ISSN 2085-4552

ULTIMATICS

Jurnal Teknik Informatika

Engineering | Vol. 16 | No. 2 | Hal. 76-154 | December 2024 | ISSN 2085-4552

UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2085-4552

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 : ultimatics@umn.ac.id

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 Computer Science and Informatics, as well as the latest theoretical and practical issues, including Analysis and Design of Algorithm, Software Engineering, System and Network security, 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. Ultimatics : Jurnal Teknik Informatika is published regularly twice a year (June and December) and is managed by the Informatics Study Program at Universitas Multimedia Nusantara.

ISSN 2085-4552

Call for Papers

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. ULTIMATICS is published twice a year by Faculty of Engineering and Informatics of Universitas Multimedia Nusantara in cooperation with UMN Press.

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.

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 2085-4552

FOREWORD

ULTIMA Greetings!

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 Computer Science and Informatics, as well as the latest theoretical and practical issues, including Analysis and Design of Algorithm, Software Engineering, System and Network Security, 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. Ultimatics: Jurnal Teknik Informatika is published regularly twice a year (June and December) and is published by the Faculty of Engineering and Informatics at Universitas Multimedia Nusantara.

In this December 2024 edition, Ultimatics enters the 2nd Edition of Volume 16. In this edition there are ten scientific papers from researchers, academics and practitioners in the fields of Computer Science and Informatics. Some of the topics raised in this journal are: Application Fuzzy AHP-TOPSİS Hybrid Method in Facility Location Selection for Software Systems, Fuzzy TOPSIS Implementation for The Determination of Priority Scale in Improving Service Quality, Data Mining Motorbike Sales Classification Using a Combination of K-Means and Naïve Bayes Algorithms, Application of Deep Learning Techniques for Enhancing Arabic Vocabulary Acquisition in Students at Mts Darun-Najah, Sentiment Analysis in E-Commerce: Beauty Product Reviews, Prostate Cancer Screening for Specific Races Using Bioinformatics and Artificial Intelligence on Genomic Data, Improved SVM for Website Phishing Detection Through Recursive Feature Elimination, Implementation of Deep Learning Model for Identification of Skin Diseases by Utilizing Convolutional Neural Network, Comparison Of Multilinear Regression and Adaboost Regression Algorithms in Predicting Corrosion Inhibition Efficiency Using Pyridazine Compounds, Implementation of Gamification Method and Fisher-Yates Shuffle Algorithm for Design and Development Django Learning Application, Leveraging Content-Based Filtering for Personalized Game Recommendations: A Flutter-Based Mobile Application Development.

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 ultimatics@umn.ac.id and the webpage of our Journal here.

Finally, we would like to thank all contributors to this December 2024 Edition of Ultimatics. We hope that scientific articles from research in this journal can be useful and contribute to the development of research and science in Indonesia.

December 2024,

Fenina Adline Twince Tobing, S.Kom., M.Kom.
Editor-in-Chief

ISSN 2085-4552

DAFTAR ISI

ISSN 2085-4552

Application of Fuzzy AHP-TOPSİS Hybrid Method in Facility Location Selection for Software Systems

Kamala Aliyeva [1]
[1]Department of Instrument Engineering, Azerbaijan State Oil and Industry University
[1]Azadlig Avenue, AZ1010, Baku, Azerbaijan, kamala.aliyeva@asoiu.edu.az

Accepted 07 May 2024
Approved 08 January 2025

***Abstract*— Facility location is an integral part of the strategic planning process of almost every organization. Selecting the right location for software systems facilities involves considering various factors to ensure optimal performance, reliability, and cost-effectiveness. For business success, and competitive advantage there are some critical factors that very highly affect facility location. They are proximity to customers, infrastructure, labor quality, total cost, suppliers, etc. The criteria for selecting a facility location may be vaguely defined or open to interpretation. External factors such as economic conditions, political stability, and environmental risks may introduce vagueness and unpredictability into facility location decisions. In this paper we apply fuzzy AHP-TOPSIS hybrid method for facility location in software systems. Fuzzy AHP (Analytic Hierarchy Process) and fuzzy TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) are both decision-making methods commonly used in facility location selection. Fuzzy AHP is particularly useful in situations where decision criteria are subjective and uncertain, providing a more robust framework for making well-informed decisions. Fuzzy TOPSIS is useful in complex decision-making where uncertainty and subjectivity play a significant role, offering a flexible and comprehensive approach to evaluating and ranking alternatives. In the first part of the facility location selection process, we use fuzzy AHP method for determining weights of criteria that are important in selection process. Then by using fuzzy TOPSIS we rank alternatives and select appropriate location for facility. Selection of the best location for software systems provided according to three attributes and three alternatives (A, B, C). Relative closeness of each alternative to the ideal solution represents that alternative C is the best alternative.**

***Index Terms*— *Fuzzy numbers, facility location selection, software systems, ideal solution, fuzzy AHP-TOPSIS*.**

I. INTRODUCTION

Facility location typically refers to the problem of selecting the location of facilities (for example, warehouses, factories, offices, production centers, stores) based on various criteria, such as minimizing transportation costs or maximizing customer service [1]. Facility location in the context of software systems often refers to the strategic placement of data centers, servers, or other infrastructure to optimize performance, reliability, and cost-effectiveness. Key considerations for facility location in software systems are proximity to users, redundancy and disaster recovery, cost optimization, regulatory compliance, network connectivity, scalability, security, environmental considerations etc. Placing facilities closer to many users can reduce latency and improve response times. This is particularly important for real-time applications like online gaming or video streaming. Distributing facilities across different geographic locations helps ensure redundancy and disaster recovery. In the event of a natural disaster or network outage in one location, services can be quickly restored from another location. Considerations such as the cost of real estate, electricity, cooling, and labor vary by location. Choosing locations with lower operational costs can result in significant savings over time. Compliance with data protection regulations may require storing data in specific geographic regions. It's essential to choose locations that comply with relevant laws and regulations. Facilities should be in areas with robust network connectivity to ensure high-speed and reliable connections to the internet backbone and other networks. Choose locations that can accommodate future growth and scalability needs. This includes factors such as available space for expansion and access to skilled labor. Security considerations, including physical security and access controls, are crucial for protecting data and infrastructure. Facilities should be in areas with low crime rates and have appropriate security measures in place. Choosing locations with access to renewable energy sources or implementing energy-efficient technologies can help reduce environmental impact. Facility location decisions for software systems require careful consideration of technical, regulatory, financial, and environmental factors to ensure optimal performance, reliability, and compliance. Some attributes are so important in selection of location that they control all

decision process. These are capacity, successful labor climate, distances, accessibility, service, proximity to suppliers and resources. Capacity considerations are an important consideration in facility location decisions that directly impact operational efficiency and customer satisfaction. Successful labor climate plays a critical role in facility location decisions by influencing workforce availability, costs, stability, productivity, regulatory compliance, and community relations. Businesses that prioritize a favorable labor climate are better positioned to establish sustainable and successful operations in their chosen locations. Labor climate is a criterion of wages, training needs, regards to work, labor performance, and union strength. Proximity to customers or locating near customers is important when customers' needs technical support, products are voluminous and shipping rates are high [2]. Proximity to suppliers and resources is a critical consideration in facility location decisions, impacting transportation costs, supply chain efficiency, quality control, collaboration, risk mitigation, and access to specialized skills. By strategically locating facilities close to suppliers and key resources, businesses can gain competitive advantages, improve operational performance, and enhance overall business resilience. These need permanent coordination and negotiation, which can become heavier as distance increases [3]. In selection process is especially important availability resources and minimal costs. Different facility location factors are determined in place choosing, including investment cost, availability of high-quality labors, shipping, infrastructure. and facility location thus obviously involves multiple criteria.

Uncertainty is a significant factor in facility location decisions due to several reasons. Market dynamics, economic conditions, consumer preferences, and market demand can fluctuate, leading to uncertainty in forecasting future sales volumes and distribution patterns. This uncertainty makes it challenging to determine the optimal location for facilities to meet evolving market demands effectively. It offers a way to model and reason about uncertain or ambiguous situations by capturing the inherent fuzziness in human reasoning and natural language [4]. Changes in regulations, zoning laws, and government policies can impact facility location decisions. Uncertainty regarding future regulatory requirements or restrictions may influence the suitability of certain locations and affect long-term investment decisions. The conditional method of approaching facility location problems like cost volume analysis that is s a managerial accounting technique used to examine the relationship between costs, volume, and profit within a business, factor rating method that is a decision-making tool utilized in place choosing or facility location analysis, and center of gravity method that is a quantitative methodology used in facility location analysis to define the optimal place for a facility, such as a warehouse, distribution center, or manufacturing plant are generally less effective at combating imprecision or vagueness in linguistic judgments [5]. In real life, data for assessing the accessibility of object locations for various subjective factors and factor weights are represented in linguistic terms [6]. For effectively resolve the vagueness that often arises in accessible information and eliminate the essential vagueness of human thinking and preferences, fuzzy set theory has been employed to identify uncertain multi-attribute decision-making problems [7]. Thus, fuzzy AHP -TOPSIS hybrid method offered in this article for facility location selection problem in software system, where the ratings of various alternative locations under different subjective attributes and the weights of all attributes are introduced by fuzzy numbers [8]. The basis for decision-making is that the main decisions are made based on the results of AHP [9]. The decision-makers require to estimate alternatives always comprise vagueness. For modeling such vagueness in the facility location selection, fuzzy sets can be integrated with binary comparisons, for example using the AHP extension [10]. The fuzzy AHP technology permits a more precise specification of the decision-making process [11]. The fuzzy TOPSIS technology is a highly employed technique in decision-making for prioritization alternatives [12].

In this article, the fuzzy AHP-TOPSIS hybrid technique used for defining weights of importance of attributes, ranking alternatives, and determining best location for software system [13-15]. Basic goal in facility location for software systems is to create a strategic infrastructure footprint that optimizes performance, reliability, and cost-effectiveness while ensuring compliance with regulations and minimizing environmental impact. This article is structured as follows. Section 2 represents the main steps of the fuzzy AHP-TOPSIS hybrid technique that is employed in facility location problem. Section 3 offered hybrid methodology with fuzzy numbers for the facility location problem. Section 4 represents discussion and Section 5 presents conclusions of this research.

II. THEORY

Fuzzy AHP-TOPSIS methodology is one of the largely used techniuqes of multi-attribute decision making and this hybrid decision-making method combines the Analytic Hierarchy Process methodology with the Technique for Order of Preference by Similarity to Ideal Solution method, while also incorporating fuzzy set theory. This method is especially useful when dealing with decision problems involving multiple criteria or attributes that are subjective, imprecise, or uncertain in nature. Basic functions of fuzzy AHP-TOPSIS hybrid methodology are hierarchy formation, fuzzy pairwise comparison, aggregation of weights, fuzzy normalization, determining fuzzy similarity to ideal solution. Hierarchy formation is identifying the decision hierarchy, which consists of the main objective, attributes, and alternatives. Break down the main objective into multiple criteria and further decompose each criterion into sub-criteria if necessary. Fuzzy AHP use fuzzy pairwise comparison matrices to assess the comparative significancy or weights of attributes and

sub-attributes. Decision-makers assign linguistic terms such as "weak", "strong", "very strong", "extremly" or fuzzy numbers to express the pairwise comparisons between criteria based on their perceived importance or preference. Aggregation of weights function is aggregating the fuzzy pairwise comparison matrices to calculate the overall weights of attributes and sub-attributes. Various aggregation methods, such as fuzzy geometric mean, fuzzy arithmetic mean, or fuzzy weighted average, can be used to compute the aggregated weights. Fuzzy normalization normalize the decision matrix for each criterion to convert linguistic assessments or fuzzy numbers into crisp values. Fuzzy normalization techniques, such as triangular or trapezoidal fuzzy numbers, can be employed to handle vagueness and imprecision in the data. Fuzzy TOPSIS technique determine fuzzy positive ideal solution and fuzzy negative ideal solution is determined [13]. Then, computed the fuzzy closeness coefficient or similarity score for each alternative relative to the fuzzy positive ideal solution and fuzzy negative ideal solution using fuzzy TOPSIS. Alternatives with higher similarity scores to the fuzzy positive ideal solution and lower similarity scores to the fuzzy negative ideal solution are considered more preferable. Ranking and sensitivity analysis is odering the alternatives on base of their fuzzy closeness coefficients to identify the most preferred alternative(s). Performance sensitivity analysis to assess the robustness of the rankings to changes in the criteria weights or input data and evaluate the stability of the decision. Fuzzy AHP-TOPSIS provides a structured and systematic approach to decision-making in complex and uncertain environments, allowing decision-makers to incorporate subjective judgments, imprecise data, and uncertainty into the decision process. By integrating fuzzy set theory with AHP and TOPSIS, this method enables more comprehensive and nuanced decision analysis, leading to more informed and robust decisions. Some of the superiorities of this hybrid technique are determining weights of criterias by calculating consistency ratio, rationality, comprehensibility, well computational efficiency and the ability to measure the relative performance of each alternative in a simple mathematical form.

The basic steps in multi-attribute decision-making technique fuzzy AHP-TOPSIS are the following:

Step 1. Creating a fuzzy comparison matrix. The scale of linguistic terms is determined. The scale used is the triangular fuzzy numbers scale from one to nine and determined by the membership function that represented in Table I

TABLE I. SCALE OF INTEREST

Scale of interest	Linguistic term	Member function
1	Equally important	(1,1,1)
3	Weakly important	(1,3,5)
5	Strongly more important	(3,5,7)
7	Very strongly important	(5,7,9)
9	Extremely important	(7,9,9)

Then, using the triangular fuzzy numbers to make pair-wise comparison matrix for the basic attribute and sub-attribute. The form of fuzzy pairwise comparison matrix represented in formula (1) [14].

$$\tilde{A} = \begin{bmatrix} 1 & \cdots & \tilde{a}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{a}_{1n} & \cdots & 1 \end{bmatrix} \quad (1)$$

Step 2. Determining fuzzy geometric mean. The fuzzy geometric mean calculated by using formula (2) [14]:

$$\tilde{r}_i = \left(\tilde{a}_{i1} \otimes \tilde{a}_{i2} \otimes \cdots \otimes \tilde{a}_{in}\right)^{1/n} \quad (2)$$

where $\tilde{a}_{in}$ is estimation of fuzzy comparison matrix from attributes i to n. The outcome of the fuzzy geometric mean will be later to called local fuzzy number.

Step 3. Determining the fuzzy weight for each attribute. Calculate the global fuzzy number for each evaluated attribute with formula (3).

$$\tilde{w}_i = \tilde{r}_1 \otimes \left(\tilde{r}_1 \oplus \tilde{r}_2 \oplus \cdots \tilde{r}_n\right)^{-1} = \left(lw_i, mw_i, uw_i\right) \quad (3)$$

Step 4. Determining the best non fuzzy performance. The global fuzzy number transformed to crisp weight value using the sentry of area method to find the value of best non fuzzy performance (BNP) from the fuzzy weight in each attribute, determined using formula (4).

