing historical data from other user. This makes it particularly useful in "cold start" scenario where collaborative filtering method might struggles. However, a significant drawback from CBF is overspecialization. This happens because recommendation are limited to items very similar with the user's past preferences, which can reduce diversity and prevent discovering new and various item. This limitation often arise when item description or category is incomplete [8].

In item-based content similarity, each item are associated with its respective category. If two item share the same categories, they considered related or similar to each other. All items are represents as vectors with values of [0, 1].

The representation of an item a as a vector (v_a) is given by:

$$v_a = \left(v_{a,1}, v_{a,2}, \ldots, v_{a,c}\right) \quad (5)$$

$$v_{a,c} = \begin{cases} 1 & \text{if item a belongs to category C} \\ 0 & \text{if item a does not belong to category C} \end{cases}$$

Subsequently, vector-based cosine similarity is employed to compute the content-based similarity between pairs of items. This measurement quantifies the cosine from the angle between two vectors, indicating their similarity.

$$\text{Sim}_{a,b}^{\text{Content}} = \frac{\sum_{u=1}^{n} V_a \times V_b}{\sqrt{\sum_{u=1}^{n} (V_a)^2} \times \sqrt{\sum_{u=1}^{n} (V_b)^2}} \quad (6)$$

E. Hybrid Content-Based and Collaborative Filtering

Hybrid recommendation systems represent an advanced approach that combines Collaborative Filtering (CF) and Content-Based Filtering (CBF) to overcome the individual limitations of each method and achieve higher recommendation accuracy. By integrating the strengths of both, these systems can leverage general knowledge derived from user interactions (CF) and the detailed item characteristics (CBF) to broaden the scope of recommendations. This comprehensive approach allows the system to consider a wider range of factors, such as user preferences and item attributes, thereby enhancing the overall quality and relevance of recommendations.

Various techniques can be employed to implement hybrid recommendation systems:

- Switching: Selects the best recommendation method based on the specific situation or context.
- Mixed: Integrates recommendations from various methods into a single list.
- Feature Combination: Merges features from different methods. This method is often chosen for its ability to optimally combine the advantages of both CF and CBF.
- Feature Augmentation: Adds new features to improve recommendations.

- Cascade: Helps resolve conflicts when recommendations from different methods contradict each other.
- Meta-level: Uses the output of one method as input for another.

This research specifically utilizes the Feature Combination method for the food recommendation system. This choice was motivated by its proven effectiveness in optimally merging the strengths of both CF and CBF. CF relies on past user behaviors and preferences for recommendations, while CBF uses item or food information and their characteristics. By combining the features from both methods, the recommendation system can generate more accurate and relevant suggestions for each user. The effectiveness of the Feature Combination method in improving recommendation accuracy has been shown in previous studies. For example, one study applied feature combination in a recommendation system by integrating a user-based CF approach with demographic information. Another research focused on a music playlist recommendation system using feature combination, which merged collaborative information from music playlists with song feature vectors from different sources. This integration helped solve data sparsity and improved song representations, leading to better recommendations, especially in cold-start cases and for songs that were not popular.

The hybrid recommendation process, particularly using the feature combination method, typically involves three stages:

1. Item-based Collaborative Filtering Similarity: Calculates the similarity between items using CF principles.
2. Item-based Content Similarity: Calculates the similarity between items using CBF principles.
3. Hybrid Prediction: Combines the similarities calculated in the previous two stages to generate final predictions.

In the hybrid prediction step, the predicted rating for an unknown item a by user u is typically derived through a two-step process. First, a Weighted Sum approach calculates the total score for an item by combining the predicted scores from both Collaborative Filtering (CF) and Content-Based Filtering (CBF). This weighted sum is based on the ratings provided by user u for items b that are most similar to item a, as used in both item-based CF and item-based content methods for prediction. The predicted score from Content-Based Filtering ($P_{u,a}^{\text{Content}}$) is given by Equation (7):

$$P_{u,a}^{\text{Content}} = \frac{\sum_{b \in I} \left(r_{u,b} \times \text{Sim}_{a,b}^{\text{Content}} \right)}{\sum_{b \in I} \left| \text{Sim}_{a,b}^{\text{Content}} \right|} \quad (7)$$

Meanwhile, the predicted score from Collaborative Filtering ($P_{u,a}^{CF}$) is given by Equation (8):

$$P_{u,a}^{CF} = \frac{\sum_{b \in I} \left(r_{u,b} \times \text{Sim}_{a,b}^{CF} \right)}{\sum_{b \in I} \left| \text{Sim}_{a,b}^{CF} \right|} \quad (8)$$

Subsequently, linear weighted hybridization is applied to combine these predicted ratings from both item-based CF and item-based content methods to produce a final hybrid prediction. This is represented by Equation (9):

$$P_{u,a}^{\text{Hybrid}} = \lambda \cdot P_{u,a}^{CF} + (1 - \lambda) \cdot P_{u,a}^{\text{Gontent}} \quad (9)$$

In Equation (9), λ and 1−λ (where λ∈[0,1]) represent the relative significance of the item-based CF and item-based content predictions in the final hybrid prediction.

F. Vector Space Model (VSM)

The Vector Space Model (VSM) is a conceptual framework frequently used in information retrieval to represent documents in a way that makes comparison and searching easier. In VSM, each document is conceptualized as a point inside a multi-dimensional space. Every dimension in this space corresponds to a unique word or term that is present across the document collection [20].

The presence and significance of a word in a document are quantified and represented by a numerical weight along its corresponding dimension. When a search query is initiated, it is also transformed into a vector within the same multi-dimensional space. The similarity or relevance between documents, or between a document and a query, is then determined by the "closeness" of their points (vectors) in this space. For example, a smaller angle between two vectors indicates a higher similarity [21].

The main goal of VSM is to place documents with similar topics or content close to each other in the vector space. This arrangement makes it easier to find relevant documents during a search. VSM proves very effective because it allows the calculation of similarity using mathematical formulas, making the search process both efficient and more effective.

G. Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) is a powerful mathematical technique used to break down a rectangular matrix into three simpler and more interpretable matrices. The three matrices consist of two orthogonal matrices and one diagonal matrix [22].

The main utility of SVD lies in its ability to reveal hidden patterns and structures in data by transforming it into a form that is easier to analyze. The two orthogonal matrices resulting from this decomposition represent the rows and columns of the original matrix, but in a way that highlights the relationships between them. Meanwhile, the diagonal matrix contains singular values, which show the "strength" or importance of those identified relationships. This decomposition

makes SVD an a useful tool for dimensionality reduction, noise reduction, and identifying latent semantic factors in data, which can be really beneficial in various data analysis or machine learning applications [23].

H. End-User Computing Satisfaction (EUCS)

End-User Computing Satisfaction (EUCS) is a recognized method designed to measure the level of satisfaction users have with an information system. The main purpose for using EUCS is to determine how effectively a system fulfills user requirements by comparing their expectations with the actual performance and features of the system [24].

The EUCS model is structured around five key aspects that are considered influential in shaping user satisfaction [25]:

- Content: This aspect pertains to the quality and completeness of the information provided by the system.
- Accuracy: This refers to the correctness and truthfulness of the information generated and presented by the system.
- Format: This dimension evaluates how easily information displayed by the system can be read and understood.
- Ease of Use: This focuses on how simple and straightforward the system is for users to learn and interact with.
- Timeliness: This relates to the promptness and efficiency with which the system delivers necessary information to the user.

By evaluating these five criteria, EUCS provides a comprehensive assessment of user satisfaction, making it a valuable tool for system developers and researchers to understand user perception and identify areas for improvement.

III. METHODOLOGY

The research employs a structured approach encompassing several distinct phases to achieve its objective. This systematic progression begins with a comprehensive literature review, followed by meticulous data collection, a well-defined system design phase, subsequent implementation, and concludes with a thorough evaluation of the developed system. This method is well-suited for the development and assessment of a mobile-based recommendation system, allowing both the creation of the system and the measurement of its effectiveness from a user satisfaction perspective.

Data collection and preprocessing were critical steps in preparing the necessary information for the restaurant recommendation system. The primary data source of this research was the Grab Food API. Through API calls, a comprehensive dataset was acquired, including merchant info, detailed menu descriptions, restaurant locations, pricing, categories, and user ratings. This raw data was then subjected to a rigorous preprocessing pipeline to ensure quality, consistency, and suitability for filtering algorithms. Key preprocessing steps involved cleaning and transforming the raw data to fix any inconsistencies or errors. Specifically, the 'tags' column, which was initially stored in JSON format in the restaurant table, was converted into a string. Furthermore, for the purpose of Content-Based Filtering (CBF), the 'chain' column and the now-string 'tags' column were combined to create a new 'combined' feature, enriching the dataset for similarity calculation. Throughout this phase, missing values and inconsistencies were handled carefully to maintain data integrity and avoid potential bias in the recommendation process. Finally, the acquired and preprocessed data, including restaurant info, user ratings, and user preferences, was transformed into Pandas DataFrames for efficient manipulation and analysis in the system.

The system design outlines the architectural blueprint and operational flow of the mobile-based restaurant recommendation system, specifying its core components, user interactions, and the underlying data structures. The system architecture consists of a frontend, a backend, and a database, working together to deliver personalized restaurant recommendations. The mobile application, serving as the frontend, is developed using Expo and React Native, ensuring cross-platform compatibility and a responsive user interface. The backend services, which are responsible for data processing and recommendation logic, are built using Express and Flask.

The application's processes and user interactions are meticulously illustrated through various flowcharts. The key operational flows are described as follows:

1. Landing Page and Sign In/Sign Up Process: This flow (Fig. 1) begins by presenting users with the option to either sign up for a new account or sign in if they are existing users. For new sign-ups, user data is securely stored in the database, and the user is then directed to the home page. Existing users input their email and password, which are validated against the database; successful authentication leads them to the home page.
2. Home Page Flow: This flow (Fig. 2) dynamically checks for the presence of user ratings data in the database. If ratings data exists, the system displays hybrid-filtered restaurant recommendations; otherwise, an empty recommendation message is presented.
3. Hybrid Filtering Recommendation Process: This core process (Fig. 3) orchestrates the generation of personalized recommendations. It starts with the initial acquisition of comprehensive user data,

ratings, restaurant details, and menu information. Subsequently, data preprocessing and feature extraction are performed to prepare the data. In the Collaborative Filtering (CF) phase, the system creates a matrix between users and all restaurants/menus based on their ratings. User similarity is then calculated based on their given ratings. Concurrently, the Content-Based Filtering (CBF) phase calculates restaurant and menu similarity based on their content or features, such as name and category. The similarities derived from both CF and CBF are then combined, and the recommendations are further refined by applying user-defined preferences. Finally, the system sorts the combined items and retrieves the top 10 items as recommendations to be displayed to the user.

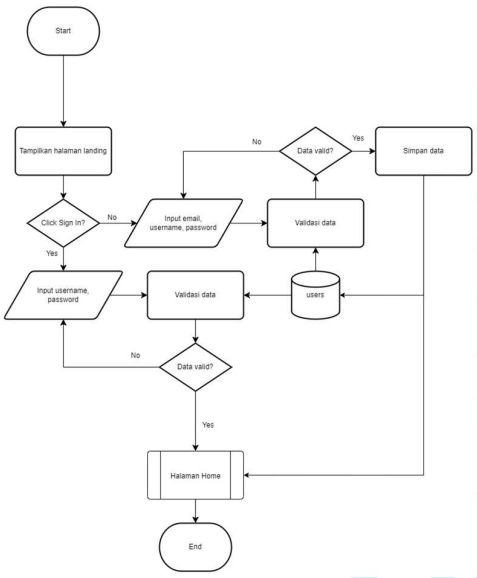


Fig. 1. Flowchart of Landing Page and Sign In/Sign Up Process

4. This flow (Fig. 4) facilitates user interaction with unrated restaurants and menus. It involves fetching data from the API, and any new restaurant or menu data not already in the database is added. Unrated restaurants and menus are presented to the user in a stacked card format. Users can interact with these cards by swiping left to dislike or right to like an item; these interactions update their preferences and ratings in the database.
5. Profile Page Flow: This flow (Fig. 5) allows users to manage their account and recommendation preferences. Users can modify their recommendation preferences, such as maximum price, minimum rating, and maximum distance, through interactive sliders. The system also provides options to reset all existing ratings, which deletes previously performed ratings , and to log out, returning the user to the landing page.

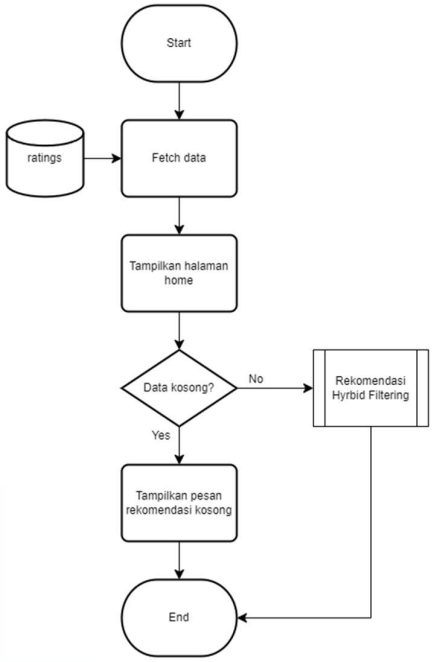


Fig. 2. Flowchart of Home Page

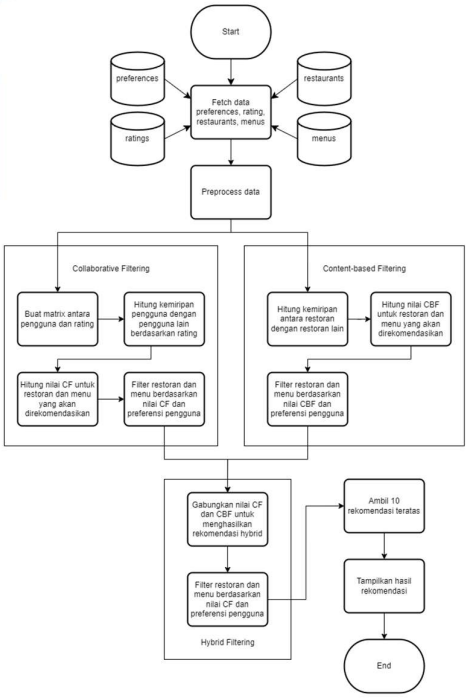


Fig. 3. Flowchart of Hybrid Filtering Recommendation

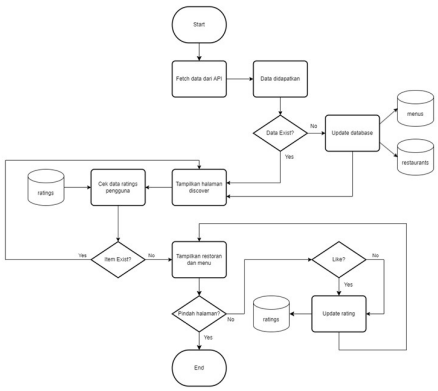


Fig. 4. Flowchart of Discover Page

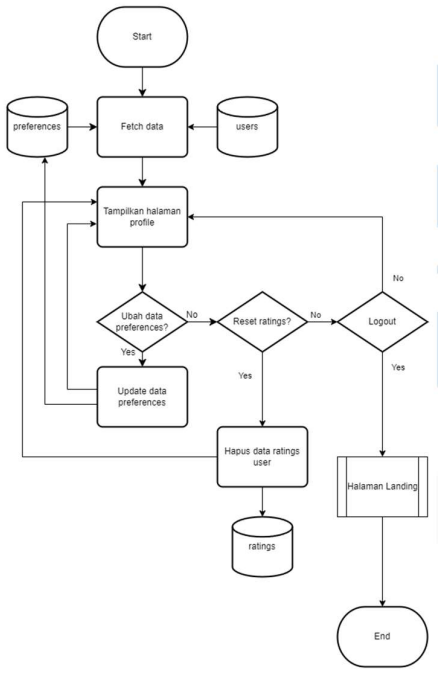


Fig. 5. Flowchart of Profile Page

The underlying Database Structure (Table 1) is meticulously designed to support the restaurant recommendation system's functionality. It comprises several key tables:

1. The users table stores user account information (id: UUID, created_at: timestamp, email: varchar, password: varchar).
2. The preferences table manages user-specific recommendation settings (id: INT, user_id: UUID FK, price_high: INT, min_rating: FLOAT, max_distance: INT, halal_only: BOOL).
3. The restaurants table holds detailed information about each restaurant (resto_id: varchar PK, name: varchar, latitude: float, longitude: float, rating: float [0-5], image_url: varchar, chain: varchar, halal: bool, tags: json).
4. The menus table stores specific menu item details (id: varchar PK, name: varchar, resto_id: varchar FK, image_url: varchar, price: int, description: text).
5. The ratings table records user feedback on restaurants and menus (id: INT PK, created_at: timestamp, user_id: UUID FK, resto_id: varchar FK, menu_id: varchar FK, is_liked: bool).

TABLE I. DATABASE STRUCTURE

Table	Column	Data Type
users	id	uuid
	created_at	timestamp
	email	varchar
	password	varchar
preferences	id	int
	user_id	uuid
	price_high	int
	min_rating	float
	max_distance	int
	halal_only	bool
restaurants	resto_id	varchar
	name	varchar
	latitude	float
	longitude	float
	rating	float
	image_url	varchar
	halal	bool
	chain	varchar
	tags	json
menus	id	Varchar
	name	varchar
	resto_id	varchar
	image_url	varchar
	price	int
	description	text
ratings	id	int
	created_at	timestamp
	user_id	uuid
	resto_id	varchar
	is_liked	bool

IV. RESULT AND IMPLEMENTATION

The User Interface Implementation provides the user-facing components of the mobile application, ensuring intuitive and engaging user experience.

1. The Landing Page (Fig. 6) serves as the initial screen upon application launch, prominently displaying the application's name, "FoodieMatch," alongside clear "Sign In" and "Sign Up" buttons.
2. The Sign In Page (Fig. 7) facilitates user login, prompting for email and password input. Upon successful authentication, users are directed to the home page.

3. For new users, the Sign Up Page (Fig. 7) allows for account registration by requiring email, password, and password confirmation. Valid new user data is then registered, and the user is automatically authenticated and navigated to the home page.