Step 5. Normalization decision matrix provided by using formula (4) [14].

$$r_{ij} = \frac{a_{ij}}{\sqrt{\sum_{k=1}^{m} a_{kj}^2}} \quad (4)$$

Step 6. Formulated weighted normalized decision matrix by using equation (5) [14].

$$V_{ij} = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \cdots & w_n r_{1n} \\ w_1 r_{21} & w_2 r_{22} & \cdots & w_n r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_1 r_{m1} & w_2 r_{m2} & \cdots & w_n r_{mn} \end{bmatrix} \quad (5)$$

Step 7. Determine fuzzy positive ideal solution A^* and fuzzy negative ideal solution A^-. The positive ideal solution technique represents that each evaluated factor has a monotonically increasing or decreasing characteristic. For determination of the fuzzy positive ideal solution set used equation (6) [15]:

$$A^* = \left\{\max v_{ij} \mid j \in J\right\}, \left\{\min v_{ij} \mid j \in J'\right\}$$
$$A^* = \left\{v_1^*, v_2^*, \cdots, v_n^*\right\} \quad (6)$$

Fuzzy negative ideal solutions set selects the smallest of the column values in the matrix. For determination of the fuzzy negative ideal solution set used equation (7) [15]:

$$A^{-} = \left\{\min v_{ij} \mid j \in J\right\}, \left\{\max v_{ij} \mid j \in J'\right\}$$
$$A^{-} = \left\{v_1^{-}, v_2^{-}, \cdots, v_n^{-}\right\} \quad (7)$$

Step 8. Calculation positive ideal and negative ideal the seperation measures.

Positive ideal separation measure is calculated by using formula (8).

$$S_i^{*} = \sqrt{\sum_{j=1}^{n}\left(v_{ij} - v_j^{*}\right)^2} \quad (8)$$

Negative ideal separation measure is calculated by using formula (9) [16].

$$S_i^{-} = \sqrt{\sum_{j=1}^{n}\left(v_{ij} - v_j^{-}\right)^2} \quad (9)$$

Step 9. Calculation the relative closeness to the positive ideal solution [16].

$$C_i^{*} = \frac{S_i^{-}}{S_i^{-} + S_i^{*}} \quad (10)$$

Step 10. Ranking alternatives.

III. METHOD

Selecting the appropriate facility for software systems involves choosing the infrastructure or platform where the software will be developed, deployed, and maintained. There are different functions for facility selection process in the context of software systems [17].

- Defining requirements. Clearly defining the requirements and objectives of the software system. Consider factors such as scalability, performance, security, reliability, compliance, and integration needs. Determine any specific infrastructure or platform requirements based on the nature of the software and its intended use.
- Evaluating hosting options. Assess different hosting options for the software system, such as on-premises, cloud, or hybrid solutions. Evaluate the advantages and disadvantages of each option in terms of cost, scalability, flexibility, reliability, security, and maintenance requirements.
- Considering cloud providers. If opting for a cloud-based solution, evaluate various cloud service providers based on factors such as pricing, services offered, performance, availability, security, compliance certifications, and geographical coverage.
- Assessing infrastructure requirements. Determine the infrastructure requirements for the software system, including computing resources, networking, databases, development tools, and middleware. Consider whether specialized hardware or software components are needed to support specific functionalities or performance requirements.
- Evaluating data storage and management. Assess requirements for data storage, management, and backup. Consider factors such as data volumes, access patterns, data consistency, latency, replication, disaster recovery, and compliance with data protection regulations.
- Considering development environment. Evaluate options for the development environment, including programming languages, frameworks, extension tools, control systems, continuous integration/continuous deployment pipelines, and collaboration platforms. Choose tools and technologies that align with the skills and preferences of the development team and support efficient software development practices.
- Assessing security and compliance. Ensure that the selected facility meets security and compliance requirements for the software system. Consider factors such as data encryption, access controls, identity management, audit logging, vulnerability management, and regulatory compliance.
- Evaluating support and maintenance. Consider support and maintenance requirements for the software system, including monitoring, troubleshooting, patch management, upgrades, backups, and technical support. Evaluate the availability of support services from vendors or service providers and assess their responsiveness and expertise.
- Cost analysis. Conduct a comprehensive cost analysis of different facility options, taking into account upfront costs, ongoing operational expenses, licensing fees, subscription costs, and potential cost savings or cost avoidance associated with each option.
- Risk assessment. Identify and assess potential risks associated with each facility option, such as vendor lock-in, service outages, security breaches, data loss, regulatory non-compliance, and changes in business requirements or market conditions. Develop mitigation strategies to address identified risks.
- Finalizing decision. Based on the evaluation criteria, prioritize the facility options and make a final decision on the most suitable facility for the software system. Document the rationale behind the decision and communicate it to stakeholders involved in the software development and deployment process. By following this systematic approach, organizations can make informed decisions when selecting facilities for software systems, ensuring that the chosen facility meets the requirements of the software system and supports its successful development, deployment, and maintenance. There may be a lack of comprehensive data or uncertainty surrounding key factors such as demand patterns, infrastructure availability, regulatory requirements, making it challenging to make well-informed decisions. Stakeholders may have different preferences, priorities, and risk tolerances when it comes to facility location selection. Vague or conflicting preferences can lead to uncertainty and disagreement in decision-making. The business and technological landscape is constantly evolving, introducing uncertainty and vagueness into facility location decisions. Factors such as market trends,

technological advancements, and regulatory changes can impact the suitability of a chosen location over time.

Assume that a multi criteria decision making problem of facility location selection for software systems involves three criteria - C_1, C_2, C_3, (C_1 - Labor climate, C_2 - Proximity to customers, C_3 -Proximity to suppliers and resources) and 3 alternatives - A, B, C. For determining best location of software systems integrated fuzzy AHP-TOPSIS approach is used. Graphical structure of this methodology given in fig 1.

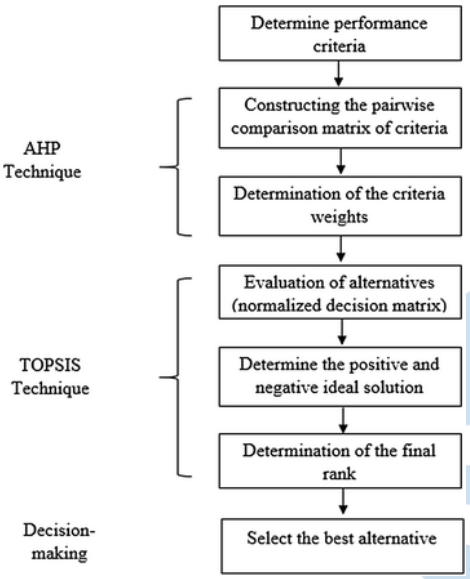


Fig 1. Graphical structure of fuzzy AHP-TOPSIS hybrid methodology

Step 1. Determining labor climate, proximity to customers and proximity to suppliers and resources as criteria for facility location selection [18]. Labor climate criteria in software systems refer to the conditions, environment, and factors that affect the morale, productivity, and satisfaction of software development teams. By considering and actively managing these labor climate criteria, organizations can create an environment that supports the well-being and productivity of their software development teams, ultimately leading to better outcomes for projects and the organization as a whole. These criteria play a crucial role in determining the overall success of software projects. Proximity to customers in software systems refers to the degree to which software development teams interact, understand, and collaborate with end-users or customers throughout the development process. This criterion is essential for building successful software products that meet the needs and expectations of customers effectively. By prioritizing proximity to customers in software systems, organizations can build products that are not only technically robust but also resonate with users, drive adoption, and ultimately contribute to the success of the business. Proximity to suppliers and resources in software systems refers to the accessibility and availability of necessary resources, tools, and external partners that support the software development process. By effectively managing proximity to suppliers and resources in software systems, organizations can enhance their agility, scalability, and competitiveness in delivering high-quality software solutions that meet customer needs and expectations.

Step 2. Constructing the pairwise comparison matrix of criteria. For determining weights of criteria we use fuzzy triangle numbers. Using scale of interest that represented in Table I we construct pairwise comparison matrix that represented in Table II.

TABLE II. PAİRWİSE COMPARİSON MATRİX OF CRİTERİA

	C_1	C_2	C_3
C_1	(1,1,1)	(1,3,5)	(1/7,1/5,1/3)
C_2	(1/5,1/3,1)	(1,1,1)	(1/9,1/7,1/5)
C_3	(3,5,7)	(5,7,9)	(1,1,1)

Step 3. Determining weights of each criteria. The geometric mean of fuzzy comparison values of each attribute is determined by using formula (2). $\tilde{r}_i$ - geometric mean of fuzzy comparison values of "Labor climate" criterion calculated as down.

$$\tilde{r}_i = \left(\prod_{j=1}^{n} \tilde{d}_{ij}\right)^{1/n} = \left[\left(1*1*\tfrac{1}{7}\right)^{\frac{1}{3}}; \left(1*\tfrac{1}{3}*\tfrac{1}{5}\right)^{\frac{1}{3}}; \left(1*5*\tfrac{1}{3}\right)^{\frac{1}{3}}\right] =$$
$$= [0.53, 0.41, 1.18]$$

r_i -geometric mean of fuzzy comparison values of "Proximity to customers" criterion calculated as down.

$$\tilde{r}_i = \left(\prod_{j=1}^{n} \tilde{d}_{ij}\right)^{1/n} = \left[\left(\tfrac{1}{5}*1*\tfrac{1}{9}\right)^{\frac{1}{3}}; \left(\tfrac{1}{3}*1*\tfrac{1}{7}\right)^{\frac{1}{3}}; \left(1*1*\tfrac{1}{5}\right)^{\frac{1}{3}}\right] =$$
$$= [0.29, 0.36, 0.59]$$

r_1 -geometric mean of fuzzy comparison values of "Proximity to suppliers and resources" criterion is calculated as down.

$$\tilde{r}_i = \left(\prod_{j=1}^{n} \tilde{d}_{ij}\right)^{1/n} = \left[(3*5*1)^{\frac{1}{3}}; (5*7*1)^{\frac{1}{3}}; (7*9*1)^{\frac{1}{3}}\right] =$$
$$= [2.44, 3.23, 3.92]$$

The geometric means of fuzzy comparison values for different attributes shown in Table III.
In addition, the total values and the reverse values are also present.

TABLE III. THE GEOMETRIC MEANS OF FUZZY COMPARISON VALUES

Criteria	$\tilde{r}_i$		
Labor climate	0.53	0.41	1.18

Proximity to customers	0.29	0.36	0.59
Proximity to suppliers and resources	2.44	3.23	3.92
Total	3.26	4	5.69
Reverse (power of -1)	0.31	0.25	0.17
Increasing Order	0.17	0.25	0.31

The fuzzy weight of „Labor climate" criteria $\tilde{w}_1$ found by using of equation (3).

$$\tilde{w}_1 = \left[(0.53*0.17);(0.41*0.25);(1.18*0.31)\right] = \\ = \left[0.09, 0.1, 0.36\right]$$

The fuzzy weight of „Proximity to customers" criterion $\tilde{w}_2$ calculated as down.

$$\tilde{w}_2 = \left[(0.29*0.17);(0.36*0.25);(0.59*0.31)\right] = \\ = \left[0.05, 0.09, 0.18\right]$$

The fuzzy weight of „Proximity to suppliers and resources" criterion calculated as down.

$$\tilde{w}_3 = \left[(2.44*0.17);(3.23*0.25);(3.92*0.31)\right] = \\ = \left[0.41, 0.8, 0.9\right]$$

The weight of each criterion represented by fuzzy numbers, such as,

$$\tilde{w}_1 = (0.09,\ 0.1,\ 0.36)$$
$$\tilde{w}_2 = (0.04, 0.09, 0.2)$$
$$\tilde{w}_3 = (0.41, 0.8, 0.9)$$

A performance evaluation matrix, also known as performance assessment matrix, is an organized tool that used to estimate and evaluate the performance of individuals, teams, projects, or processes against predefined criteria or objectives. It provides a systematic framework for measuring performance, identifying strengths, drawbacks, and facilitating decision-making related to performance improvement, recognition, rewards, corrective actions. Performance rating decision matrix presented in Table IV.

TABLE IV. DECISION MATRIX OF PERFORMANCE RATING

	C_1	C_2	C_3
	w_1	w_2	w_3
	0.09, 0.1, 0.36	0.04,0.09,0.2	0.41,0.8,0.9
A	0.06, 0.2, 0.8	0.3, 0.8, 0.9	0.4, 0.65, 0.9
B	0.21, 0.6, 0.9	0.15, 0.5, 0.75	0.05, 0.07, 0.09
C	0.23,0.5,0.8	0.15,0.25,0.65	0.65, 0.74, 0.85

By implementing a performance evaluation matrix, organizations can effectively assess and manage performance, foster accountability, transparency, and drive continuous improvement and excellence across individuals, teams, and organizational units.
For solution of this problem and selection better location for facility used TOPSIS steps [19].
Step 4. Constructing the weighted decision matrix (Table V).

TABLE V. WEIGHTED DECISION MATRIX

	C_1	C_2	C_3
A	0.005, 0.02, 0.3	0.01, 0.07, 0.18	0.16, 0.52, 0.81
B	0.02, 0.06, 0.32	0.006, 0.04, 0.15	0.02, 0.06, 0.08
C	0.02, 0.05, 0.29	0.006, 0.02, 0.13	0.27, 0.6, 0.76

Step 5. Calculating the fuzzy positive ideal solution and fuzzy negative-ideal solution:

$$A^* = \left\{(0.02,\ 0.06,\ 0.32), (0.01,\ 0.07,\ 0.18), (0.27, 0.6, 0.76)\right\}$$
$$A^- = \left\{(0.005,\ 0.02,\ 0.29), (0.006, 0.02, 0.13), (0.02,\ 0.06,\ 0.08)\right\}$$

Determining separation measure for each alternative. For example, separation measure for first alternative can be determined as

$$S_A^* = \left\{\left[(0.005,\ 0.02,\ 0.29) - (0.005,\ 0.02,\ 0.29)\right]^2 + \right.$$
$$+\left[(0.01,\ 0.07,\ 0.18) - (0.005,\ 0.02,\ 0.29)\right]^2 +$$
$$\left.+\left[(0.16,\ 0.52,\ 0.81) - (0.005,\ 0.02,\ 0.29)\right]^2\right\}^{1/2} =$$
$$= (0.091, 0.34, 1.45)$$

$$S_A^* = (0.091, 0.34, 1.45), \qquad S_A^- = (0.08, 0.3, 1.05)$$

Similarly, for other alternatives separation measures are determined.

$$S_B^* = (0.19, 0.65, 1.98), \qquad S_B^- = (0.04, 0.19, 0.76)$$
$$S_C^* = (0.18, 0.60, 1.96), \qquad S_C^- = (0.035, 0.185, 0.73)$$

Step 6. Determination of the final rank by defining relative closeness to the ideal solutions for each alternative:

$$C_A^* = \frac{S_A^-}{S_A + S_A^-} = \frac{(0.091, 0.34, 1.45)}{(0.091, 0.34, 1.45) + (0.08, 0.31, 1.05)} = \\ = (0.05, 0.3, 5.6)$$

$$C_B^* = \frac{S_B^-}{S_B^* + S_B^-} = (0.02, 0.3, 3.4)$$

$$C_C^* = \frac{S_C^-}{S_C^* + S_C^-} = (0.07, 0.9, 9.1)$$

Step 7. Selecting best alternative. Relative closeness of each alternative to the ideal solution represents that alternative C is best alternative - $C > A > B$

IV. RESULT

Facility location for software systems in uncertain conditions involves determining optimal locations for data centers, servers, or other computing facilities considering potential uncertainties such as fluctuations in demand, network latency, power outages, and natural disasters. There are some basic strategies for addressing uncertainty in facility location for software systems. They are conducting a comprehensive risk analysis to identifying potential sources of uncertainty, including environmental factors, market dynamics, and technical issues, assessing the probability and potential impact of each risk to inform decision-making, designing facilities with flexibility in mind to adapt to changing conditions [20]. This could involve modular designs that allow for easy expansion or relocation of computing resources based on demand fluctuations or unexpected events. Implementing redundancy and backup mechanisms to mitigate the impact of failures or disruptions is important in facility location [21]. This may include duplicating critical infrastructure components across multiple locations to ensure continuous operation in the event of failures [22]. Game-theoretic models can be used to address uncertainties in competitive environments where the actions of other agents (e.g., competitors, regulators) are uncertain [23]. This approach helps in making strategic decisions regarding facility locations. These methods, such as genetic algorithms and simulated annealing, are used to find near-optimal solutions for complex facility location problems under uncertainty. They are particularly useful when exact solutions are computationally infeasible [24]. Distributing computing facilities across geographically diverse locations for minimization the risk of localized disruptions such as natural calamity or regional infrastructure failures, considering factors such as proximity to target markets, regulatory requirements, and network connectivity when selecting locations also important issues in facility location of software systems. Implementation real-time monitoring systems to track key performance indicators and environmental conditions and using this data to dynamically adjust resource allocation, routing decisions, and failover strategies to optimize performance and resilience in uncertain conditions are significant of facility location for software systems in uncertainty conditions. Cloud services, leverage cloud computing services are offered built-in redundancy, scalability, and geographic diversity. Cloud providers typically operate data centers in multiple regions, providing inherent resilience to failures and disruptions. Foster collaboration and communication between stakeholders, including IT teams, facility managers, and business units, ensure alignment of goals and priorities in addressing uncertainty. Regularly review and update facility location strategies based on evolving risks and opportunities. By incorporating these strategies into facility location decisions, software systems can better withstand uncertainty and maintain optimal performance and reliability in dynamic environments.

V. CONCLUSIONS

The selection of a facility location is a multifaceted decision that affects various operational, financial, and strategic aspects of a business. A thorough analysis considering these practical implications can lead to a more informed and effective location decision, ultimately contributing to the business's success. The application of the fuzzy AHP-TOPSIS hybrid method in facility location selection for software systems is not only theoretically sound but also practically applicable. It provides decision-makers with a systematic and structured approach to evaluate and select the most suitable location based on their specific requirements and constraints. To make decision on facility location selection for software systems, the methodology of fuzzy AHP-TOPSİS with fuzzy numbers is applied in this paper to take high vagueness that appropriates to the considered problem. Selection of the best location for software systems provided according to three attributes and three alternatives- A, B, C . The first attribute for selection is labor climate, the second attribute is proximity to customers, and third attribute is proximity to suppliers and resources. Results defined from the relative closeness to the positive ideal solutions used for ranking the preference order and determined that alternative C is best alternative for facility location.

In concluding the application of the Fuzzy Analytic Hierarchy Process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) hybrid method in facility location selection for software systems, several key points can be highlighted. The hybrid method provides a robust framework for decision-making by incorporating both fuzzy AHP and TOPSIS. This integration allows for handling the inherent uncertainties and complexities involved in facility location selection for software systems. Facility location selection involves multiple criteria such as labor climate, proximity to customers, proximity to suppliers and resources proximity. The fuzzy AHP-TOPSIS hybrid method enables the consideration of these diverse criteria and their relative importance in decision-making. Fuzzy logic is effective in capturing the subjectivity and imprecision in expert judgments during pairwise comparisons of criteria and alternatives. This ensures a more accurate representation of decision-makers' preferences and enhances the reliability of the decision-making

process. Sensitivity analysis allows for assessing the robustness of the selected facility location against changes in criteria weights and alternative rankings. This helps decision-makers understand the stability of their decisions and identify potential risks associated with the chosen location. Validation of the selected facility location through simulation or real-world experimentation validates the effectiveness of the hybrid method. Moreover, the method can be adapted and customized to suit different contexts and decision-making scenarios, making it a versatile tool for facility location selection in various industries.

In summary, the Fuzzy AHP-TOPSIS hybrid method offers a comprehensive and effective approach to facility location selection for software systems, integrating fuzzy logic, AHP, and TOPSIS to address the complexities and uncertainties inherent in decision-making processes. Its systematic framework, coupled with sensitivity analysis and validation, empowers decision-makers to make informed and reliable decisions regarding facility location selection

REFERENCES

[1] Chen, Y. S., Chu, H. H., and Sangaiah, A. K., "Identifying Ambient Service LocationProblems and Its Application Using a Humanized Computing Model", *Journal of Am-bient Intelligence and Humanized Computing*, vol. 10, pp. 2345-2359, 2019. https://doi.org/10.1007/s12652-018-0.

[2] Yang J, Lee H., "An AHP decision model for facility location selection", *Facilities*, vol. 15, pp. 241–254, 1997. http://dx.doi.org/10.1108/02632779710178785.

[3] Krajewski L. J., Ritzman L. P., "*Operations Management: Strategy and Analysis*," 3rd Edition, Addison-Wesley, Boston, 1993.

[4] Kahraman C., Cebeci U., Ulukan Z. "Multi-criteria supplier selection using fuzzy AHP",*Logistics Information Management*, vol.16, no. 6, pp. 382–394, 2003. http://dx.doi.org/10.1108/09576050310503367.

[5] Hwang C., Yoon K. "Multiple attributes decision making methods and applications", *Springer*, Berlin, 1981. https://doi.org/10.1007/978-3-642-48318-9.

[6] Liang G.S., "Fuzzy MCDM based on ideal and anti-ideal concepts", *European Journal of Operational Research*, vol.112, pp. 682–691, 1999. https://doi.org/10.1016/S0377-2217(97)00410-4.

[7] Kahraman C., Ruan D., Doğan İ., "Fuzzy group decision making for facility location selection", *Information Sciences*, vol. 157, pp. 135–153, 2003. https://doi.org/10.1016/S0020-0255(03)00183-X

[8] Zadeh L.A., "Fuzzy sets", *Information and control*, vol. 8, pp. 338–353, 1965. http://dx.doi.org/10.1016/S0019-9958(65)90241-X

[9] Chang D.Y. "Applications of the extent analysis method on Fuzzy AHP", *European Journal of Operational Research* , vol. 95, no. 3, pp. 649–655, 1996. https://doi.org/10.1016/0377-2217(95)00300-2

[10] Chu T.C. "Selecting plant location via a fuzzy TOPSIS approach", *International Journal of Advanced Manufacturing Technology*, vol. 20, pp. 859–864, 2002. https://doi.org/10.1007/s001700200227.

[11] Ishii H., Lee Y.L., Yeh K.Y., "Fuzzy facility location problem with preference of candidate sites", *Fuzzy Sets Systems*, vol. 158, pp. 1922-1930, 2007. https://doi.org/10.1016/j.fss.2007.04.022.

[12] Sachin, P., Ravi K., "A fuzzy AHP-TOPSIS framwork for ranking the solutions of Knowledge Management adoption in Supply Chain to overcome its barriers", *Expert Systems with Applications*, vol. 41, pp. 679-693, 2014. https://doi.org/10.1016/j.eswa.2013.07.093.

[13] Kusumawardani R., Agintiara M., "Application of Fuzzy AHP-TOPSIS Method for Decision Making in Human Resource Manager Selection Process", *Procedia Computer Science*, vol. 72, pp. 638-646, 2015. https://doi.org/10.1016/j.procs.2015.12.173.

[14] Saaty, T.L., "The Analytic Hierarchy Process", *McGraw-Hill International*, New York, NY, USA, 1980.

[15] Hwang, C.L., Lai, Y.J., Liu, T.Y., "A new approach for multiple objective decision making", *Computers and Operational Research*, vol. 20, no. 8, pp. 889–899, 1993. doi:10.1016/0305-0548(93)90109-v.

[16] El Alaoui, M., "Fuzzy TOPSIS: Logic, Approaches, and Case Studies", *New York: CRC Press*, 2021. doi:10.1201/9781003168416.

[17] Liang G.S., Wang M.J., "A fuzzy multi-criteria decision making method for facility site selection", *The International Journal of Advanced Manufacturing Technology*, vol. 29, no. 11, pp. 2313–2330, 1991.

[18] Aliyeva K. R., "Applying Type-2 Fuzzy TOPSIS Method to Selection of Facility Location", *International Conference on Theory and Applications of Fuzzy Systems and Soft Computing*, pp. 180 – 185, 2022.

[19] Aliyeva K. R., "Decision on Facility Location on Base of Interval Data", *11th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions and Artificial Intelligence*, pp. 371-374, 2021.

[20] Snyder, L. V., "Facility location under uncertainty: A review", *IIE Transactions*, vol. 38, no. 7, pp. 547-564, 2006. https://doi.org/10.1080/07408170500216480.

[21] Ben-Tal, A., El Ghaoui, L., & Nemirovski, A., "Robust optimization", *Princeton University Press*, 2009.

[22] Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A., "A stochastic programming approach for supply chain network design under uncertainty", *European Journal of Operational Research*, vol. 167, no. 1, pp. 96-115, 2005.

[23] Aghassi, M., Bertsimas, D., "Robust game theory", *Mathematical Programming*, vol. 107, no. 1, pp. 231-273, 2006. https://doi.org/10.1007/s10107-005-0686-0.

[24] Guo, D., Lim, G. J., Smith, J. C., "A multi-objective multi-matrix decomposition algorithm for large-scale facility location problems under uncertainty", *European Journal of Operational Research*, vol. 241, no.2, pp. 335-348, 2015.

ISSN 2085-4552

Fuzzy TOPSIS Implementation for the Determination of Priority Scale in Improving Service Quality

Novrindah Alvi Hasanah[1], Muhammad Faisal[2], Linda Salma Angreani[3]
[1]Electrical Engineering, UIN Maulana Malik Ibrahim Malang, Indonesia
[2, 3] Informatics Engineering, UIN Maulana Malik Ibrahim Malang, Indonesia
[1]novrindah@uin-malang.ac.id, [2]mfaisal@ti.uin-malang.ac.id, [3]linda@uin-malang.ac.id

Accepted 07 May 2024
Approved 08 January 2025

***Abstract*— Service quality plays a crucial role in economic development, particularly in the service industry, such as hotel services. Despite this, many hotels lack a systematic approach to help management identify areas that require improvement based on customer feedback. This research aims to develop a system that supports efforts to enhance service quality, utilizing the Fuzzy TOPSIS method. The study incorporates 150 data points obtained from questionnaires distributed to hotel service customers. The research involves two trials: service improvement priority and service eligibility. The results indicate an 84.45% accuracy level for service improvement priority testing, based on 120 out of 150 data points. Additionally, the accuracy level for service eligibility testing is 85.34%, derived from 131 data points out of the total 150. The research findings highlight the cafeteria as a significant area requiring improvement in service quality, aligning with the insights of hospitality experts. These results can serve as a foundation for management to enhance service quality based on selected criteria and alternatives.**

***Index Terms*— service; industry; quality; fuzzy; TOPSIS**.

I. INTRODUCTION

In the era of globalization, the main factor that makes a significant contribution to a country's economic progress is the hospitality and tourism sector, especially the lodging sector (hotels, hostels, apartments, etc.) [1] So, the competition between service providing companies to increase. This requires every company to have strategies and innovations to attract customer interest, one of which is by improving service quality. Service quality has an important role in economic development, including the service industry [2]. Strategies to improve service quality can attract customer interest in determining the hotel services they want to use. However, differences in customer interests make the facilities and quality of services offered not necessarily suitable for customers. Hotel businesses are required to always be sensitive to changing needs and attitudes from customers. In the service industry, especially lodging, customers must pay attention to customer needs and hotel efficiency in order to increase sales [3][4]. Several studies have been conducted on improving service quality and its impact on customer satisfaction, focusing on various subjects. For example, research have been conducted to explores the effect of hotel service quality on customer loyalty [5][6][7].

Based on several previous studies, one of the strategies in service quality control is to incorporate digitization in line with current advancements. So far, many hotel services have only provided the option for customers to submit complaints to the front office, which are then manually forwarded to the management. Meanwhile, the management of the hotel that we used as the object of research does not yet have its own system that can facilitate the identification of facilities and services requiring improvement based on customer complaints. In fact, reviews on Google regarding the hotel we researched show that several customers provided feedback. Examples include the front office service not being available 24 hours, the hotel not providing breakfast, and issues with the cafeteria only offering snacks rather than other food options. So far, there have been 280 Google reviews [8] revealing repeated customer complaints, yet the absence of any response from the hotel could lead to a further drop in its rating. Despite Google reviews being one of the benchmarks for customers seeking temporary accommodations, 1-star and 2-star ratings persist. Therefore, we propose to create a system that can accommodate customer complaints with the aim that management can improve hotel services. This ineffectiveness hampers the management of consumer complaints. Thus, the strategy for improving service quality is impeded. Consequently, there is a need for a digitalization-based system with a methodology that can assist management in identifying services and facilities requiring improvement based on customer complaints. Several previous studies have explored various decision-making methods developing thus far. For example, a utilizes the Fuzzy method combined with AHP to determine ranking criteria for analyzing

strategic service quality [2]. The study is conducted from the perspective of digital transformation in the hospitality industry. Another research study [9] examines the comparison between the TOPSIS and Fuzzy TOPSIS methods in evaluating financial performance analysis of banks. Additionally, a study [10] performs a stability analysis of the method by altering the service quality ranking for different metro lines. Other research on transportation service case studies has also been carried out in this study [11]. Controlling service quality was also discussed regarding passenger satisfaction in assessing service quality in the airline industry [12]. Another study [13] focused on construction of a service quality scale. Lastly, the research study [1] also discussed the issue of service quality and human resources in providing services.