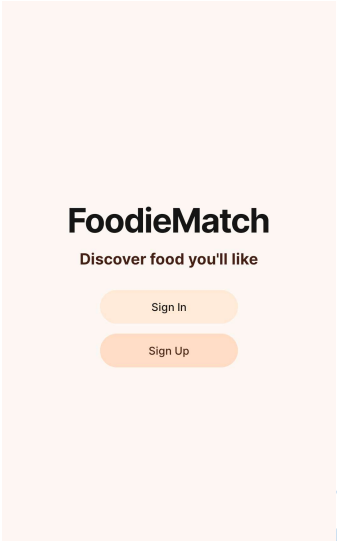


Fig. 6. Implementation Result of Landing Page

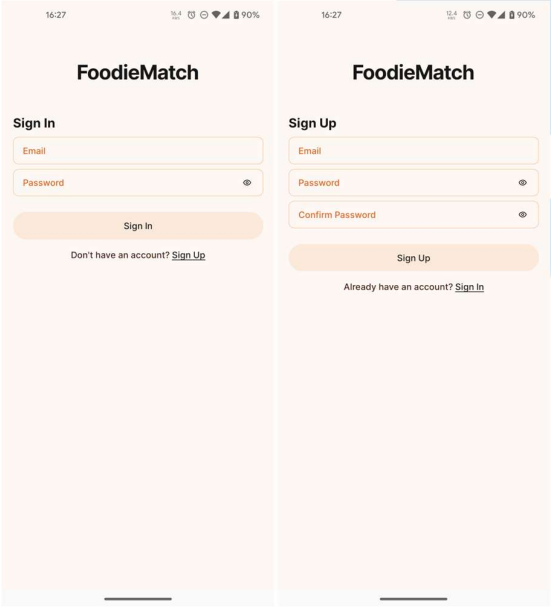


Fig. 7. Implementasion Result of Sign In and Sign Up Pages

4. The Home Page (Fig. 8) is designed to dynamically display restaurant recommendations or an empty message if no ratings data exists. It incorporates a navigation bar with tabs for "Home," "Discover," and "Profile" to facilitate seamless navigation. Users can also view a list of recently liked restaurants (Fig. 9) and access detailed information about specific restaurants and their menus by tapping on a recommended item.

5. The Discover Page (Fig. 10) presents unrated restaurants and menus in a stacked card format, allowing users to interact by swiping left to dislike or right to like an item. These interactions dynamically update their preferences and ratings in the database.

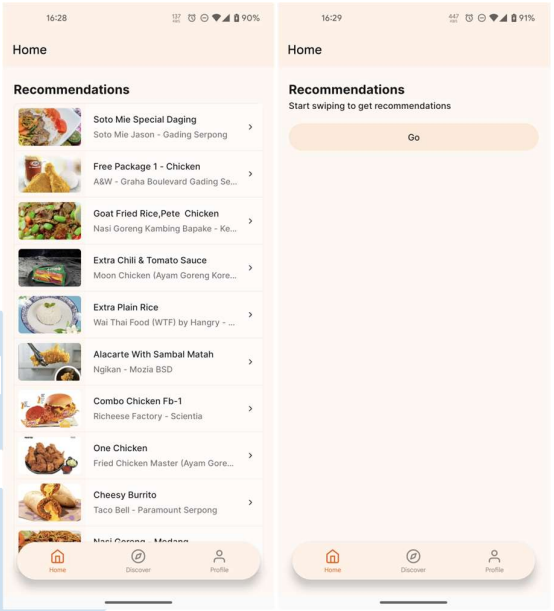


Fig. 8. Implementation Result of Home Page (Recommendations and Empty State)

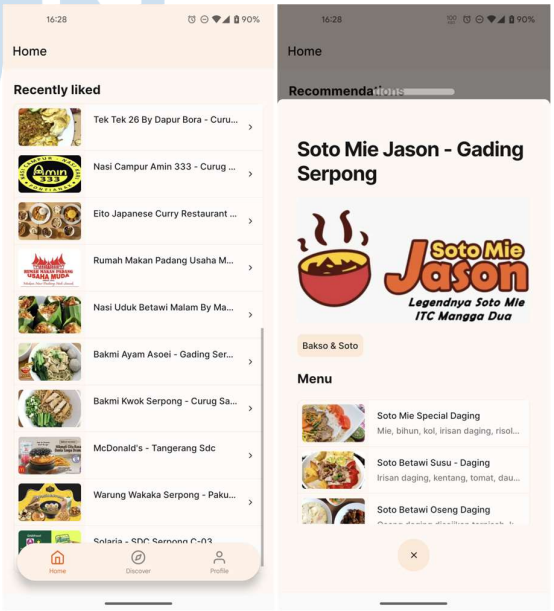


Fig. 9. Implementation Result of Home Page (Recently Liked and Restaurant Details)

Fig. 10. Implementation Result of Discover Page

6. The Profile Page (Fig. 11) provides users with access to their account and recommendation preferences. Here, users can modify parameters such as maximum price, minimum rating, and maximum distance through interactive sliders. It also includes options to reset all previous ratings, reverting preferences to default values, and a logout feature to return to the landing page.

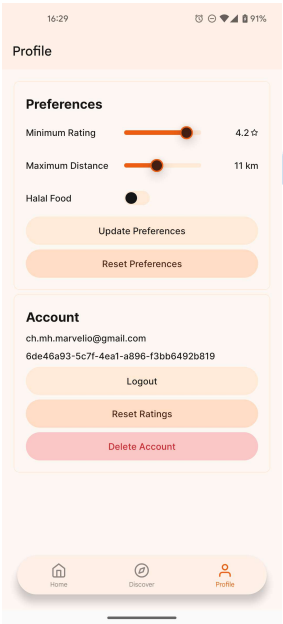


Fig. 11. Implementation Result of Profile Page

The Hybrid Recommendation System Implementation is a core component, demonstrating how the theoretical algorithms are translated into functional code that drives the application's recommendation engine. The Haversine Function (Fig. 12) is implemented to accurately calculate the great-circle distance between two geographical points using their longitude and latitude coordinates, which is essential for location-based filtering.

```
def haversine(lon1, lat1, lon2, lat2):
    lon1, lat1, lon2, lat2 = map(radians, [lon1, lat1, lon2, lat2])
    dlon = lon2 - lon1
    dlat = lat2 - lat1
    a = sin(dlat / 2) ** 2 + cos(lat1) * cos(lat2) * sin(dlon / 2) ** 2
    c = 2 * asin(sqrt(a))
    r = 6371
    return c * r
```

Fig. 12. Code Snippet for Haversine Function

For Collaborative Filtering (CF) Implementation, the system creates a user-restaurant matrix from the collected rating data. Each entry in this matrix indicates whether a user liked a particular restaurant. Subsequently, user similarity is computed using cosine similarity (Fig. 13), a crucial step in identifying users with similar tastes.

```
user_resto_matrix = ratings.pivot(index='user_id', columns='resto_id', values='is_liked').infer_objects(copy=False).fillna(0)
user_similarity = cosine_similarity(user_resto_matrix)
user_similarity = pd.DataFrame(user_similarity, index=user_resto_matrix.index, columns=user_resto_matrix.index)
```

Fig. 13. Code Snippet for Collaborative Filtering

In the Content-Based Filtering (CBF) Implementation, the TfidfVectorizer is employed to transform restaurant tags into a TF-IDF (Term Frequency-Inverse Document Frequency) feature matrix. Restaurant similarity is then calculated based on these processed tags using cosine similarity (Fig. 14), enabling the system to recommend items with similar content.

```
tfidf = TfidfVectorizer()
tfidf_matrix = tfidf.fit_transform(restaurants['combined'])
resto_similarity = cosine_similarity(tfidf_matrix, tfidf_matrix)
resto_similarity = pd.DataFrame(resto_similarity, index=restaurants['resto_id'], columns=restaurants['resto_id'])
```

Fig. 14. Code Snippet for Content-Based Filtering

User Preference Preparation involves defining and preparing user-specific preferences such as liked and disliked restaurants, minimum rating thresholds, maximum distance, and halal/non-halal food preferences (Fig. 15). This is crucial for tailoring recommendations. Filtering Based on Preferences then applies these user-defined criteria to filter the recommended items, ensuring that only relevant options are displayed. This step also ensures that restaurants previously liked or disliked by the user are excluded from the new recommendation list (Fig. 16).

Fig. 15. Code Snippet for User Preference Preparation

Fig. 16. Code Snippet for Preference Filtering

The Score Calculation (CF, CBF, Hybrid) phase (Fig. 17) involves a detailed process where CF scores are derived from user similarity and the ratings provided by other users. CBF scores are computed based on the content similarity between items and items previously liked by the user. Finally, these CF and CBF scores are combined to form a comprehensive hybrid score.

Fig. 17. Code Snippet for Collaborative and Content-Based Filtering Score Calculation

For Recommendation Generation, the system processes the items based on their calculated CF, CBF, and hybrid scores. These recommendations are then sorted by their highest score, and the number of recommendations displayed is limited as per the system's design. This process also includes handling the exclusion of duplicate restaurants or those from the same chain to ensure variety, returning the final recommendations in a structured dictionary format (Fig. 18).

Fig. 18. Code Snippet for Recommendation Results

The System Evaluation quantifies the effectiveness and user satisfaction of the developed restaurant recommendation system. The Evaluation Method employed was the End-User Computing Satisfaction (EUCS) [22], a widely accepted approach for measuring user satisfaction with information systems. EUCS is used to determine how well the system fulfills user requirements by comparing their expectations with the actual system performance. Evaluation Data Collection was conducted by distributing a questionnaire via Google Forms to 27 respondents who had used the application. The questionnaire comprised 10 questions divided into five key EUCS aspects, with responses collected using a 1-5 Likert scale, ranging from "Strongly Disagree" (1) to "Strongly Agree" (5). The specific questions were:

- Content: "Do you think this application provides content and information about restaurants and food that is appropriate?" (P1) and "Do you think the content in this application is clear and easy to understand?" (P2).
- Accuracy: "Do you think the recommendations given by this application are in accordance with your preferences?" (P3) and "Does the navigation in the application lead to the correct page?" (P4).
- Format: "Do you think the application display is attractive?" (P5) and "Do you think the application has an easy-to-understand structure and layout?" (P6).
- Ease of Use: "Do you think this application is easy to use?" (P7) and "Can you easily access all features in this application?" (P8).
- Timeliness: "Does the system display information with a fast response?" (P9) and "Does the application display the latest information?" (P10).