Based on some of these studies, the Fuzzy TOPSIS method theoretically exists and can be used to assess service quality. This is supported by the theory proposed by Yoon & Hwang as cited in [14], which states that Fuzzy TOPSIS is the appropriate method for decision-making in determining priorities, TOPSIS is a popular method in a mathematical approach that selects the optimal solution by measuring the distance of each alternative to the positive ideal solution and negative ideal solution simultaneously. In decision making systems, positive ideal solutions are popular solutions used to maximize benefit criteria and minimize cost criteria. In contrast to the positive ideal solution, the negative ideal solution is a less popular solution in maximizing the criteria [15][16] such a research [17] and [18]. The method solves problems based on the fundamental idea that the chosen alternative has the closest distance to the positive ideal solution and the farthest distance from the negative ideal solution. Additionally, research conducted by [1] suggests that the Fuzzy TOPSIS method can be served as a measure of the performance of alternative decisions using simple computational forms, the same as the use of the Fuzzy TOPSIS method in research [19], then [20] and the last one is [21].

This research focuses on developing a system that can support efforts to improve service quality. The system was built by utilizing the Fuzzy TOPSIS method, which has been proven to effectively address problems by emphasizing the shortest distance to a positive ideal solution and the farthest distance from a negative ideal solution. The study focuses on a hotel, which is one of the businesses of the hotel businesses of state universities in Malang Indonesia. Data was collected by distributing questionnaires to customers over approximately one month. The study aims to assist the inn management team in identifying areas where services may be suboptimal according to customers. Consequently, the findings can facilitate immediate improvements in service quality.

Several previous studies have been conducted to measure the level of service quality. For instance, research conducted in [5] focused on examining customer satisfaction levels and customer involvement behavior (CEB). The study investigates the mediating relationship between website quality, customer satisfaction, and CEB. The results indicate that customer satisfaction has a significant impact on CEB and customer loyalty. Furthermore, additional research on the impact of service quality on customer satisfaction was conducted in [6]. The study assessed hotel service quality during the COVID-19 pandemic using data from 400 hotels in Thailand. The research findings demonstrate that service quality has a significant influence on customer loyalty. Additionally, research conducted in [7] aimed to investigate the relationship between service quality (SQ) and customer loyalty (CL), with a focus on the mediating role of customer satisfaction (CS) and customer delight (CD). The study involved 313 hotel customers. The results indicated that service quality had a significant positive effect on both customer satisfaction and delight. Furthermore, the findings revealed that customer satisfaction and delight effectively mediated the relationship between service quality and customer loyalty.

Based on the literature briefly mentioned in the background, this study employs the Fuzzy TOPSIS method as a decision-making approach to identify service criteria that require improvement. Previous studies have demonstrated the effectiveness of this method in addressing various issues, as exemplified by research conducted [1]. The study conducted a comparison between the TOPSIS and Fuzzy TOPSIS methods for financial performance analysis. The object of the research is banks. Another research study [10] applied Fuzzy theory analysis in conjunction with TOPSIS to evaluate service quality in rail services. A survey was conducted to assess service quality based on several criteria, including access (egress), security checks, ticket purchases or recharges, card swiping, transfers, waiting for boarding, in-vehicle experience, and other extended services. The research presents the findings of several criteria that require improvement. Research on evaluating business performance of homestay in China is also conducted [22]. The research proposes an alternative selection method using Fuzzy TOPSIS. It has been found that after the COVID-19 pandemic, homestays operating in rural areas have optimal performance. These results can be used as a benchmark by experts in assessing the quality of homestays. Lastly, the research study [20] served as the basis for choosing the method employed in this study. The research implementing Fuzzy TOPSIS by utilizing quality dimensions as input for the quality of public transportation services.

TABLE 1
ALTERNATIVE BASED ON REGARDING HOTEL BUSINESS STANDARDS

No	Alternative
1.	Rooms
2.	Bathrooms
3.	Cafetaria
4.	Front office
5.	Worship place
6.	Parking lots

TABLE 2
CRITERIA BASED ON REGARDING HOTEL BUSINESS STANDARDS

No	Criteria
1.	Products
2.	Service
3.	Cleanliness
4.	Security
5.	Management
6.	Health

Several previous studies that have been conducted focused on improving service quality through programs built using the same method. However, expert opinion must be an important consideration because experts have knowledge of the case studies used. This research combines surveys conducted on customers who are service recipients with the opinions of experts in the hotel sector who are deemed to have more knowledge in improving service quality in the hotel business. The experts used in this research were three hotel industry experts from three different hotels in Indonesia. The data collected from customers and experts is processed by the system and produces output which is expected to help determine the service quality improvement decision support system.

The Fuzzy TOPSIS method operates by selecting the alternative with the highest value, which represents the chosen alternative that requires service quality improvement. The implementation of the Fuzzy TOPSIS method is integrated into a web-based system that can effectively handle customer complaints. The data that is processed is 150 data resulting from the answers of customers who are hotel customers for one month.

II. METHOD

This research commences with a survey aimed at studying and analyzing the condition of the research object, enabling the definition and formulation of the problems faced by the research object. In this step, brief interviews were conducted with experts in the research object field to identify the criteria and alternatives considered for establishing a reference in determining the priority scale for improving the service quality. The interview results revealed the utilization of six criteria, including service, cleanliness, management, and health. The selection of the six

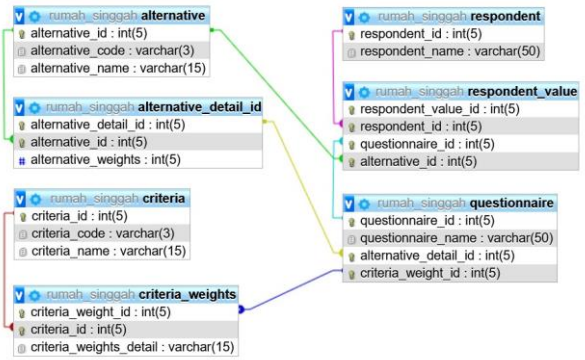


Fig. 1. Design of database system

criteria was based on interviews with experts, taking into account the Minister of State Apparatus Empowerment and Bureaucratic Reform Regulation Number 29 of 2022 concerning Monitoring and Evaluation of the Performance of Public Service Delivery [23] regarding General Guidelines for the Implementation of Public Services and the Regulation of the Minister of Tourism and Creative Economy of the Republic of Indonesia Number PM.53/HM. 001/MPEK/2013 [24] regarding Hotel Business Standards.

Experts and respondents were asked several questions regarding hotel services and facilities used as research objects. Experts answered questions based on previously mentioned government regulations. Some of the questions answered by experts and respondents included ranking service quality criteria from most important to least important. Additionally, there were questions about each alternative, assessed based on the provided criteria. The assessment consists of 5 points: very low, low, sufficient, good, and very good.

As for the determined alternatives, there are six of them, as shown in Table 1. While the alternative will be assessed based on the criteria used in the study. The criteria are shown in Table 2.

These criteria and alternatives were identified by experts based on the available facilities and common customer complaints. The study utilizes data collected from customers over one month, resulting in a successful acquisition of 150 data points, which serve as input for the system. The system was built on a web basis using the Code Igniter framework and programming language using PHP. The database uses phpMyAdmin and contains a total of seven tables. These tables include alternatives, alternative details, criteria, criteria weights, respondents, respondent values, and questionnaires. The seven tables are related to each other as shown in Figure 1. Then, the design of the web page for inputting customer assessments of hotel services is shown in Figure 2.

The method in this study aims to identify the alternative with the highest value to improve service quality. The research methodology employed in this

Fig. 2. Questionnaire form of the quality assessment system

determines the priority for service improvement. Each point outlined above will be discussed in greater detail in the subsequent sub-chapters.

A. Criteria

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.

The application of Fuzzy TOPSIS to determine priority scales involves several steps, including criteria evaluation, alternative eligibility assessment, decision-making, and ranking decision rules. The criteria utilized are measurements, rules, and standards that aid in decision-making. The eligibility of alternatives is determined by various constraints, such as physical availability, resource availability, and information constraints, among others. Subsequently, the evaluation of each available alternative's criteria must be conducted to assess their attractiveness to the criteria weight values or weight values. The weight value of each alternative $Ai(i = 1, 2, \ldots, m)$ for each criterion $Cj(j = 1, 2, \ldots, n)$ the weight values of each alternative

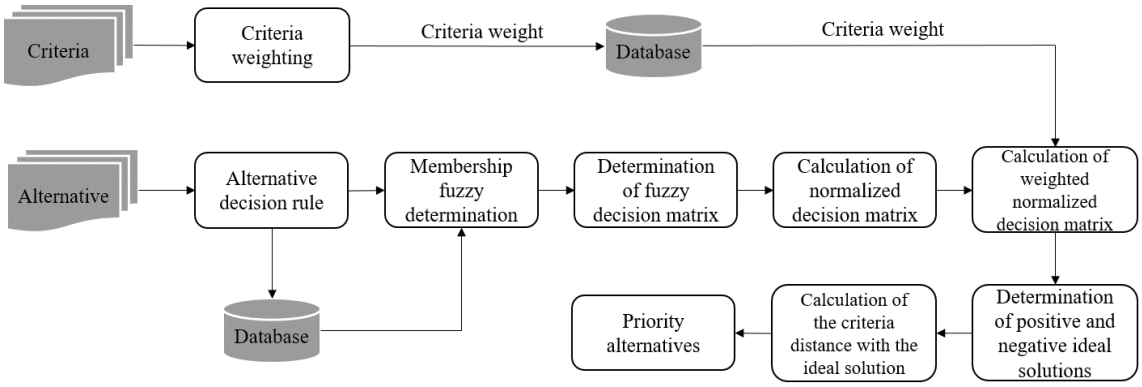


Fig. 3. Block diagram of the method

study is illustrated in Figure 3. The figure illustrates the flow of the research system, which incorporates the implementation of the Fuzzy TOPSIS method. The system database stores the weights assigned to each predefined criterion. The subsequent step involves determining alternative decision rules for each predefined alternative. These rules are utilized to calculate the value of alternative decisions based on assessments provided by respondents. Fuzzy membership is then determined to facilitate the creation of a fuzzy decision matrix in the subsequent stage. The decision matrix is further transformed into a normalized form. Next, the predetermined criteria weights are multiplied by the normalized decision matrix, resulting in a weighted normalized decision matrix. With the weighted normalized decision matrix in hand, it becomes possible to identify the positive ideal solutions and negative ideal solutions. The final step involves calculating the distances between the criteria and the positive ideal solutions, as well as the negative ideal solutions. This calculation yields the outcome, which can be represented as a decision matrix, which can be written as: $D = [xij]m \times n, i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$.

In this system, the weighting of the criteria is determined by experts and then calculated by the system using the Fuzzy TOPSIS method to establish the ranking of each alternative. The objective is to identify the prioritized alternatives for service improvement. The criteria weighting is divided into five categories: very low, low, sufficient, good, and very good. Each expert provides an assessment for the six criteria, indicating their importance on a scale ranging from very unimportant to very important.

Table 3 displays the criteria weight rule, indicating the value of the alternative decision rule for each weight. The alternative decision rules range from 0.0 to 1.0, representing fuzzy rules. Based on the criteria weighting rules, the importance assessment for criteria weighting conducted by the hospitality experts can be converted into a matrix, as illustrated in Table 4. The

TABLE 3
CRITERION WEIGHT RULES

No.	Criterion Weight	Value 1	Value 2	Value 3
1	Very Low	0.0	0.0	0.2
2	Low	0.0	0.2	0.4
3	Enough	0.2	0.4	0.6
4	Good	0.4	0.6	0.8
5	Very Good	0.6	0.8	1.0

TABLE 5
CRITERION WEIGHT

No.	Criteria's Name	Value 1	Value 2	Value 3
1	Product	0.00	0.15	0.35
2	Service	0.50	0.70	0.90
3	Cleanliness	0.40	0.60	0.80
4	Security	0.20	0.40	0.60
5	Management	0.30	0.50	0.70
6	Health	0.00	0.05	0.25

TABLE 4
TRIANGULAR FUZZY NUMBER VALUE

Criterion	Expert Hotels Object 1			Expert Hotels Object 2			Expert Hotels Object 3			Expert Hotels Object 4		
Product	0	0.2	0.4	0	0.2	0.4	0	0.2	0.4	0.	0	0.2
Service	0.6	0.8	1	0.6	0.8	1	0.4	0.6	0.8	0.4	0.6	0.8
Cleanliness	0.4	0.6	0.8	0.4	0.6	0.8	0.2	0.4	0.6	0.6	0.8	1
Security	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6
Management	0.2	0.4	0.6	0.2	0.4	0.6	0.6	0.8	1	0.2	0.4	0.6
Health	0	0	0.2	0	0	0.2	0	0	0.2	0	0.2	0.4

table depicts the conversion of the values provided by the hospitality experts, who serve as the objects, into a matrix by replacing each expert's judgment with a predetermined value based on the criteria weight rule, which follows the Fuzzy rule. By evaluating the level of importance for each criterion through the perspectives of the four hospitality experts, the weighted importance for each criterion is obtained by summing up the values assigned by each expert and dividing it by the number of hospitality experts, as illustrated as in (1).

$$\frac{\sum value\ of\ all\ the\ hotel's\ object\ expert}{\sum hotel's\ object\ expert} \quad (1)$$

Table 5presents the calculation results of the importance weight assessment for each criterion. The calculation results consist of three weights, as they correspond to the triangular curve representation that utilizes three parameters.

B. Alternative

Based on the survey results and discussions held by the researchers with experts, taking into account the facilities available at the research object, six alternatives were identified: rooms, bathrooms, cafeterias, front offices, worship place, and parking lots. The system conducts calculations using the Fuzzy TOPSIS method, resulting in a ranking for each alternative to be selected. The alternative with the lowest preference value is given the highest priority for service improvement. The alternative decision rules are presented in Table 6.

The described alternative represents the object of the study. The system testing utilizes 150 data obtained from customer questionnaires. These data will be compared with the opinions of three experts from hotel services who will serve as a basis for comparison. The ranking of alternatives will then be used as a reference for further calculations in the research, employing the Fuzzy TOPSIS method. The flowchart depicting the ranking of priority alternatives for service improvement using Fuzzy TOPSIS is presented in Figure 4. The figure illustrates the first step that needs to be taken,

TABLE 6
ALTERNATIVE DECISION RULES

No.	Criterion Weight	Value 1	Value 2	Value 3
1	Very Low	0.0	0.0	2.0
2	Low	0.0	2.0	4.0
3	Enough	2.0	4.0	6.0
4	Good	4.0	6.0	8.0
5	Very Good	6.0	8.0	10.0

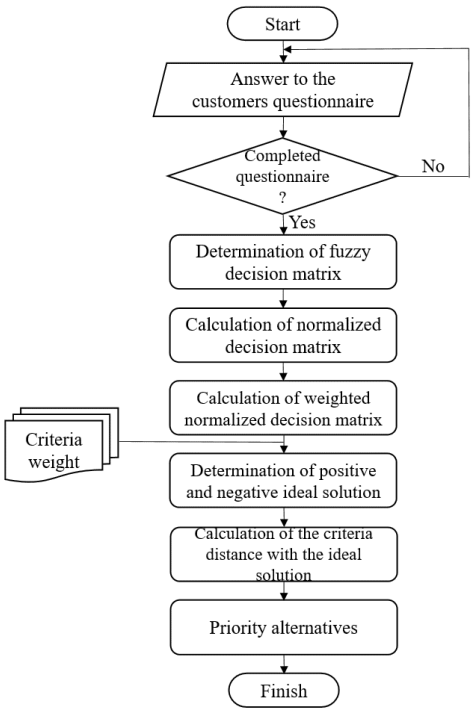


Fig. 4. Flowchart proposed method of Fuzzy TOPSIS

which is to assign a triangular fuzzy number value for each visitor's response. The data from the visitor questionnaires will then be compared with the opinions of three experts in the hotel service industry. The subsequent step involves creating a fuzzy decision matrix, which is later transformed into a normalized decision matrix. These results are further multiplied by the criteria weights, resulting in a weighted normalized decision matrix. The values for positive and negative ideal solutions are then calculated. From these calculations, the distance between each alternative and the ideal point can be determined. The distance between the alternatives and the ideal point represents the final result in the form of a preference value.

C. Method Design

The data, which has been transformed into alternative rules, will be evaluated and converted into a matrix using the predetermined criteria weights as references. Subsequently, the data is processed to generate a matrix by dividing the count of each criterion by the total number of respondents. The formula for this calculation is presented as in (2).

$$\frac{\sum value\ of\ all\ the\ respondent}{\sum respondent} \quad (2)$$

After completing all the calculations, a fuzzy decision matrix and a matrix divisor are obtained. The matrix divisor is determined by taking the square root of each criterion raised to the power of two, resulting in a fuzzy decision matrix. Each alternative is assigned a value for each criterion. The fuzzy decision matrix is subsequently transformed into a normalized decision fuzzy matrix by dividing it by the matrix divisor. Once the fuzzy decision matrix is obtained, the next step involves converting the matrix into a normalized decision matrix. This is achieved by dividing each weight assigned to each alternative for each criterion by the corresponding matrix divisor. The subsequent step involves converting the matrix into a normalized decision matrix by dividing each weight assigned to each alternative for each criterion by the matrix divisor. Then, the normalized decision matrix is further transformed into a weighted decision matrix by multiplying each weight with the corresponding criterion weight that was determined during the criteria weighting stage.

The subsequent step after obtaining the weighted normalized decision matrix is to calculate the positive ideal solution and the negative ideal solution. The positive ideal solution is determined by identifying the alternative with the highest value across all criteria, while the negative ideal solution is determined by identifying the alternative with the lowest value across all criteria.

Once the alternative's distance to the ideal points is calculated, the final result, which is the preference value, can be determined. The alternative's distance to the positive ideal solution is in (3), (4), and (5):

$$D_i^+ = \sqrt{\sum_{j=1}^{n}\left(y_{ij} + y_i\right)^2} \quad i = 1, 2, \ldots, m \quad (3)$$

The alternative's distance to the negative ideal solution is formulated as follows:

$$D_i^- = \sqrt{\sum_{j=1}^{n}\left(y_{ij} - y_i\right)^2} \quad i = 1, 2, \ldots, m \quad (4)$$

To calculate the preference value for each alternative, the following formulation is used:

$$CC_i = \frac{D_i^-}{D_i^- + D_i^+} \quad i = 1, 2, \ldots, m \quad (5)$$

III. Result and Discussion

This study utilizes questionnaire data collected from customers, resulting in a total of 150 data points. Additionally, the responses provided by three hospitality experts were used for comparison with the customers' answers. The customer data, collected successfully from the respondents, is processed using

TABLE 7
FUZZY DECISION MATRIX

Alt	Criterion 1			Criterion 2			Criterion 3			Criterion 4			Criterion 5			Criterion 6		
1	5.36	7.36	9.36	5.20	7.20	9.20	4.56	6.56	8.56	4.48	6.48	8.48	5.52	7.52	9.52	3.52	3.52	7.52
2	4.72	6.72	8.72	4.08	6.08	8.08	4.16	6.16	8.16	3.60	5.60	7.60	4.48	6.48	8.48	3.52	3.52	7.52
3	2.08	4.08	6.08	3.28	5.28	7.28	3.36	5.36	7.36	3.28	5.28	7.28	2.64	4.40	6.40	5.12	5.12	7.12
4	5.28	7.28	9.28	5.68	7.68	9.68	4.80	6.80	8.80	4.08	6.08	8.08	5.44	7.44	9.44	5.76	5.76	7.76
5	4.40	6.40	8.40	4.16	6.16	8.16	3.20	5.20	7.20	3.36	5.36	7.36	4.24	6.24	8.24	5.70	5.70	7.60
6	2.80	4.80	6.80	2.88	4.88	6.88	3.28	7.28	5.28	2.80	4.80	6.80	3.52	5.52	7.52	4.72	4.72	6.72

TABLE 8
NORMALIZED DECISION MATRIX

Alt	Criterion 1			Criterion 2			Criterion 3			Criterion 4			Criterion 5			Criterion 6		
1	0.51	0.48	0.47	0.49	0.47	0.45	0.47	0.45	0.44	0.50	0.47	0.45	0.51	0.48	0.47	0.42	0.42	0.42
2	0.45	0.44	0.43	0.38	0.39	0.40	0.43	0.42	0.42	0.40	0.41	0.41	0.41	0.42	0.42	0.42	0.42	0.42
3	0.20	0.27	0.30	0.31	0.34	0.35	0.34	0.36	0.35	0.37	0.38	0.39	0.24	0.28	0.31	0.38	0.39	0.39
4	0.50	0.48	0.46	0.54	0.50	0.50	0.50	0.47	0.45	0.46	0.44	0.43	0.50	0.48	0.46	0.45	0.44	0.43
5	0.42	0.42	0.42	0.39	0.40	0.33	0.33	0.36	0.37	0.38	0.39	0.39	0.39	0.40	0.40	0.43	0.42	0.42
6	0.27	0.27	0.34	0.27	0.32	0.34	0.34	0.36	0.38	0.31	0.35	0.36	0.32	0.35	0.37	0.33	0.36	0.37

TABLE 9
WEIGHTED NORMALIZED DECISION MATRIX

Alt	Criterion 1			Criterion 2			Criterion 3			Criterion 4			Criterion 5			Criterion 6		
1	0	0.07	0.16	0.25	0.33	0.41	0.19	0.27	0.35	0.10	0.19	0.27	0.15	0.24	0.33	0	0.02	0.10
2	0	0.07	0.15	0.19	0.28	0.36	0.17	0.25	0.34	0.08	0.16	0.24	0.12	0.21	0.29	0	0.02	0.10
3	0	0.04	0.11	0.15	0.24	0.32	0.14	0.22	0.30	0.07	0.15	0.23	0.07	0.14	0.22	0	0.02	0.10
4	0	0.07	0.16	0.27	0.35	0.43	0.20	0.28	0.36	0.09	0.18	0.26	0.15	0.24	0.32	0	0.02	0.10
5	0	0.06	0.15	0.20	0.28	0.36	0.13	0.21	0.30	0.08	0.16	0.24	0.12	0.20	0.28	0	0.02	0.11
6	0	0.05	0.12	0.14	0.22	0.31	0.14	0.22	0.30	0.06	0.14	0.22	0.10	0.18	0.26	0	0.02	0.09

TABLE 10
POSITIVE IDEAL POINT AND NEGATIVE IDEAL POINT

	Criterion 1			Criterion 2			Criterion 3			Criterion 4			Criterion 5			Criterion 6		
FPIS	0	0.07	0.16	0.25	0.33	0.41	0.19	0.27	0.35	0.10	0.19	0.27	0.15	0.24	0.33	0	0.02	0.10
FNIS	0	0.07	0.15	0.19	0.28	0.36	0.17	0.25	0.34	0.08	0.16	0.24	0.12	0.21	0.29	0	0.02	0.10

a fuzzy decision matrix as depicted in Table 7, enabling further analysis in subsequent steps. Moreover, the assessment of the hotel service expert regarding the existing facilities at the subject in is presented in**.**

Table 7 displays the fuzzy decision matrix, which was derived by transforming the respondents' answer matrix into a fuzzy decision matrix. Subsequently, the fuzzy decision matrix is converted into a normalized decision fuzzy matrix by dividing it by the matrix divisor.

After transforming the questionnaire responses into a fuzzy decision matrix, the next step is to calculate the normalized decision matrix. This involves dividing the values in the fuzzy decision matrix by the matrix divisor. The resulting normalized decision matrix is presented in Table 8. The table presents the normalized decision matrix, which is obtained by dividing each weight of each alternative in each criterion by the matrix divisor. Following that, the normalized decision matrix is multiplied by the predetermined criteria weights to produce the weighted normalized decision matrix, as depicted in Table 9.

Next, the positive ideal point and negative ideal point are determined, as illustrated in Table 10. In the table, FPIS represents the positive ideal point, while FNIS represents the negative ideal point. The positive ideal point is the highest value among all values in each criterion column, while the negative ideal point is the lowest value among all values in each criterion column.

After obtaining the positive ideal point and negative ideal point from the alternative matrix multiplication table against the next criteria, the next step is to determine the separation measures or the distance of each alternative to the positive ideal point and negative ideal point, as shown in Table 11. The table shows the D+ value, which represents the alternative distance to the positive ideal point, and the D- value, which represents the alternative distance to the negative ideal point. Based on these values, the weight or priority of each alternative can be determined. This weight determines the most prioritized alternative for service improvement. The preference values can be seen in Table 12. The table presents the final result of the calculation, which is the preference value. By sorting the preference values from highest to lowest, we can determine the alternative with the highest priority for service improvement. The test results indicate that the cafeteria has the highest priority for service improvement, as it obtained the lowest preference value.

TABLE 11
ALTERNATIVE DISTANCE TO THE IDEAL POINT

Alternative Code	Alternative	D+	D-
A1	Bedroom	0.01	0.10
A2	Bathroom	0.05	0.06
A3	Cafetaria	0.10	0.01
A4	Front-Office	0.01	0.10
A5	Worship Place	0.06	0.05
A6	Parking Area	0.10	0.02

TABLE 12
PREFERENCE VALUE

Alternative Code	Alternative	Preference Value
A1	Bedroom	0.87
A2	Bathroom	0.53
A3	Cafetaria	0.13
A4	Front-Office	0.94
A5	Worship Place	0.44
A6	Parking Area	0.17

After obtaining the results of the system trial with the input, which are the questionnaire responses from the respondents, the accuracy of the system can be calculated to determine its effectiveness. The accuracy calculation involves three steps. First, calculating the priority of service engagement by visitors. Second, evaluating the eligibility of the service. Third, calculating the overall accuracy of the system.

1) Service Improvement Priority: The assessment results of service improvement from the system were compared with the responses obtained from 150 data of customer's questioner answer, as well as the opinions of three hospitality experts. The purpose of this comparison was to determine the level of agreement or discrepancy between the two sets of data. The results of the analysis show that of the 150 data

obtained, 120 data show a value "match" with the results shown by the system, the “cafeteria” is an alternative that requires service improvement. Meanwhile, 30 other data show "not match" with the results shown by the system. Thus, the accuracy of service improvement priorities by customers shows accuracy:

$$\frac{120}{150} \text{ x } 100\ \% = 90\ \%$$

TABLE 13
SERVICE IMPROVEMENT QUALITY TRIALS BY EXPERT

Expert	Alternative	Priority		
		System	Hospitality Expert	Output
Expert-1	Bedroom	5	5	Match
	Bathroom	4	4	Match
	Cafetaria	1	2	Not Match
	Front-Office	6	6	Match
	Worship Place	3	3	Match
	Parking Area	2	1	Not Match
Expert-2	Bedroom	5	5	Match
	Bathroom	4	4	Match
	Cafetaria	1	1	Match
	Front-Office	6	6	Match
	Worship Place	3	3	Match
	Parking Area	2	2	Match
Expert-3	Bedroom	5	5	Match
	Bathroom	4	4	Match
	Cafetaria	1	1	Match
	Front-Office	6	6	Match
	Worship Place	3	3	Match
	Parking Area	2	2	Match

Meanwhile, the data generated by the expert's assessment showed that 16 data were "match" with the results shown by the system (as shown in the Table 13) while only two data showed "not match" with the results shown by the system. Therefore, the accuracy of service improvement priorities by experts shows an accuracy of:

$$\frac{16}{18}\ x\ 100\ \% = 88.89\ \%$$

Measuring the accuracy level of service improvement priority trials shows the final accuracy to be:

$$\text{Service Improvement Priority Accuracy} = \frac{Accuracy\ 1 + Accuracy\ 2}{2} = \frac{80\ \% + 84.45\%}{2}$$
$$= 84.45\ \%$$

1) Service Eligibility: The results of the service eligibility trial of the system were compared with the results of the service eligibility trial from 150 data generated by customers who were respondents and three hospitality experts. The purpose of comparing the two data is to determine the suitability or incompatibility of the data. The eligibility test resulted in two choices: to be repaired or maintained. The output shows that the choice is to be repaired if the preference value is less than 0.50, whereas if the preference value is more than 0.50, then the output shows that the alternative is to be maintained. The results of the analysis show that out of the 150 data obtained, 131 data show a value "match" with the results shown by the system, indicating that the cafeteria is an alternative that requires service improvement. Meanwhile, 19 other data show "not match" with the results shown by the system. Thus, the accuracy of service improvement priorities by customers is:

$$\frac{131}{150} \text{ x } 100\ \% = 8.34\ \%$$

TABLE 14
SERVICE ELIGIBILITY TRIALS BY EXPERT

Expert	Alternative	Eligibility		
		System	Hospitality Expert	Output
Expert-1	Bedroom	Maintained	Maintained	Match
	Bathroom	Maintained	Maintained	Match
	Cafetaria	Repaired	Repaired	Match
	Front-Office	Maintained	Maintained	Match
	Worship Place	Repaired	Maintained	Not Match
	Parking Area	Repaired	Repaired	Match
Expert-2	Bedroom	Maintained	maintained	Match
	Bathroom	Maintained	maintained	Match
	Cafetaria	Repaired	Maintained	Not Match
	Front-Office	Maintained	Maintained	Match
	Worship Place	Repaired	Maintained	Not Match
	Parking Area	Repaired	Repaired	Match
Expert-3	Bedroom	Maintained	Maintained	Match
	Bathroom	Maintained	Maintained	Match
	Cafetaria	Repaired	Repaired	Match
	Front-Office	Maintained	Maintained	Match
	Worship Place	Repaired	Repaired	Match
	Parking Area	Repaired	Repaired	Match

Meanwhile, the data generated by the expert's assessment showed 15 data "match" with the results shown by the system, indicating agreement. While only three data show "not match" with the results shown by the system (as shown in the Table 14). Thus, the accuracy of service improvement priorities by experts is:

$$\frac{15}{18}\ x\ 100\ \% = 83.34\ \%$$

Measurement of the level of service eligibility accuracy shows the final accuracy to be:

$$\text{Serviceability Accuracy} = \frac{Accuracy\ 1 + Accuracy\ 2}{2}$$
$$= \frac{87.34\ \% + 83.34\ \%}{2}$$
$$= 85.34\ \%$$

Based on the conducted tests and the accuracy measurements of each test, the accuracy level of the Fuzzy TOPSIS method when implemented in an application to determine priority scales for improving the quality of hotel business services can be observed. In this study, the criteria weights were determined by

four experts from the management of the research object. These criteria weights serve as a reference for obtaining a weighted normalized decision matrix. Upon obtaining the weighted normalized decision matrix, positive ideal solutions and negative ideal solutions are derived. Subsequently, the distances between the criterion values and the positive and negative ideal solutions are determined.