The Satisfaction Score Calculation for each question and overall criterion involved multiplying the frequency of responses by their respective scale weights, dividing by the total number of respondents, and then converting the result to a percentage. This calculation process is exemplified by the following equations:

- Content (P1):

$$\frac{(0\times1)+(0\times2)+(0\times3)+(9\times4)+(18\times5)}{27\times5}\times 100 \qquad (10)$$

- Content (P2):

$$\frac{(0\times1)+(0\times2)+(0\times3)+(6\times4)+(21\times5)}{27\times5}\times 100 \qquad (11)$$

- Percentage Content:

$$\frac{(93.3+95.5)}{2} = 94.4 \qquad (12)$$

Similar calculations were performed for Accuracy, Format, Ease of Use, and Timeliness. The detailed questionnaire results are presented in Table 2. The Results of this evaluation indicated an overall user satisfaction level of 93.9% for the recommendation system, derived from the average of all five criteria.

This high satisfaction rate across all aspects (Content: 94.4%, Accuracy: 93.6%, Format: 92.6%, Ease of Use: 97%, Timeliness: 92.2%) underscores the system's effectiveness in meeting user expectations.

TABLE II. EUCS QUESTIONAIRE RESULT

Question no.	Answer				
	1	2	3	4	5
1	0	0	0	9	18
2	0	0	0	6	21
3	0	0	0	16	11
4	0	0	0	1	26
5	0	0	2	11	14
6	0	0	0	5	22
7	0	0	0	4	23
8	0	0	0	4	23
9	0	0	1	11	15
10	0	0	0	8	19

V. CONCLUSION

This research successfully designed and developed a mobile-based restaurant recommendation system using a hybrid approach that combines Collaborative Filtering (CF) and Content-Based Filtering (CBF). The system was implemented as a mobile app utilizing Expo, React Native, Express, and Flask as its core frameworks and libraries. The evaluation of the developed system, which was conducted using the End-User Computing Satisfaction (EUCS) method, indicated that the system effectively met user expectations across all five key aspects: Content, Accuracy, Format, Ease of Use, and Timeliness. The overall user satisfaction for the system was remarkably high at 93.9%. This high satisfaction rate shows that the design and implementation of the restaurant recommendation system, which integrates hybrid collaborative and content-based filtering methods into a mobile app, was well received by its users.

REFERENCES

[1] A. Szymkowiak, B. Melović, M. Dabić, K. Jeganathan and G. S. Kundi, "Information technology and Gen Z: The role of teachers, the internet, and technology in the education of young people," *Technology in society,* vol. 65, p. 101565, 2021.

[2] R. Williams, "The technology and the society," in *Popular Fiction*, Routledge, 2023, p. 9–22.

[3] M. Ahmadi, S. N. Shahrokhi, M. Khavaninzadeh and J. Alipour, "Development of a mobile-based self-care application for patients with breast cancer-related lymphedema in Iran," *Applied Clinical Informatics,* vol. 13, p. 935–948, 2022.

[4] M. H. Azahari and F. A. H. Ali, "The development of an online food ordering system for JomMakan restaurant," *Applied Information Technology and Computer Science,* vol. 3, p. 369–376, 2022.

[5] Y. Bohang, "HUBUNGAN ANTARA MOTIVASI MEMBELI DAN PERILAKU PEMESANAN MAKANAN MELALUI APLIKASI ONLINE," 2020.

[6] S. G. Pillai, W. G. Kim, K. Haldorai and H.-S. Kim, "Online food delivery services and consumers' purchase intention: Integration of theory of planned behavior, theory of perceived risk, and the elaboration likelihood model," *International journal of hospitality management,* vol. 105, p. 103275, 2022.

[7] N. Thongsri, P. Warintarawej, S. Chotkaew and W. Saetang, "Implementation of a personalized food recommendation system based on collaborative filtering and knapsack method," *Int. J. Electr. Comput. Eng,* vol. 12, p. 630–638, 2022.

[8] U. Javed, K. Shaukat, I. A. Hameed, F. Iqbal, T. M. Alam and S. Luo, "A review of content-based and context-based recommendation systems," *International Journal of Emerging Technologies in Learning (iJET),* vol. 16, p. 274–306, 2021.

[9] R. Widayanti, M. H. R. Chakim, C. Lukita, U. Rahardja and N. Lutfiani, "Improving recommender systems using hybrid techniques of collaborative filtering and content-based filtering," *Journal of Applied Data Sciences,* vol. 4, p. 289–302, 2023.

[10] G. Parthasarathy and S. Sathiya Devi, "Hybrid recommendation system based on collaborative and content-based filtering," *Cybernetics and Systems,* vol. 54, p. 432–453, 2023.

[11] L. Li, Z. Zhang and S. Zhang, "Hybrid algorithm based on content and collaborative filtering in recommendation system optimization and simulation," *Scientific Programming,* vol. 2021, p. 7427409, 2021.

[12] C. Janiesch, P. Zschech and K. Heinrich, "Machine learning and deep learning," *Electronic markets,* vol. 31, p. 685–695, 2021.

[13] Q. Zhang, J. Lu and Y. Jin, "Artificial intelligence in recommender systems," *Complex & Intelligent Systems,* vol. 7, p. 439–457, 2021.

[14] S. M. Pande, P. K. Ramesh, A. Anmol, B. R. Aishwarya, K. Rohilla and K. Shaurya, "Crop recommender system using machine learning approach," in *2021 5th international conference on computing methodologies and communication (ICCMC)*, 2021.

[15] F. Ricci, L. Rokach and B. Shapira, "Recommender systems: Techniques, applications, and challenges," *Recommender systems handbook,* p. 1–35, 2021.

[16] H. Papadakis, A. Papagrigoriou, C. Panagiotakis, E. Kosmas and P. Fragopoulou, "Collaborative filtering recommender systems taxonomy," *Knowledge and Information Systems,* vol. 64, p. 35–74, 2022.

[17] L. Wu, X. He, X. Wang, K. Zhang and M. Wang, "A survey on accuracy-oriented neural recommendation: From collaborative filtering to information-rich recommendation," *IEEE Transactions on Knowledge and Data Engineering,* vol. 35, p. 4425–4445, 2022.

[18] K. Mao, J. Zhu, J. Wang, Q. Dai, Z. Dong, X. Xiao and X. He, "SimpleX: A simple and strong baseline for collaborative filtering," in *Proceedings of the 30th ACM international conference on information & knowledge management*, 2021.

[19] S. Mohmmad, R. Kanakam, R. Dadi, Shabana and S. N. Pasha, "Content and history based movie recommendation system," in *AIP Conference Proceedings*, 2022.

[20] K. Denistia, E. Shafaei-Bajestan and R. H. Baayen, "Exploring semantic differences between the Indonesian prefixes PE-and PEN-using a vector space model," *Corpus Linguistics and Linguistic Theory,* vol. 18, p. 573–598, 2022.

[21] A. Nazir, R. N. Mir and S. Qureshi, "Idea plagiarism detection with recurrent neural networks and vector space model," *International Journal of Intelligent Computing and Cybernetics,* vol. 14, p. 321–332, 2021.

[22] X. Cai, C. Huang, L. Xia and X. Ren, "LightGCL: Simple yet effective graph contrastive learning for recommendation," *arXiv preprint arXiv:2302.08191,* 2023.

[23] P. J. Schmid, "Dynamic mode decomposition and its variants," *Annual Review of Fluid Mechanics,* vol. 54, p. 225–254, 2022.

[24] R. M. Maisari, M. N. Alamsyah and L. Sunardi, "Analisis pengukuran tingkat kepuasan pengguna aplikasi OVO menggunakan metode End User Computing Satisfaction (EUCS)," *ESCAF,* p. 1405–1414, 2024.

[25] W. J. Doll and G. Torkzadeh, "The measurement of end-user computing satisfaction," *MIS quarterly,* p. 259–274, 1988.

ISSN 2355-0082

Monte Carlo Algorithm Applications in Shrimp Farming: Monitoring Systems and Feed Optimization

Vincentius Kurniawan[1], Atanasius Raditya Herkristito[2], Maria Irmina Prasetiyowati[3],
[1,2,3]Dept. of Informatics: Universitas Multimedia Nusantara, Jakarta, Indonesia
vincentius.kurniawan@lecturer.umn.ac.id[1], atanasius@student.umn.ac.id[2], maria@umn.ac.id[3]

Accepted 2 July 2025
Approved 20 February 2022

***Abstract*— Indonesia is one of the world's leading maritime nations, ranking second in fishery export value in 2020. Shrimp stands out as the most lucrative commodity, with an export value of USD 1,997.49 million. Feeding shrimp plays a vital role in their growth and cultivation; however, overfeeding can result in feed residue that negatively impacts the quality of pond water and represents the biggest operational after capital expenditure. The profitability of shrimp farming heavily depends on the feeding cost. This study uses the Monte Carlo algorithm to track feed in shrimp and provides an optimal feeding plan. The algorithm can be used to provide feed recommendations for shrimp start from 33 days of cultivation (DoC), with a best range around 85kg to 92kg. The findings show the potential of the Monte Carlo algorithm in enhancing feeding plans in shrimp farming industries.**

***Index Terms*— Cultivation; Feeding; Feed Recommendations; Margin; Monte Carlo; Operational Cost; Pond Water; Shrimp Farming**.

I. INTRODUCTION

Shrimp farming plays a crucial role in the aquaculture's global industry. It has made a big contribution to the seafood market. As one of the fastest growing industries, shrimp farming drives large economic value around the world, with exports reaching into billions of dollars annually [1]. Indonesia, as a maritime country, holds an important position in the global shrimp market. With the longest coastline and favorable climatic conditions, Indonesia able to capitalize on shrimp aquaculture to become one of the top exporters of shrimp. In 2023, the industry contributed USD 2.2 billion in export value, showing its importance to Indonesia's economy [2]. This growth is driven by increasing global demand, particularly for high-value shrimp species, and Indonesia's commitment to improving shrimp farming practices. However, shrimp farmers are still facing the continuous challenge of increasing productivity and cost-effectiveness.

Litopenaeus vannamei is also known as whiteleg shrimp. It is one of the most popular shrimp species in Indonesia, due to the relatively lower growth period and disease resistance compared to other species [3]. Shrimp uses its sensory organs and appendages to detect food particles dispersed in the water. They scrape or grasp feed particles, more largely by the pereiopods (walking legs). Overfeeding, however, can be both a source of water contamination and an increase in operational cost because feeding that is not utilized is simply wasted. Therefore, controlled feeding is very vital to create an economic-ecological balance with respect to maximized efficiency of feed utilization and water quality maintenance in shrimp culture [4].

Shrimp feed plan and estimation is usually done based on mathematical approaches. Two commonly used techniques to estimate the amount of feed are index feeding and feed rate methods, both with their disadvantages and advantages. In practical situations, factors like water quality, temperature, pH, salinity, and alkalinity certain from environmental causes must be considered by the shrimp farmer, since these play a significant role in influencing shrimp appetite [5]. Whenever environmental parameters are kept to their optimum level, the shrimp shows healthy appetite, growth and development [6].

Although index feeding and feed rate methods are commonly used, most systems still overlook daily environmental changes as well as biological variations in shrimp behavior. Furthermore, even though automatic feeding systems and biofloc-based approaches have potential [7], their combination with probabilistic forecasting models is still underexplored. There is limited research on short-term feed prediction models that following to changes in shrimp biomass and survival rate that can be extracted based on the combination of index and feeding rate calculation.

Monte Carlo algorithm can be used to detemine an optimal shrimp feed calculations. The Monte Carlo algorithm provides a method for forecast of shrimp feed quantities. While other approach like genetic algorithms are better for long-term optimization of feed projection over an extended period of time, Monte Carlo is more suitable for short-range forecast due to its transparency and flexibility [8]. The other approach like regression models are straightforward, but they lack the

ability to adjust in real-time for complex, adaptive dynamic systems [6]. Monte Carlo enables processing random data and an ideal approach since it decreases uncertainty via simulation [9]. In shrimp farming, statistical data for feed calculation changes daily. These factors include such as average shrimp weight, biomass, survival rate, density, water conditions, pH levels, and salinity.

Based on those aforementioned factors, a monitoring and feed calculation system was developed using the Monte Carlo algorithm. The aim of this research is to develop and validate a Monte Carlo simulation system for estimating feed quantities of L. vannamei, intended to help farmers to make informed decisions based on changing conditions, enhance feed efficiency, and reduce operational costs.

II. LITERATURE REVIEW

A. Shrimp Farming and Feeding Methodologies

A particular feeding practice is required in order to achieve optimal growth. Blind feeding method is typically used for the first 30 to 35 days of cultivation. It is a feed method where feeding amount is given without considering shrimp biomass. This method gives shrimp fry the right nutrition in their early stages of growth. However, if it is not managed properly, it could cause overfeeding and deteriorate the quality of the water [10]. As shrimp grow and gain more weight, feeding methods use calculated approach such as index feeding or feed rate feeding. Index feeding method is formulated based on calculations on shrimp biomass, survival rate, and average weight, whereas feeding rate follows a set value established through research and experience in shrimp culture [7]. Both methods have a similar goal which is optimized feed utilization. In order to maintain shrimp appetite and growth, both approaches would still require close observation of environmental factors like temperature, pH, salinity, and water quality [11].

The index feeding method for shrimp uses percentage indexing calculation. The method considers many parameters, including the age of shrimp in days of cultivation (DoC), the quantity of shrimp in one pond, and the feed index. Before applying this technique, farmers need to determine the goal of average daily growth (ADG) of the shrimp. After the ADG number is decided, it is possible to calculate the feed index percentage according to an equation (1) [6].

$$Index\ (\%) = 2\ \times ADG\ \times 100\% \quad (1)$$

Once the index is decided, the feed quantity can be calculated using the equation (2).

$$Feed(Kg) = \frac{Index(\%) \times DoC \times Num.of\ Larvae}{1000} \quad (2)$$

Meanwhile, the feed rate method calculates shrimp feed based on the average body weight (ABW) and the biomass of the shrimp. The Feed Rate (FR) table from previous observation is used as a reference for this feed rate method. Before determining the feed quantity using this method, farmers need to do sampling to obtain the ABW value. When the ABW number is identified from the sampling process, they can calculate the shrimp biomass using equation (3) [6].

$$Biomass = ABW\ \times Population \quad (3)$$

After both the ABW and biomass values have been determined, the feed amount can be calculated using the FR Feeding method and its corresponding formula (4).

$$Feed = Biomass \times FR(\%) \quad (4)$$

B. Monte Carlo Simulation

Monte Carlo Simulation is a method used to demonstrate how sample data in the simulation may be applied and forecast its distribution. This simulation is conducted using the previous farming and feeding log data. The main idea of the Monte Carlo Simulation is to create a model variable and derive its value from the other variables that are analyzed. Monte Carlo Simulation enables calculating the average feed quantity recommendations and the standard deviation of feed amount [9].

The formula to calculate the average feed quantity is described in equation (5).

$$\mu = \frac{1}{n}\sum_{i=1}^{n} R_i \quad (5)$$

μ represents the mean obtained from the Monte Carlo Simulation. σ represents the standard deviation derived from the Monte Carlo Simulation. n indicates the number of Monte Carlo simulations to be conducted. Ri is the feed quantity recommendation for the i-th iteration.

Equation (6) is used to calculate the standar deviation of the feed quantity.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_i - \mu)^2} \quad (6)$$

After obtaining the average and standard deviation, the next step is to calculate the confidence interval. Before calculating the confidence interval, it is necessary to define the confidence level and then identify the z-score corresponding to the chosen confidence level using the z-table. The formula for calculating the confidence interval is described in equation (7).

$$CI = \mu\ \pm z\ \times \sigma \quad (7)$$

σ represents the standard deviation derived from the Monte Carlo Simulation. z represents the z-score based on the selected confidence level, which can be found in the z-table [12].

III. DESIGN AND ANALYSIS

The business process analysis for the system module designed to monitor and calculate shrimp feed using the Monte Carlo algorithm is divided into two

aspects. The first aspect addresses input-output data requirements. Input data for the system includes user and role information for system access, master data encompassing suppliers, supplier types, ponds, clusters, feeding rate presets, and feeding rate parameters, as well as feeding entries and active pond data.

The second aspect focuses on functionality. Traditionally, feed calculations are managed using whiteboards and Google Spreadsheets. The system is expected to handle more complex feed recommendation calculations while replacing manual recording methods. Based on the identified needs, a web-based system can provide enhanced functionality for managing complex feed calculations and serve as an effective substitute for manual data tracking.

The system development process has adopted the Monte Carlo algorithm as the primary method for calculating feed recommendations. The web-based system utilizes the ReactJS framework for the frontend and Django for the backend. The Monte Carlo algorithm calculates feed recommendations on the backend, sending the results to the frontend via Application Programming Interface (API)s.

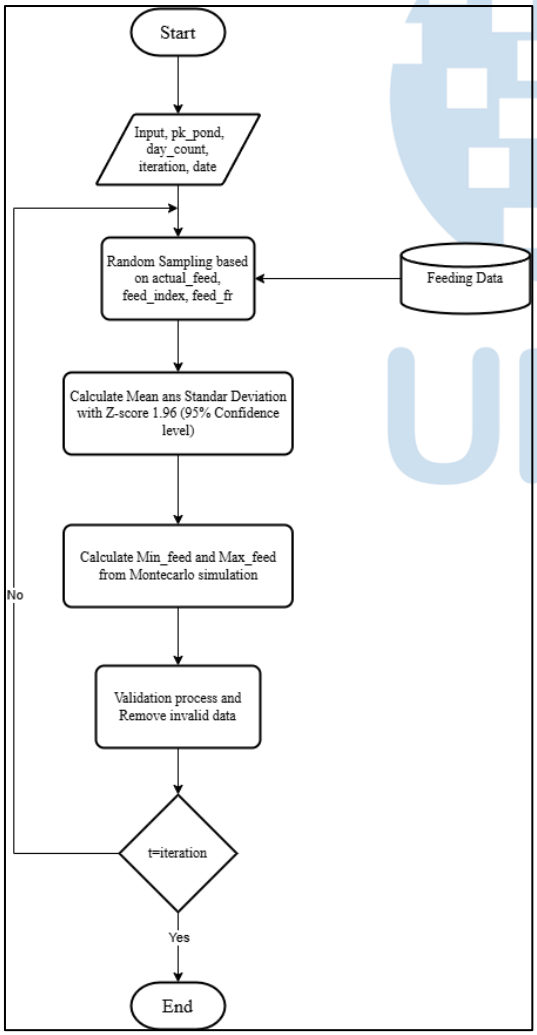


Fig 1. Flowchart of Monte Carlo Simulation

Figure 1 flowchart illustrates the process of analyzing the feeding data in Monte Carlo simulation. The process starts by collecting input parameters such as pond identification (*pk_pond*), cultivation days (*day_count*), iteration count, and reference date. Then it takes a random sample of the feeding data such as including feed quantities, feed index values, and feed rate (FR) number. From the obtain sample, the mean and standard can be calculated. For establishing a confidence level of 95%, a Z-score of 1.96 is used which ensures reliability from a statistical perspective. To support the simulation flowchart illustrated in Figure 1, Table 1 presents the schema of the feeding data utilized during the Monte Carlo analysis.