Based on these distances, the final result is the preference value. The results from the system serve as a basis for measuring the accuracy level by comparing them with the results derived from the data obtained from 150 respondents and three of hospitality experts.

In this trial, a comparison was made between the data obtained from the system's trial and the 150 data collected from customer respondents. The accuracy level was measured by evaluating the agreement between the available data. Based on the test results using the 150 existing data, the accuracy level is determined to be 80% according to the precision formula.

Then, the results of the service improvement priority trial by comparing the data obtained from the system's trial with the data provided by three hospitality experts. The accuracy level was measured by evaluating the agreement between the data from the hospitality experts. Based on the trial results using data from three hospitality experts with 18 available data, an accuracy level of 88.89% was obtained according to the precision formula.

After conducting the service improvement priority test, the second test carried out is the service eligibility trial. The service eligibility trials were performed twice, comparing the system's data with data from both hospitality experts and customers. The first service eligibility test involved comparing the system's trial results with the analysis results obtained from the questionnaire data. The accuracy level was measured by evaluating the agreement between the system's data and the data from the questionnaire. Based on the trial results and the analysis of the questionnaire data, an accuracy level of 87.34% was achieved using the precision formula.

The second service eligibility test is conducted by comparing the system's trial results with the trial results using data from three hospitality experts. The comparison of data from the experts will determine the accuracy level. Based on the test results using data from the three hospitality experts, with a total of 18 available data points, an accuracy level of 83.34% was achieved.

This research succeeded in showing that the results provided by the system built showed the same results as the expert's opinion. This proves that expert opinion is very influential and can be used as a basis for the system being built to have a high level of accuracy. This research has advantages that have not been demonstrated in previous studies like the research we mentioned in the background section. This research's advantage is by comparing the opinions of experts with the results managed by the system. Previous research used expert scores only on triangular fuzzy number values without comparing the final system results with expert opinions regarding the object under study. Several studies show that Fuzzy TOPSIS can work well if used as a method in decision support systems.

IV. CONCLUSIONS

Based on the trials conducted by implementing the Fuzzy TOPSIS method in the application for determining the priority scale to improve the quality of hotel business services, it can be concluded that the implementation of the Fuzzy TOPSIS method is effective in supporting decision-making for service improvement priorities. The high accuracy levels achieved, as measured by precision formulas, further validate the usefulness of the Fuzzy TOPSIS method in this context. This research also indicates that the system successfully yielded results consistent with the opinions of selected experts. This implies that the system built can assist hotel management in identifying areas for service quality improvement.

For future research, we aim to compare the results of our work with other methods and with method developments that have been carried out. This will enable us to construct an even more refined system

ACKNOWLEDGMENT

We would like to express our gratitude to the Management of Rumah Singgah UIN Maulana Malik Ibrahim Malang for providing research data. The interpretations and conclusions presented here do not necessarily reflect the views of these agencies. The views or opinions expressed in this article are those of the writers and should not be construed as the opinions of these agencies. Mention of trade names, vendor names, or commercial products does not constitute endorsement or recommendation by these agencies

REFERENCES

[1] M. Waqanimaravu and C. N. Arasanmi, "Employee training and service quality in the hospitality industry," *J. Foodserv. Bus. Res.*, vol. 23, no. 3, pp. 216–227, 2020, doi: 10.1080/15378020.2020.1724850.

[2] G. Büyüközkan, O. Feyzioğlu, and C. A. Havle, "Intuitionistic fuzzy AHP based strategic analysis of service quality in digital hospitality industry," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 1687–1692, 2019, doi: 10.1016/j.ifacol.2019.11.443.

[3] M. S. Hossain, S. N. Kannan, and S. K. K. Raman Nair, "Factors Influencing Sustainable Competitive Advantage in the Hospitality Industry," *J. Qual. Assur. Hosp. Tour.*, vol. 22, no. 6, pp. 679–710, 2021, doi: 10.1080/1528008X.2020.1837049.

[4] B. J. Ali *et al.*, "Hotel Service Quality: The Impact of Service Quality on Customer Satisfaction in Hospitality," *Int. J. Eng. Bus. Manag.*, vol. 5, no. 3, pp. 14–28, 2021.

[5] N. T. VO, M. Chovancová, and H. T. Tri, "The Impact of E-service Quality on the Customer Satisfaction and Consumer Engagement Behaviors Toward Luxury Hotels," *J. Qual. Assur. Hosp. Tour.*, vol. 21, no. 5, pp. 499–523, 2020, doi: 10.1080/1528008X.2019.1695701.

[6] J. Kanyama, W. Nurittamont, and S. Siripipatthanakul, "Hotel Service Quality and Its Effect on Customer Loyalty : The Case of Ubon Ratchathani, Thailand during COVID-19 Pandemic," *J. Manag. Business, Heal. Educ.*, vol. 1, no. 2, pp. 1–20, 2022.

[7] P. Anabila, L. E. K. Ameyibor, M. M. Allan, and C. Alomenu, "Service Quality and Customer Loyalty in Ghana's Hotel Industry: The Mediation Effects of Satisfaction and Delight," *J. Qual. Assur. Hosp. Tour.*, vol. 23, no. 3, pp. 748–770, 2022, doi: 10.1080/1528008X.2021.1913691.

[8] Google, "Rumah Singgah UIN Maulana Malik Ibrahim," Google Reviews and Ratings. [Online]. Available: https://bit.ly/4akmbrX

[9] Y. A. Unvan, "Financial performance analysis of banks with topsis and fuzzy topsis approaches," *Gazi Univ. J. Sci.*, vol. 33, no. 4, pp. 904–923, 2020, doi: 10.35378/gujs.730294.

[10] J. Li, X. Xu, Z. Yao, and Y. Lu, "Improving service quality with the fuzzy TOPSIS method: A case study of the Beijing rail transit system," *IEEE Access*, vol. 7, pp. 114271–114284, 2019, doi: 10.1109/ACCESS.2019.2932779.

[11] G. Magistra Piche, H. A. Mesta, and K. Kunci, "OMISS Operations Management and Information System Studies x Analysis of Service Quality on Train Passenger's Perception by Applying Fuzzy Topsis Method," *Oper. Manag. Inf. Syst. Stud.*, vol. 1, no. 3, pp. 179–190, 2021, doi: 10.24036/omiss.v1i3.23.

[12] S. Tahanisaz and S. shokuhyar, "Evaluation of passenger satisfaction with service quality: A consecutive method applied to the airline industry," *J. Air Transp. Manag.*, vol. 83, no. December 2019, p. 101764, 2020, doi: 10.1016/j.jairtraman.2020.101764.

[13] C. C. Cheng, Y. Y. Chang, and C. T. Chen, "Construction of a service quality scale for the online food delivery industry," *Int. J. Hosp. Manag.*, vol. 95, no. November 2020, p. 102938, 2021, doi: 10.1016/j.ijhm.2021.102938.

[14] S. Kusumadewi, S. Hartati, A. Harjoko, and R. Wardoyo, *Fuzzy Multi-Attribute Decision Making (Fuzzy MADM)*. Yogyakarta: Graha Ilmu, 2006.

[15] M. M. Salih, B. B. Zaidan, A. A. Zaidan, and M. A. Ahmed, "Survey on fuzzy TOPSIS state-of-the-art between 2007 and 2017," *Comput. Oper. Res.*, vol. 104, pp. 207–227, 2019, doi: 10.1016/j.cor.2018.12.019.

[16] P. Manivel and R. Ranganathan, "An efficient supplier selection model for hospital pharmacy through fuzzy AHP and fuzzy TOPSIS," *Int. J. Serv. Oper. Manag.*, vol. 33, no. 4, pp. 468–493, 2019, doi: 10.1504/IJSOM.2019.101588.

[17] A. Shamsuzzoha, S. Piya, and M. Shamsuzzaman, *Application of Fuzzy TOPSIS Framework for Selecting Complex Project in a Case Company*, vol. 14, no. 3. 2021. doi: 10.1108/JGOSS-07-2020-0040.

[18] D. Ulfiana, Y. E. Windarto, N. Bashit, and N. S. Ristianti, "Analysis of Flood Vulnerability as a Support to Water Sensitive Urban Design Planning in Klaten Regency," *Media Komun. Tek. Sipil*, vol. 26, no. 2, pp. 183–193, 2021, doi: 10.14710/mkts.v26i2.31929.

[19] B. Sahin, T. L. Yip, P. H. Tseng, M. Kabak, and A. Soylu, "An Application of a Fuzzy TOPSIS Multi-Criteria Decision Analysis Algorithm for Dry Bulk Carrier Selection," *Inf.*, vol. 11, no. 5, 2020, doi: 10.3390/INFO11050251.

[20] J. T. de Aquino, F. J. C. de Melo, T. de B. Jerônimo, and D. D. de Medeiros, "Evaluation of Quality in Public Transport Services: The Use of Quality Dimensions as an Input for Fuzzy TOPSIS," *Int. J. Fuzzy Syst.*, vol. 21, no. 1, pp. 176–193, 2019, doi: 10.1007/s40815-018-0524-1.

[21] R. Khambhati, H. Patel, and S. Kumar, "A performance evaluation and comparison model for Urban Public Healthcare Service Quality (UrbPubHCServQual) by Fuzzy TOPSIS Method," *J. Nonprofit Public Sect. Mark.*, vol. 34, no. 3, pp. 291–310, 2022, doi: 10.1080/10495142.2020.1865232.

[22] L. Y. Peng, H. Xu, J. Lu, J. J. Luo, and H. H. Zhan, "An Integrated Approach with CLFPR-Based ANP and Fuzzy TOPSIS for Evaluating Business Performance of Rural Homestays: Study from China," *Sustain.*, vol. 15, no. 1, 2023, doi: 10.3390/su15010538.

[23] "Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 29 Tahun 2022 tentang Pemantauan dan Evaluasi Kinerja Penyelenggaraan Pelayanan Publik."

[24] "Peraturan Menteri Pariwisata dan Ekonomi Kretaif Republik Indonesia Nomor PM.53/HM.001/MPEK/2013 tentang Standar Usaha Hotel."

ISSN 2085-4552

Data Mining Motorbike Sales Classification Using a Combination of K-Means and Naïve Bayes Algorithms

Eka Sofiati[1], Dori Gusti Alex Candra[2], Irzon Meiditra[3]
[1,2] Dept. of Information System, Mitra Gama Institute of Technology, Duri, Indonesia
[3] Dept. of Information Technology, Rokan Hilir Institute of Technology, Rokan Hilir, Indonesia
[1]ekasfti@gmail.com, [2]dorigustialexcandra@gmail.com, [3]meiditrairzon@gmail.com

Accepted 04 October 2024
Approved 08 January 2025

***Abstract*— The showroom, which has been established since 2006 and is located in Lapai, Padang City, has problems, namely the difficulty of analyzing consumer demand and the amount of sales data that accumulates. In addition, many stock items are not available when consumer demand is high. From these problems, a data mining application system is needed to improve sales patterns and process sales data to determine items that are often purchased and not using data mining methods, namely K-Means and Naïve Bayes. The data is obtained directly from CV Unique Motor in the form of motorcycle sales data and motorcycle inventory data. The stages of the research include several stages, namely problem identification, data collection, data analysis, clustering stage analysis, classification stage analysis, system analysis, and system implementation. At the system analysis stage, a design will be carried out using data mining by applying the K-Means and Naive Bayes algorithms, where the program will be executed in the PHP programming language and MySQL database. The calculation results show that the highest probability value is in the many variable, namely 0.0607, so it can be concluded that the sales level is a lot. With this system, it can speed up the showroom in making decisions from the data taken so that the showroom can increase the amount of stock that has a best-selling classification so that the showroom does not run out of stock and can help the showroom classify motorcycle sales and can harmonize the availability and inventory of existing motorbikes by classifying sales volume.**

***Index Terms*— Data Mining; Classification; Sales; K-Means; Naïve Bayes**.

I. INTRODUCTION

CV. Unique Motor is a motorcycle buying and selling showroom, which is the object of the author's research this time. This showroom, which has been established since 2006 and is located in Lapai, Padang City, has a problem, namely the difficulty in analyzing consumer demand and a lot of accumulation of sales data based on motorcycle inventory stock sales transaction data from October 2021 to March 2022, totaling 37 motorcycle data. In addition, many stock items are not available when consumer demand is high. From these problems, a data mining application system is needed to improve sales patterns and process sales data to determine what is often purchased and not using data mining methods, namely K-Means and Naïve Bayes.

Data mining is a step in performing knowledge discovery in databases. Knowledge discovery as a process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation [1]. The KMNB approach is made using a combination of clustering and classification techniques. For data classification based on categories using the Naïve Bayes method. And clustering helps to identify groups that have the same characteristics or show characteristics at the beginning [2]. Clustering is the classification of unsupervised patterns such as observations, data items, and vectors into groups called clusters [3]. The goal is to minimize the objective function set by the clustering process [4]. The basic concept of clustering is to group a large number of objects into a cluster. A good cluster is a cluster that has a high degree of similarity between objects in the cluster [5]. Clustering is a data grouping method that starts by grouping two or more of the most similar objects [6]. K-Means is a non-hierarchical data clustering method that attempts to divide existing data into two or more groups, where data with similar characteristics [2].

Classification is a technique used to predict the class or property of each data instance and classify the data itself based on the topic [7]. Classification performs the process of finding a model or function that describes and characterizes a class of data for a specific purpose [8]. Each classification algorithm used produces a best-fit model that relates the input data to a known classification class [9]. Classification consists of three steps, namely model building, model application, and evaluation [10]. Naïve Bayes is a classification

proposed by British scientist Thomas Bayes and is known as Bayes' theorem because it uses probalistic and statistical methods [11]. Naïve Bayes is a simple algorithm and is used to predict a probability in the membership of a class [12].

Some of the research conducted previously includes the Comparison of Naive Bayes and C4.5 Algorithms in Determining the Sales Level of Honda Motorbikes This research was conducted to determine the sales level of Honda motorbikes using motorcycle sales datasets using data mining [13]. Another study on Product Sales Level Analysis Using K-Nearest Neighbor (KNN) and K-Means explains that high sales require information and identify which products are the most potential and to sell, so that goods in the warehouse do not accumulate as needed [14].

Other research using the K-Means and Naïve Bayes algorithms has also been done before. Research on Determining Livestock Aid Recipients Using K-Means & Naive Bayes Algorithms explains that the combination of classification and clustering is believed to be able to provide accurate classification results when determining data [2]. Another study on Literacy Classification Using K-Means-Naïve Bayes Algorithm explains that to get the optimal number of classes through the clustering process and find out the results of the classification process [15]. The combination of data mining methods in this study hopes to produce accurate classification by combining clustering and classification

II. METHOD

This research proposes a combination of clustering and classification methods. The clustering method is used to obtain several groups of motors. The clustering process is carried out using the K-Means algorithm. After the motorcycle groups are obtained, the classification method is used to obtain the sales classification. The classification process is carried out using the Naïve Bayes algorithm. The research methodology proposed in this study can be seen in Figure 1.

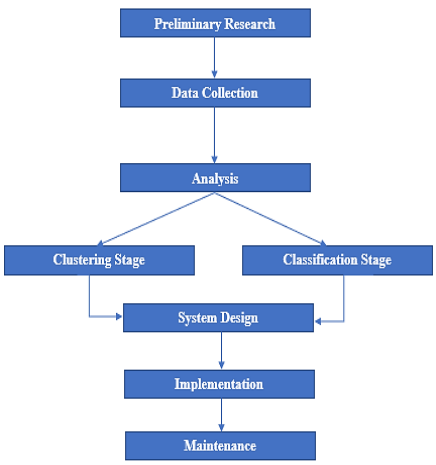


Fig. 1. Research Framework

A. *Preliminary Research*

First, confirm that you have the correct template for your paper size. This template is for ULTIMATICS. It has been tailored for output on the A4 paper size.

B. *Data Collection*

Make direct observations and conduct interviews with motorcycle showrooms and collect literature related to this research.

C. *Analysis*

Analyze the data obtained from the showroom and determine what data mining techniques and methods are used, then analyze the data mining application to be built.

D. *Clustering Stage*

In this clustering stage is the first stage of this research, the clustering process is carried out using the K-Means algorithm. The clustering process begins by determining the number of clusters to be formed, where the motorcycle data will be grouped into clusters, namely “Little Sold” and “Much Sold”. Then determine the initial centroid randomly. Then calculate the distance of each data to each centroid using the Euclidean Distance formula as shown in equation (1).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)}2 \qquad (1)$$

E. *Classification Stage*

The classification stage is the second stage of this research; the classification process is carried out using the Naïve Bayes algorithm. Motorcycle data that has been clustered from the first stage is used as input data for the second stage. The data used is the “sold a lot” result data. The classification process is grouped into several criteria, where these criteria will be the determinant in decision-making. using the Naïve Bayes formula as shown in equation (2).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \qquad (2)$$

F. *Implementation*

To build data mining applications in this study, using the PHP programming language and supported by the MySQL database.

III. RESULTS AND DISCUSSION

The data used in this study is the total inventory data and motorcycle sales for October 2021-March 2022. The sample of motorcycle data that is the object of research is as in table 1 below:

TABLE I. MOTORCYCLE INVENTORY AND SALES DATA

No	Motor Brand	Total Inventor	Total Sales
1	Beat	51	29
2	CB 150 R	9	4
3	PCX	1	1
4	Supra X	10	9
5	Vario 125	12	11
6	Vario CW	6	1
7	Vario fi	7	4
8	Satria FU	10	2
9	Mio	4	3
10	Mio J	5	4
….	….	….	….
…	…	….	….
37	Mio CW	1	0

A. K-Means

The first step is to determine the cluster center or centroid which is obtained randomly or randomly for the initial determination of the cluster assumed:

TABLE II. INITIAL CLUSTER CENTER

Motor Brand	Total Inventor	Total Sales	Claster
NMax	6	5	C1
Revo	3	2	C2

Next, the distance of each data with the cluster center is calculated using the euclidien distance formula (equation 1) from all data to each first center point:

$d_{11} = \sqrt{(51-6)^2} + \sqrt{(29-5)^2} = 51$
$d_{12} = \sqrt{(9-6)^2} + \sqrt{(4-5)^2} = 3{,}1623$
$d_{13} = \sqrt{(1-6)^2} + \sqrt{(1-5)^2} = 6{,}4031$
$d_{14} = \sqrt{(10-6)^2} + \sqrt{(9-5)^2} = 5{,}6569$
$d_{15} = \sqrt{(12-6)^2} + \sqrt{(11-5)^2} = 8{,}4853$
……..
……..
$d_{137} = \sqrt{(1-6)^2} + \sqrt{(0-5)^2} = 7{,}0711$

In the same way calculate the distance of each 2nd center point and we will get:

$d_{21} = \sqrt{(51-3)^2} + \sqrt{(29-2)^2} = 55{,}0727$
$d_{22} = \sqrt{(9-3)^2} + \sqrt{(4-2)^2} = 6{,}3246$
$d_{23} = \sqrt{(1-3)^2} + \sqrt{(1-2)^2} = 2{,}2361$
$d_{24} = \sqrt{(10-3)^2} + \sqrt{(9-2)^2} = 9{,}8995$
$d_{25} = \sqrt{(12-3)^2} + \sqrt{(11-2)^2} = 12{,}7279$
……
……...
$d_{237} = \sqrt{(1-3)^2} + \sqrt{(0-2)^2} = 2{,}8284$

The second is to calculate the new center point of the 1st iteration. By calculating the mean of cluster member data from each cluster, the result of the cluster 2 center point value is:

TABLE III. NEW CENTROID CENTER

C1	12,9091	8,2727
C2	1,9615	1,3462

If in the second iteration no cluster member moves to another cluster then the iteration stops but if any cluster member moves to another cluster then proceed to the next iteration. In this study the iteration stops until the 8th iteration. Here are the final results of the cluster calculation at the 8th iteration:

TABLE IV. NEW CENTROID CENTER

No	Motor Brand	C1	C2	Cluster
1	Beat	0	53,9093	C1
2	CB 150 R	48,8774	5,2214	C2
3	PCX	57,3062	3,3972	C2
4	Supra X	45,6180	8,7424	C2
5	Vario 125	42,9535	11,5704	C2
6	Vario CW	35	2,6639	C2
7	Vario fi	50,6063	3,3228	C2
8	Satria FU	49,0918	6,0952	C2
9	Mio	53,7122	0,3106	C2
10	Mio J	52,3546	1,6789	C2
….	….	….	….	….
…	…	….	….	….
37	Mio CW	57,8014	3,9912	C2

Furthermore, the data that will be forwarded to continue the Naive Bayes classification is C2, namely “Many Sold”.

B. Naïve Bayes

The decision-making process in the classification of motorcycle sales will be grouped into 5 criteria, where these criteria will determine decision-making in the motorcycle showroom. These criteria are:

1) Motor Type

TABLE V. CLASSIFICATION OF MOTOR TYPES

Motor Type	Honda
	Yamaha
	Suzuki
	Kawasaki

2) Color

TABLE VI. COLOR CLASSIFICATION

Monoton	Hitam, Putih, Biru Merah
Variasi	Putih, Biru, Merah Putih
	Hitam, Putih, Biru Merah

3) Year

TABLE VII. YEAR CLASSIFICATION

2021 Down	2010-2021
2006 Above	2006-2017

4) Category

TABLE VIII. YEAR CLASSIFICATION

Category	Matic
	Manual

5) Sales

Sedikit	<=2
Banyak	>=3

The training data used for classification amounted to 36, where the training data will then be processed using the Naïve Bayes method. The training power used can be seen in the table:

TABLE IX. TRAINING DATA

Brand	Type	Color	Year	Category	Sales
CB 150 R	Honda	Variasi	2021 Kebawah	Manual	Banyak
PCX	Honda	Monoton	2006 Keatas	Matic	Sedikit
Supra X	Honda	Variasi	2006 Keatas	Manual	Banyak
Vario 125	Honda	Variasi	2021 Kebawah	Matic	Banyak
Vario CW	Honda	Monoton	2006 Keatas	Matic	Sedikit
Vario fi	Honda	Varasi	2021 Kebawah	Matic	Banyak
Satria FU	Suzuki	Variasi	2006 Keatas	Manual	Sedikit
Mio	Yamaha	Monoton	2006 Keatas	Matic	Banyak
Mio J	Yamaha	Monoton	2021 Kebawah	Matic	Bannyak
Mio S	Yamaha	Monoton	2021 Kebawah	Matic	Sedikit
NMAX	Yamaha	Monoton	2021 Kebawah	Matic	Banyak
Vixion	Yamaha	Monoton	2006 Keatas	Manual	Banyak
Kharisma	Honda	Variasi	2006 Keatas	Manual	Sedikit
Supra X	Honda	Variasi	2006 Keatas	Manual	Sedikit
Vario 150	Yamaha	Variasi	2006 Keatas	Matic	Banyak
Spin	Suzuki	Variasi	2006 Keatas	Matic	Sedikit
Freego	Yamaha	Monoton	2021 Kebawah	Matic	Sedikit
Mio GT	Yamaha	Variasi	2006	Matic	Sedikit

Xeon	Yamaha	Variasi	2006 Keatas	Matic	Banyak
Beat Street	Honda	Monoton	2021 Kebawah	Manual	Sedikit
GSX	Suzuki	Variasi	2021 Kebawah	Manual	Sedikit
R15	Yamaha	Monoton	2006 Keatas	Manual	Sedikit
Soul GT	Yamaha	Monoton	2006 Keatas	Matic	Sedikit
Revo	Honda	Monoton	2006 Keatas	Manual	Sedikit
Fino	Yamaha	Variasi	2006 Keatas	Matic	Sedikit
Mio Soul	Yamaha	Monoton	2021 Kebawah	Matic	Sedikit
Scoopy	Honda	Variasi	2021 Kebawah	Matic	Banyak
Ninja SS	Kawasaki	Variasi	2006 Keatas	Manual	Sedikit
Ninja RR	Kawasaki	Variasi	2006 Keatas	Manual	Sedikit
MX King	Yamaha	Variasi	2006 Keatas	Manual	Sedikit
Mega Pro	Honda	Monoton	2006 Keatas	Manual	Sedikit
Supra FIT	Honda	Variasi	2006 Keatas	Manual	Sedikit
Mio Z	Yamaha	Monoton	2006 Keatas	Matic	Sedikit
Genio	Honda	Monoton	2021 Kebawah	Matic	Sedikit
Mioo M3	Yamaha	Monoton	2021 Kebawah	Matic	Sedikit
Mio CW	Yamaha	Variasi	2006 Keatas	Matic	Sedikit
Vario 150	Yamaha	Variasi	2006 Keatas	Matic	Banyak

In solving using the Naïve Bayes method if the sales level is not known as the testing data below:

TABLE X. DATA TESTING

Merek	Jenis	Warna	Tahun	Kategori	Penjualan
Mio	Yamaha	Variasi	2021 Kebawah	Matic	???

The first is to calculate the probability value of each class:

P(Penjualan = “Sedikit”) = 24/36 = 0,6667

P(Penjualan = “banyak”) = 12/36 = 0,3333

P(Jenis = Honda | Penjualan = “Sedikit”)

= 8/24 = 0,6667

P(Jenis = Honda | Penjualan = “Banyak”)

= 7/12 = 0,375

P(Warna = Variasi | Penjualan = “Sedikit”)

= 12/24 = 0,5417

P(Warna = Variasi | Penjualan = “Banyak”)

= 6/12 = 0,5833

P(Tahun = 2021 Kebawah | Penjualan = “Sedikit”)

= 6/24 = 0,2917

P(Tahun = 2021 Kebawah | Penjualan = “Banyak”)

= 8/12 = 0,75

P(Kategori = Matic | Penjualan = “Sedikit”)

= 9/12 = 0,8333

P(Kategori = Matic | Penjualan = “Banyak”)

= 13/24 = 0,5833

Selanjutnya kalikan semua variable Sedikit dan Banyak

dengan kelas yang sama:

P(Penjualan = “Sedikit”) X P(Penjualan = “Sedikit”)

= 0,6667 x 0,6667 x 0.5417 x 0,2917 x 0,8333

= 0,0307

P(Penjualan = “Banyak”) X P(Penjualan = “Banyak”)

= 0,3333 x 0,375 x 0.5833 x 0,75 x 0,5833

= 0,0607

From the results of the above calculations, it can be seen that the highest probability value is in the many variable so it can be concluded that the sales level is a lot. System Interface Testing Before entering the system, the admin must input the username and password in the database. The system then validates the username and password entered by the admin.

C. Design Implementation

at this stage is the result of the design

1) Alternative Page View

The alternative page functions to add motorcycle brands that will be clustered. On this page, the admin can view and edit motorcycle brand data.

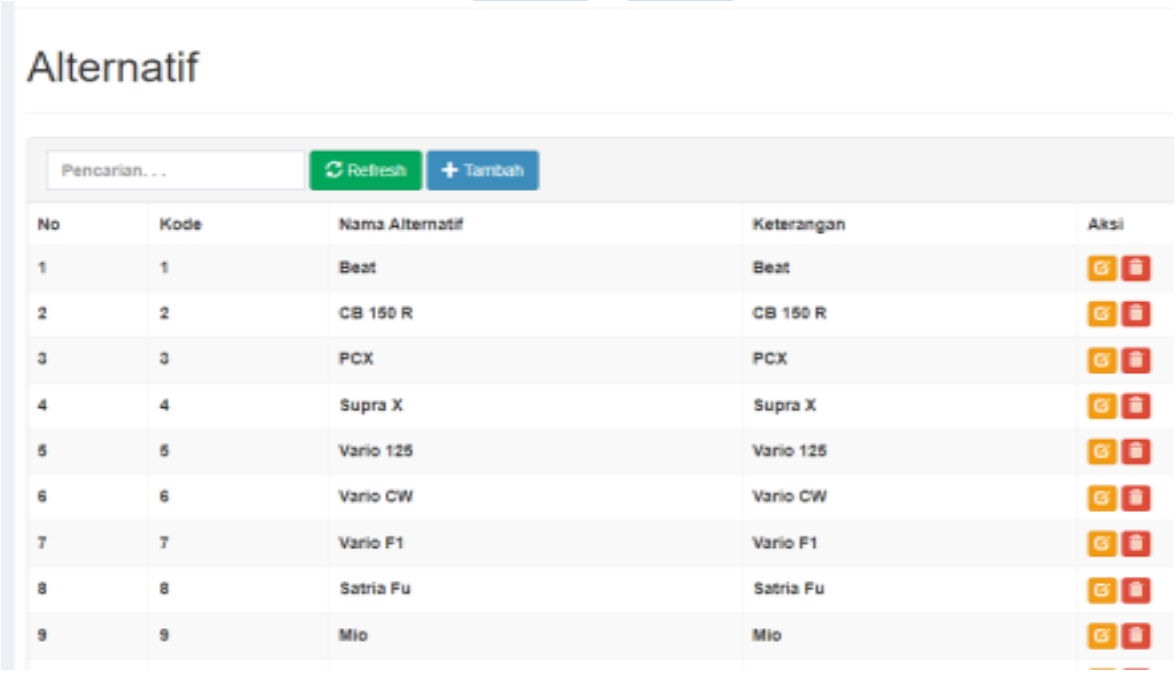
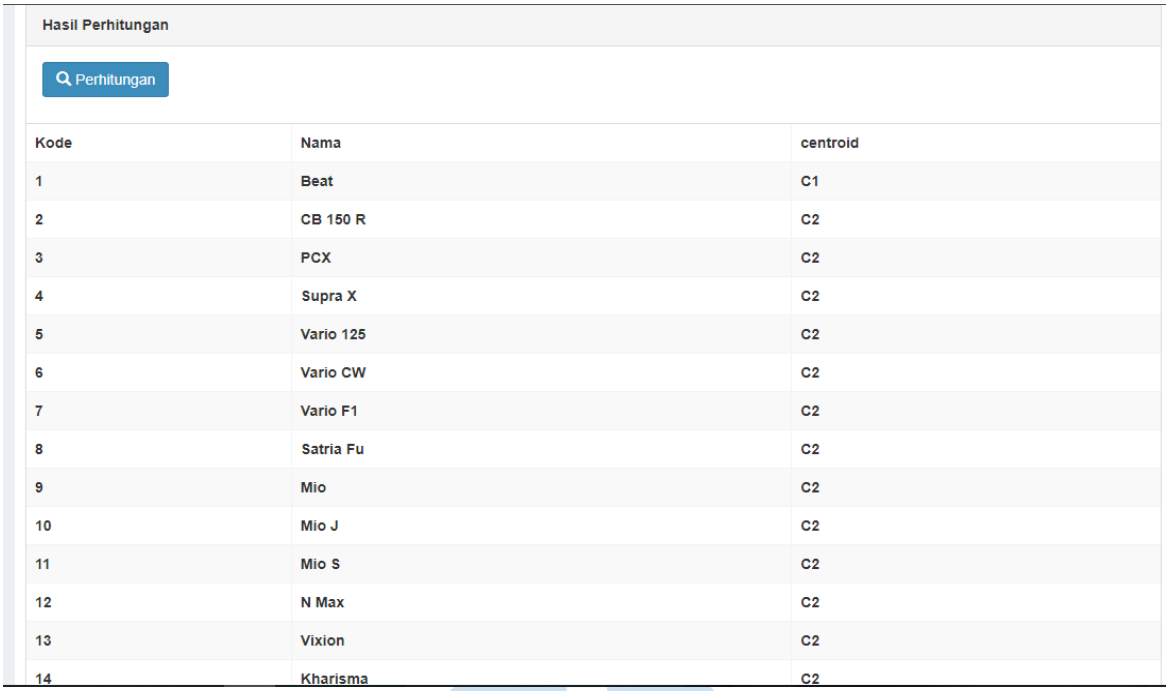