TABLE I. FEEDING DATA TABLE SCHEMA

Field	Datatype
id	uuid
active_pond_id	uuid
actual_feeding_weight	float
doc	integer
calculated_index	float
feeding_rate_preset_id	uuid
created_by	uuid
created_at	timestamp
updated_at	timestamp
deleted_at	timestamp

The feeding data schema in Table 1 includes a unique identifier `id` for each record, ensuring traceability across datasets. It associates each entry with an `active_pond_id`, which links the feeding activity to a specific pond for localized analysis. The `actual_feeding_weight` field records the total feed dispensed in kilograms, which is used for evaluating efficiency and controlling waste. The doc, or days of culture, tracks the shrimp's age since stocking and is used to adjust feed rates over time. `calculated_index` reflects a derived metric indicating feed per biomass unit, helping checks how much feed is given relative to shrimp growth. The `feeding_rate_preset_id` links to standardized feeding plans, allowing comparison and control across different cultivation setups. Each record is tagged with `created_by` to identify the user or system that logged the data, while timestamps like `created_at`, `updated_at`, and `deleted_at` support auditing, version control, and soft deletion for historical tracking and recovery.

This is followed by determining the minimum and maximum feed rates by the simulator. Invalid data is then filtered out in a validation process. The number of iterations is checked to decide if the loop is continued or end the calculation process. This method is applied to estimate feeding values: Table 1 Feeding recommendations derived from statistical analysis Results gained on the basis of Monte Carlo simulation.

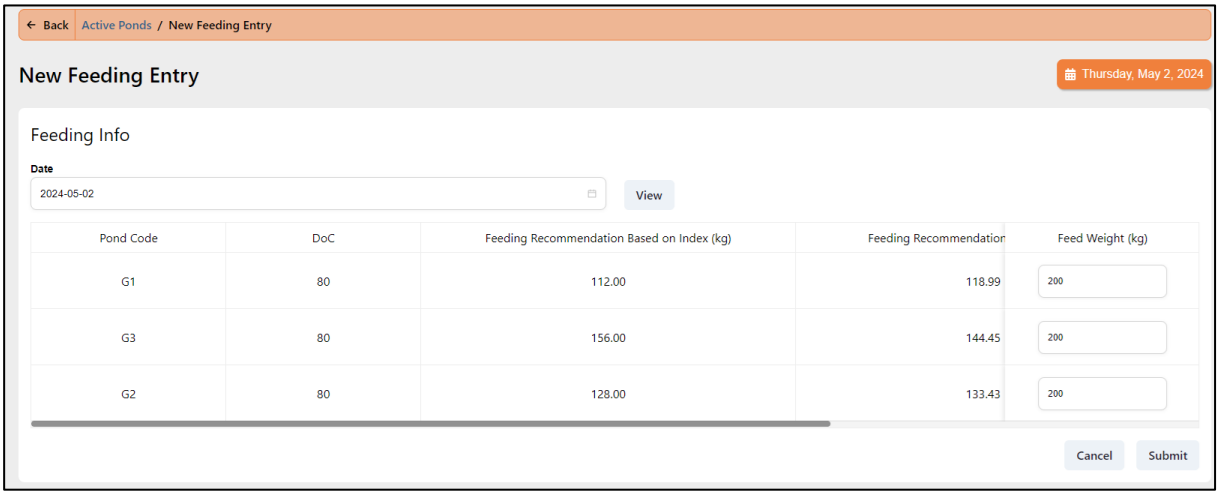


Fig 2. Feeding Entry Page

IV. Implementation and Testing

A testing and evaluation process was conducted after the system had been developed. It has a purpose to determine the accuracy and assess the functionality of the system. The evaluation was divided into two stages: functional testing is used to examine system usability and performance. The second stage is validation testing, which focuses on verifying feed recommendation calculations generated using the Monte Carlo algorithm. These assessments have a purpose to make sure that the system operates well function and provides accurate feed recommendations that are consistent with predefined models.

Figure 2 displays the interface of the feeding entries page, which is used as the input source for the Monte Carlo simulation. Within this page, users can select the specific pond and enter the actual feed quantity on a given Day of Cultivation (DoC). Once the feeding data is submitted, the system uses it as part of the historical dataset to generate projected feed requirements for the next day using Monte Carlo simulation.

At this study, the feed recommendation calculation was tested on a pond labeled H1. It started provide recommended feed quantity for Day of Cultivation (DoC) 32 or before the first shrimp sampling was done. The testing process used feed data from DoC 21 to DoC 31. The confidence level set at 95%, and simulations run with iteration sets of 30, 100, 10,000, and 100,000.

This dataset was chose because feeding during DoC 1 to DoC 20 the feeding process still followed a blind feeding approach. It serves to introduce artificial feed to the shrimp and does not using index or feed rate (FR) methods yet. The minimum iteration count used in this study was based on Sugiyono's suggestion that statistical testing should involve at least 30 iterations [13]. The feed distribution and recommended feeding values based on index and FR methods for pond H1 from DoC 21 to DoC 31 are presented in Table 2.

TABLE II. DATA FEEDING RECOMMENDATION OF DoC 21 - 31

DoC	Actual Feed (Kg)	Recommendation by Index (Kg)	Recommendation by FR (Kg)
21	67	58.8	58.5
22	63	61.6	61.42
23	69	64.4	64.35
24	72	67.2	67.28
25	75	70	70.2
26	78	72.8	73.12
27	81	75.6	76.05
28	84	78.4	78.98
29	87	81.2	81.9
30	90	84	84.82
31	87	86.8	87.75

The data presented in Table .1 was obtained from the developed system. This data was then categorized into multiple scenarios, varying in the number of iterations and the duration of historical feed data used. The number of days refers to the span of feeding data considered prior to Day of Cultivation (DoC) 32 for simulation purposes, while the iteration count represents the number of simulations performed per day. The results of the testing process are detailed as follows.

A. 30 Iteration of Monte Carlo Simulation

The initial testing phase involved running simulations 30 times for each designated number of days. In total, 330 simulations were conducted under this scenario. Upon completing all simulations, feed recommendations were obtained along with the

corresponding confidence interval, as presented in Table 3.

TABLE III. RESULT OF 30 ITERATION

Total Days	Minimum Feed (Kg)	Maximum Feed (Kg)
1	87.5000	87.5000
2	83.4376	86.3624
3	80.1359	86.6641
4	76.9458	87.7209
5	75.1825	88.0842
6	72.4937	89.1063
7	68.9931	88.6735
8	65.9086	89.2248
9	61.7309	88.0024
10	63.3541	88.0459
11	63.2646	87.8021

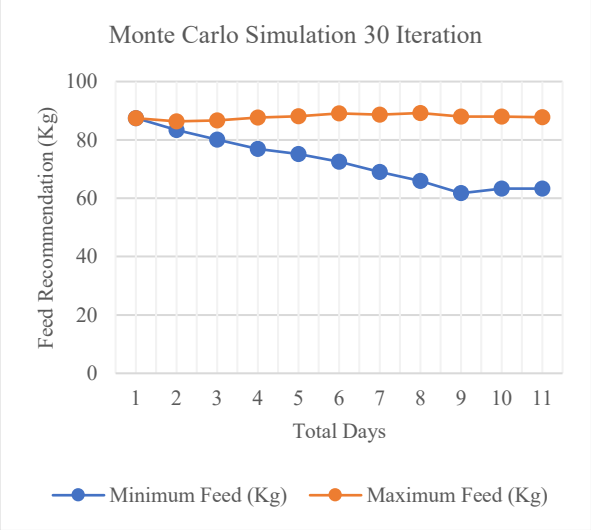


Fig 2. Line Chart of Monte Carlo 30 Iteration

A line chart was generated to facilitate the visualization of feed recommendation results from the Monte Carlo simulation, as shown in Figure 3. The chart illustrates that as the number of days considered increases, the range of recommended feed quantities expands.

Following this visualization, the next step involved comparing the simulated feed recommendations with the actual feed recorded data in Google Spreadsheets. This comparison helps identify the scenario in which the Monte Carlo simulation produces the closest feed recommendation range to the actual feed quantity provided. On Day of Cultivation (DoC) 32, the actual feed quantity administered to shrimp in pond H1 as recorded in the Google Spreadsheets was 86 kg. To generate a feed recommendation using the Monte Carlo simulation, input data from the previous 31 days (DoC 1 to 31) were used from feeding entries in Feeding Data table. This included the historical feeding plan generated using index and feeding rate approach in kilograms. Then later the simulation processed these inputs across 30 randomized iterations, producing a projected feeding plan in range between 83.4376 kg and 86.3623 kg for DoC 32. These results demonstrate a reasonable approximation to previously recorded feeding value, indicating that the simulation can produce statistically supported recommendations that reflect actual feeding behavior in the shrimp farm under dynamic cultivation conditions.

B. 100 Iteration of Monte Carlo Simulation

After completing the initial testing with 30 simulations, the process was extended to 100 simulations per day. In total, 1,100 simulations were conducted under this scenario. Following the completion of all simulations, feed recommendations were obtained along with their respective confidence intervals, as presented in Table 4.

The line chart in Figure 4 illustrates a trend similar to the previous simulation, where increasing the number of days considered results in a wider range of feed recommendations. However, the chart appears more refined compared to the earlier simulation due to the higher number of iterations, which contributes to a smoother representation of the data.

Following the visualization of the generated range, the next step involved comparing the simulated feed recommendations with actual feeding data recorded in Google Spreadsheets. This comparison helps identify the scenario in which the Monte Carlo simulation produces a feed recommendation range closest to the actual feed quantity administered. According to actual data, the feed quantity provided to shrimp in pond H1 on Day of Cultivation (DoC) 32 was 86 kg. The Monte Carlo simulation scenario with 100 iterations that yielded the closest recommendation considered feeding data from the two days prior to DoC 32, producing a range between 83.6745 kg and 86.6455 kg.

TABLE IV. RESULT OF 100 ITERATION

Total Days	Minimum Feed (Kg)	Maximum Feed (Kg)
1	87.5	87.5
2	83.6745	86.6455
3	80.0827	86.7373
4	77.0679	88.1921
5	75.7052	88.0348
6	71.7429	87.9971
7	69.3862	88.5538
8	66.6448	88.5952
9	63.5393	89.0807
10	61.635	88.1249
11	57.6148	91.4652

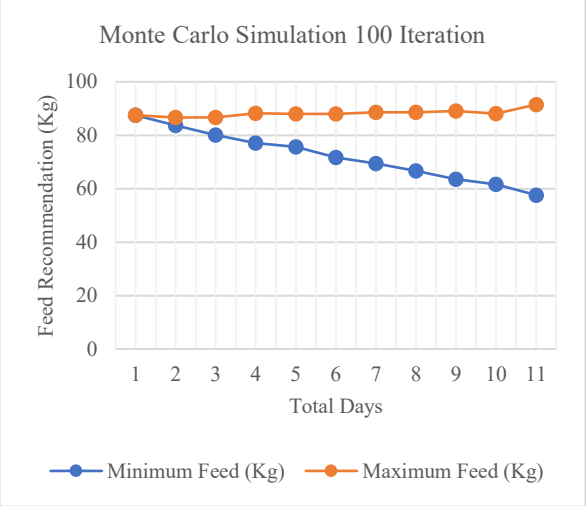


Fig 3. Line Chart of Monte Carlo 100 Iteration

C. *Testing Result Summary*

After conducting all Monte Carlo algorithm testing scenarios, the feed recommendation data that closely aligns with actual feed records from Google Spreadsheets was identified. The complete results are presented in Table 5.

The findings indicate that the Monte Carlo algorithm is applicable for feed recommendation calculations. The scenario that provides the closest recommendation range is based on data from the two days preceding the targeted Day of Cultivation (DoC), with a minimal difference of approximately 3 kg between the minimum and maximum feed recommendations. Additionally, an iteration counts of at least 10,000 ensures stability in the variation of feed recommendation values.

TABLE V. SUMMARY RESULT OF MONTE CARLO SIMULATION

Minimum Feed (kg)	Maximum Feed (kg)	Days of Data	Iterations
83.4376	86.3624	2	30
83.6745	86.6455	2	100
83.4131	86.5953	2	10,000
83.4007	86.598	2	100,000

V. CONCLUSION

Based on this research, it can be concluded that the system developed and implemented using the Monte Carlo algorithm has been successfully designed and built. The system calculates feed recommendations using Monte Carlo based on historical feeding data and provides recommendations using both the index and feed rate (FR) methods.

During Monte Carlo testing, simulations were carried out under various scenarios. Feed recommendations calculated using historical data from the previous two days produced results close to the actual feed quantity, ranging between 83 kg and 86 kg, compared to the actual value of 86 kg. Additionally, the confidence interval showed stable results across iteration ranges from 10,000 to 100,000, as larger iterations incorporated more data variability. Conversely, when iterations fell below 10,000, the results became inconsistent due to limited data variations. Meanwhile, iterations above 100,000 did not yield significantly different results compared to 100,000 iterations, only increasing computational load without improving the accuracy of feed recommendations.

Despite its effectiveness, the simulation system has certain limitations, especially its reliance on manual user input during daily feeding operations. Incorrect entries can lead to inaccurate feed projections, affecting decision-making and system reliability. Additionally, the model is currently dependent on index-based and feed rate calculations derived from historical data, which may not fully capture real-time changes in shrimp behavior or environmental conditions. To achieve more precise feed recommendations, future work should explore the integration of real-time sensor technologies such as automated feeding systems, water temperature monitoring, and pH detection sensors. Therefore allowing the model to dynamically adjust its projections based on current aquaculture conditions.

ACKNOWLEDGMENT

PT Samudra provided support for this study. A special thank you to George Samuel for contributing his expertise to this study's findings.

REFERENCES

[1] D. Jory, "Annual farmed shrimp production survey: A slight decrease in production reduction in 2023 with hopes for renewed growth in 2024," 9 October 2023. [Online]. Available: https://www.globalseafood.org/advocate/annual-farmed-shrimp-production-survey-a-slight-decrease-in-production-reduction-in-2023-with-hopes-for-renewed-growth-in-2024/. [Accessed 1 April 2025].

[2] Algo Research, "Indonesia Shrimp Sector Outlook," Algo Research, Jakarta, 2024.

[3] P. Pratiwi, M. Marzuki and B.D.H.Setyono, "Growth and survival rate of vaname shrimp (litopenaeus vannamei) pl-10 on different stocking density," *AQUASAINS,* vol. 9, 2021.

[4] Y. R. Alfiansah and A. G¨ardes, "Water quality and bacterial community management in shrimp ponds in rembang, indonesia : Towards sustainable shrimp aquaculture," *Policy Brief 2019 / 2,* vol. 2, pp. 1-5, 2019.

[5] A. Achmad, D. Susiloningtyas and T. Handayani, "Sustainable aquaculture management of vanamei shrimp (liptopenaeus vannamei) in batukaras village, pangandaran, indonesia," *International Journal of GEOMATE,,* vol. 19, no. 1, 2020.

[6] B. Madusari, H. Ariadi and D. Mardhiyana, "Effect of the feeding rate practice on the white shrimp (litopenaeus vannamei) cultivation activities," *AACL Bioflux,* vol. 15, no. 1, 2022.

[7] A. Kusmiatun, D. A. S. Utami, T. Firnaenil, Y. E. Kaborang, T. Harijono, S. Tangguda, M. S. Triyastuti, R. Djauhari, U. Tantulo and M. A. Sihombing, "Effects of Feeding Rate Reduction on the Growth Performance and Feed Utilization of Pacific White Shrimp Reared Using Biofloc System," *Journal of Aquaculture Research,* vol. 12, no. 3, pp. 205-213, 2025.

[8] M. Luna, I. Llorente and A. Cobo, "Determination of feeding strategies in aquaculture farms using a multiple-criteria approach and genetic algorithms," *Annals of Operations Research,* vol. 314, pp. 551-576, 2022.

[9] A. Mallongi, A. U. Rauf, A. Daud, M. Hatta, W. Al-Madhoun and R. Amiruddin, "Health risk assessment of potentially toxic elements in maros karst groundwater: a monte carlo simulation approach," *Geomatics, Natural Hazards and Risk,,* vol. 13, 2022.

[10] M. Fahrur, R. Syah, Makmur, H. S. Suwoyo, A. I. J. Asaad and I. Taukhid, "The Effect of Blind feeding to Increase Post Larva Growth of High-density Vaname Shrimp and Their Effect on Water Quality," *Berita Biologi,* vol. 22, no. 3, pp. 123-135, 2023.

[11] H. Ariadi, A. Wafi, Supriatna and M. Musa, "Oxygen Diffusion Rate During Blind Feeding Period of Intensive Vaname Shrimp Culture," *REKAYASA,* vol. 14, no. 2, pp. 107-115, 2024.

[12] K. K. Sabelfeld, "Monte Carlo Simulation: Concepts, Algorithms, and Applications," *Simulation Modelling Practice and Theory,* vol. 31, no. 1, 1995.

[13] S. Prof. Dr, Metode Penelitian Kuantitatif, Kualitatif dan R&D, Bandung: Alfabeta, 2016.

[14] Indonesia Research Institute Japan Co.,Ltd, "Indonesia Fisheries Market: A Brief Introduction," Indonesia Research Institute Japan Co.,Ltd, Jakarta, 2024.

[15] Adwa instruments, "The Effects of Overfeeding in an Aquariums: Changes in Water Quality and How to Fix Them," [Online]. Available: https://www.adwainstruments.com/blog/246-the-effects-of-overfeeding-in-an-aquariums-changes-in-water-quality-and-how-to-fix-them. [Accessed 22 April 2025].

[16] www.shrimpinsights.com, "Indonesia's Exports on Par with 2020 Shows Real-Time Data," 27 October 2021. [Online]. Available: https://www.shrimpinsights.com/blog/indonesias-exports-par-2020-shows-real-time-data. [Accessed 22 April 2025].

AUTHOR GUIDELINES

1. Manuscript criteria

- The article has never been published or in the submission process on other publications.
- Submitted articles could be original research articles or technical notes.
- The similarity score from plagiarism checker software such as Turnitin is 20% maximum.

2. Manuscript format

- Article been type in Microsoft Word version 2007 or later.
- Article been typed with 1 line spacing on an A4 paper size (21 cm x 29,7 cm), top-left margin are 3 cm and bottom-right margin are 2 cm, and Times New Roman's font type.
- Article should be prepared according to the following author guidelines in this template. Article contain of minimum 3500 words.
- References contain of minimum 15 references (primary references) from reputable journals/conferences

3. Organization of submitted article

The organization of the submitted article consists of Title, Abstract, Index Terms, Introduction, Method, Result and Discussion, Conclusion, Appendix (if any), Acknowledgment (if any), and References.

- Title
  The maximum words count on the title is 12 words (including the subtitle if available)
- Abstract
  Abstract consists of 150-250 words. The abstract should contain logical argumentation of the research taken, problem-solving methodology, research results, and a brief conclusion.
- Index terms
  A list in alphabetical order in between 4 to 6 words or short phrases separated by a semicolon (;), excluding words used in the title and chosen carefully to reflect the precise content of the paper.
- Introduction
  Introduction commonly contains the background, purpose of the research, problem identification, research methodology, and state of the art conducted by the authors which describe implicitly.
- Method
  Include sufficient details for the work to be repeated. Where specific equipment and materials are named, the manufacturer's details (name, city and country) should be given so that readers can trace specifications by contacting the manufacturer. Where commercially available software has been used, details of the supplier should be given in brackets or the reference given in full in the reference list.
- Results and Discussion
  State the results of experimental or modeling work, drawing attention to important details in tables and figures, and discuss them intensively by comparing and/or citing other references.
- Conclusion
  Explicitly describes the research's results been taken. Future works or suggestion could be explained after it
- Appendix and acknowledgment, if available, could be placed after Conclusion.
- All citations in the article should be written on References consecutively based on its' appearance order in the article using Mendeley (recommendation). The typing format will be in the same format as the IEEE journals and transaction format.

4. Reviewing of Manuscripts

Every submitted paper is independently and blindly reviewed by at least two peer-reviewers. The decision for publication, amendment, or rejection is based upon their reports/recommendations. If two or more reviewers consider a manuscript unsuitable for publication in this journal, a statement explaining the basis for the decision will be sent to the authors within six months of the submission date.

5. Revision of Manuscripts

Manuscripts sent back to the authors for revision should be returned to the editor without delay (maximum of two weeks). Revised manuscripts can be sent to the editorial office through the same online system. Revised manuscripts returned later than one month will be considered as new submissions.

6. Editing References

- **Periodicals**
  J.K. Author, "Name of paper," Abbrev. Title

of Periodical, vol. x, no. x, pp. xxx-xxx, Sept. 2013.

- **Book**
  J.K. Author, "Title of chapter in the book," in Title of His Published Book, xth ed. City of Publisher, Country or Nation: Abbrev. Of Publisher, year, ch. x, sec. x, pp xxx-xxx.

- **Report**
  J.K. Author, "Title of report," Abbrev. Name of Co., City of Co., Abbrev. State, Rep. xxx, year.

- **Handbook**
  Name of Manual/ Handbook, x ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. xxx-xxx.

- **Published Conference Proceedings**
  J.K. Author, "Title of paper," in Unabbreviated Name of Conf., City of Conf., Abbrev. State (if given), year, pp. xxx-xxx.

- **Papers Presented at Conferences**
  J.K. Author, "Title of paper," presented at the Unabbrev. Name of Conf., City of Conf., Abbrev. State, year.

- **Patents**
  J.K. Author, "Title of patent," US. Patent xxxxxxx, Abbrev. 01 January 2014.

- **Theses and Dissertations**
  J.K. Author, "Title of thesis," M.Sc. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year. J.K. Author, "Title of dissertation," Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

- **Unpublished**
  J.K. Author, "Title of paper," unpublished. J.K. Author, "Title of paper," Abbrev. Title of Journal, in press.

- **On-line Sources**
  J.K. Author. (year, month day). Title (edition) [Type of medium]. Available: http:// www.(URL) J.K. Author. (year, month). Title. Journal [Type of medium]. volume(issue), pp. if given. Available: http://www.(URL) Note: type of medium could be online media, CD-ROM, USB, etc.

7. Editorial Adress

Universitas Multimedia Nusantara
Jl. Scientia Boulevard, Gading Serpong
Tangerang, Banten, 15811
Email: **ultimaijnmt@umn.ac.id**

Paper Title

Subtitle (if needed)

Author 1 Name[1], Author 2 Name[2], Author 3 Name[2]
[1] Line 1 (of affiliation): dept. name of organization, organization name, City, Country
Line 2: e-mail address if desired
[2] Line 1 (of affiliation): dept. name of organization, organization name, City, Country
Line 2: e-mail address if desired

Accepted on mmmmm dd, yyyy
Approved on mmmmm dd, yyyy

***Abstract*—This electronic document is a "live" template which you can use on preparing your IJNMT paper. Use this document as a template if you are using Microsoft Word 2007 or later. Otherwise, use this document as an instruction set. Do not use symbol, special characters, or Math in Paper Title and Abstract. Do not cite references in the abstract.**

***Index Terms*—enter key words or phrases in alphabetical order, separated by semicolon (;)**

I. INTRODUCTION

This template, modified in MS Word 2007 and saved as a Word 97-2003 document, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. Margins, column widths, line spacing, and type styles are built-in here. The authors must make sure that their paper has fulfilled all the formatting stated here.

Introduction commonly contains the background, purpose of the research, problem identification, and research methodology conducted by the authors which been describe implicitly. Except for Introduction and Conclusion, other chapter's title must be explicitly represent the content of the chapter.

II. EASE OF USE

A. Selecting a Template

First, confirm that you have the correct template for your paper size. This template is for IJNMT. It has been tailored for output on the A4 paper size.

B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them.

III. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not add any kind of pagination anywhere in the paper. Please take note of the following items when proofreading spelling and grammar.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Abbreviations that incorporate periods should not have spaces: write "C.N.R.S.," not "C. N. R. S." Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units (SI units are encouraged).
- Do not mix complete spellings and abbreviations of units: "Wb/m2" or "webers per square meter," not "webers/m2." Spell units when they appear in text: "...a few henries," not "...a few H."
- Use a zero before decimal points: "0.25," not ".25."

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number the equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop.

$$\int_0^{r_2} F(r,\phi)\, dr\, d\phi = [\sigma r_2 / (2\mu_0)] \qquad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)," not "Eq. (1)" or "equation (1),"

except at the beginning of a sentence: “Equation (1) is ...”

D. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o.”
- In American English, commas, semi-/colons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset,” not an “insert.” The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively.”
- In your paper title, if the words “that uses” can accurately replace the word using, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect,” “complement” and “compliment,” “discreet” and “discrete,” “principal” and “principle.”
- Do not confuse “imply” and “infer.”
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”
- The abbreviation “i.e.” means “that is,” and the abbreviation “e.g.” means “for example.”

IV. Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention as below

IJNMT_firstAuthorName_paperTitle.

In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper. Please take note on the following items.

A. Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

B. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include ACKNOWLEDGMENTS and REFERENCES, and for these, the correct style to use is “Heading 5.”

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles, named “Heading 1,” “Heading 2,” “Heading 3,” and “Heading 4”, are prescribed.

C. Figures and Tables

Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1,” even at the beginning of a sentence.

TABLE I. TABLE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy		

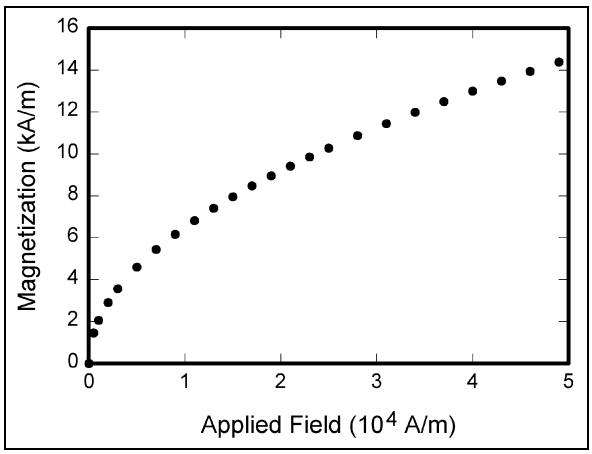


Fig. 1. Example of a figure caption

V. Conclusion

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

Appendix

Appendixes, if needed, appear before the acknowledgment.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments. Avoid expressions such as "One of us (S.B.A.) would like to thank ... ." Instead, write "F. A. Author thanks ... ." You could also state the sponsor and financial support acknowledgments here.

References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2355-0082

Universitas Multimedia Nusantara
Scientia Garden Jl. Boulevard Gading Serpong, Tangerang
Telp. (021) 5422 0808 | Fax. (021) 5422 0800