Fig. 2. Alternate Page View Page

2) Calculation Result Page Display

The page displays the training data page that will be used for the Naïve Bayes calculation process, on this page the admin can edit the categories in the dataset and can delete data in the dataset.



Hasil Perhitungan

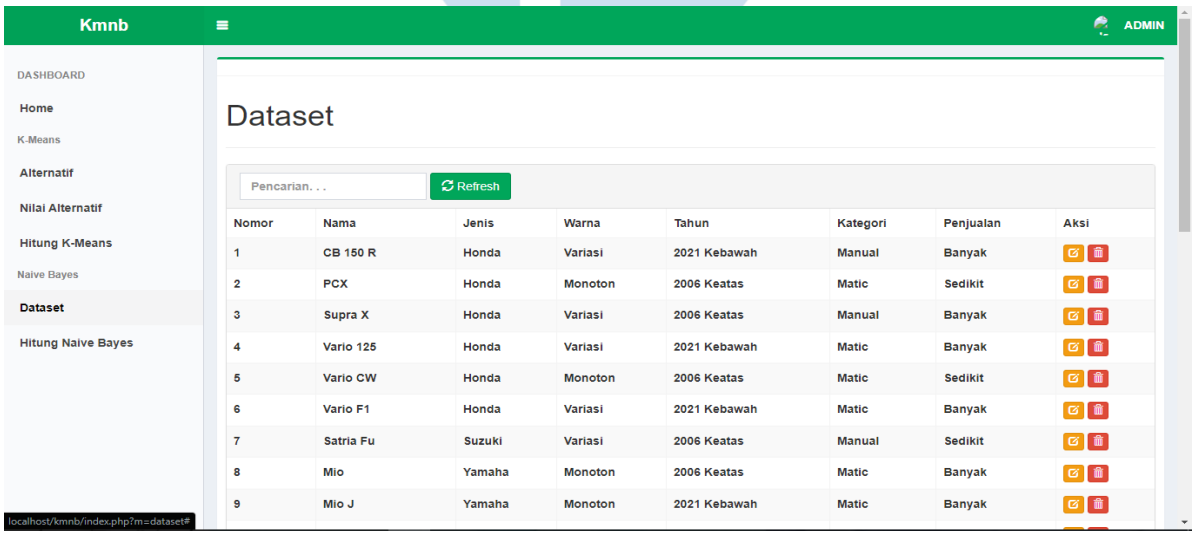
Perhitungan

Kode	Nama	centroid
1	Beat	C1
2	CB 150 R	C2
3	PCX	C2
4	Supra X	C2
5	Vario 125	C2
6	Vario CW	C2
7	Vario F1	C2
8	Satria Fu	C2
9	Mio	C2
10	Mio J	C2
11	Mio S	C2
12	N Max	C2
13	Vixion	C2
14	Kharisma	C2

Fig. 3. *Calculation Result Page Display*

3) Alternative Page View

This page displays the selection of data that will be used as test data for the Naïve Bayes process. displays the selection of data that will be used as test data for the Naïve Bayes process.



Kmnb

ADMIN

DASHBOARD
Home
K-Means
Alternatif
Nilai Alternatif
Hitung K-Means
Naive Bayes
Dataset
Hitung Naive Bayes

Dataset

Pencarian. . .

Refresh

Nomor	Nama	Jenis	Warna	Tahun	Kategori	Penjualan	Aksi
1	CB 150 R	Honda	Variasi	2021 Kebawah	Manual	Banyak	
2	PCX	Honda	Monoton	2006 Keatas	Matic	Sedikit	
3	Supra X	Honda	Variasi	2006 Keatas	Manual	Banyak	
4	Vario 125	Honda	Variasi	2021 Kebawah	Matic	Banyak	
5	Vario CW	Honda	Monoton	2006 Keatas	Matic	Sedikit	
6	Vario F1	Honda	Variasi	2021 Kebawah	Matic	Banyak	
7	Satria Fu	Suzuki	Variasi	2006 Keatas	Manual	Sedikit	
8	Mio	Yamaha	Monoton	2006 Keatas	Matic	Banyak	
9	Mio J	Yamaha	Monoton	2021 Kebawah	Matic	Banyak	

localhost/kmnb/index.php?m=dataset#

Fig. 4. Dataset Page Display Page

4) Display of Naive Bayes Result Page

This page is the result of the Naïve Bayes calculation process. Naïve Bayes result page display

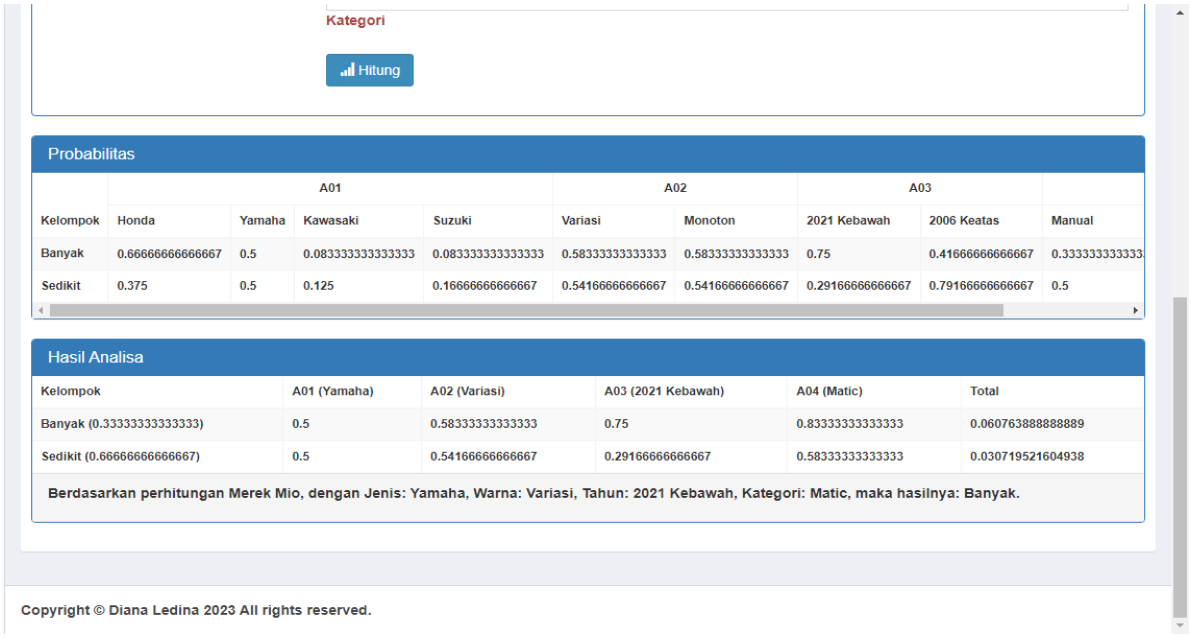


Fig. 5. Naïve Bayes Result Page Display Page

IV. CONCLUSIONS

Based on the total motorcycle inventory and sales data for October 2021 - March 2022 used in this study using the K-means algorithm, it shows that the 8th iteration with the updated cluster center has not changed, so the iteration process is stopped. So the number of clusters to be formed at this stage is two clusters of 37 total inventory and sales data. The clusters to be formed are C1, namely "Little Sold" and C2, namely "Much Sold". Furthermore, the data that will be forwarded to continue the Naive Bayes classification is C2, namely "Sold a lot". From the calculation results it can be seen that the highest probability value is in the many variable which is 0.0607 so it can be concluded that the sales level is a lot. The results of the implementation of the data mining system application are proven to be able to classify sales. This is evidenced by the data on motorcycle inventory and sales that can classify motorcycle sales with the most sales. And the existence of this system is able to solve the problem of empty motorcycle availability when there is a request from the buyer.

REFERENCES

[1] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.

[2] M. F. Asikin, D. E. Ratnawati, and M. A. Fauzi, "Penentuan Penerima Bantuan Ternak Menggunakan Algoritma K-Means & Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 2, pp. 641–649, 2018, [Online]. Available: http://download.garuda.kemdikbud.go.id/article.php?article=871445%5C&val=10384%5C&title=Penentuan Penerima Bantuan Ternak Menggunakan Algoritma K-Means Nave Bayes

[3] A. Muni, "Analisis Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Penjualan Sepeda Motor Studi Kasus PT. Alfa Scorpii," *Juti Unisi*, vol. 4, no. 1, pp. 1–8, 2020, doi: 10.32520/juti.v4i1.1087.

[4] H. Priyatman, F. Sajid, and D. Haldivany, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, pp. 62–66, 2019.

[5] Zulrahmadi, S. Defit, and Y. Yunus, "Pemetaan Wilayah Potensial Terhadap Penjualan Sepeda Motor Honda Menggunakan K-Means Clustering," *J. Inform. Ekon. Bisnis*, vol. 2, no. 2018, pp. 53–59, 2020, doi: 10.37034/infeb.v2i2.41.

[6] Haditsah Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Luxor Variasi Gorontalo)," *J. Inform. Upgris*, vol. 5, no. 1, pp. 40–45, 2019.

[7] M. P. Tana, F. Marisa, and I. D. Wijaya, "Penerapan Metode Data Mining Market

Basket Analysis Terhadap Data Penjualan Produk Pada Toko Oase Menggunakan Algoritma Apriori," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 3, no. 2, pp. 17–22, 2018, doi: 10.37438/jimp.v3i2.167.

[8] P. I. Lestari, D. E. Ratnawati, and L. Muflikhah, "Implementasi Algoritme K-Means Clustering Dan Naive Bayes Classifier Untuk Klasifikasi Diagnosa Penyakit Pada Kucing," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 968–973, 2019.

[9] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

[10] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.

[11] D. Rezekika, "Penerapan Algoritma Naïve Bayes Untuk Memprediksi Penjualan Spare Part Sepeda Motor," *J. Pelita Inform.*, vol. 8, no. 3, pp. 326–329, 2020.

[12] N. E. Akbar, P. P. Adikara, and R. C. Wihandika, "Analisis Sentimen Layanan Astra Honda Motor Menggunakan Metode Naïve Bayes dan Identifikasi Aspek pada Layanan Menggunakan DBSCAN," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 929–937, 2020, [Online]. Available: http://j-ptiik.ub.ac.id

[13] D. Prabowo, F. Hidayat, G. Gumelar, D. Qintoro, and A. Setiawan, "PERBANDINGAN ALGORITMA NAÏVE BAYES DAN C4.5 DALAM MENENTUKAN TINGKAT PENJUALAN MOTOR HONDA," *J. Inform. Komputer, Bisnis dan Manaj.*, vol. 16, no. 3, pp. 67–76, Nov. 2023, doi: 10.61805/fahma.v16i3.91.

[14] Herianto, N. Syamsiyah, and S. Anggraini, "ANALISA TINGKAT PENJUALAN PRODUK MENGGUNAKAN K-NEAREST NEIGHBOR (K-NN) DAN K-MEANS (STUDI KASUS PERUSAHAAN KAYU ELANG PERKASA)," *J. Sains Teknol.*, vol. XI, no. 1, pp. 8–18, 2021.

[15] M. F. A. Saputra, T. Widiyaningtyas, and A. P. Wibawa, "Illiteracy Classification Using K Means-NaÃ¯ve Bayes Algorithm," *JOIV Int. J. Informatics Vis.*, vol. 2, no. 3, pp. 153–158, May 2018, doi: 10.30630/joiv.2.3.129.

ISSN 2085-4552

Application of Deep Learning Techniques for Enhancing Arabic Vocabulary Acquisition in Students at MTs Darun-Najah

Misbachur Rohmatul Isnaini[1], Arvita Agus Kurniasari[2], Aji Seto Arifianto[3], Pramuditha Shinta Dewi Puspitasari[4]
[1,2,3,4]Dapartment of Informatics Technology, State Polytechnic of Jember, Jember, Indonesia
[1]aninisnaini214@gmail.com, [2]arvita@polije.ac.id, [3]ajiseto@polije.ac.id, [4]pramuditha@polije.ac.id

Accepted 29 October 2024
Approved 08 January 2025

***Abstract*— Arabic vocabulary recognition is an important aspect of learning at MTs Darun - Najah, a school that emphasizes on Islamic religious education. This research proposes the application of Convolutional Neural Network (CNN) and EfficientNet B7 to create learning media for Arabic vocabulary recognition for students. This method is implemented in the form of a web-based application. The built application offers an innovative approach in learning by utilizing deep learning. The results of several trials conducted showed that the application of Convolutional Neural Network (CNN) and EfficientNet B7 achieved 90% accuracy with an average precision of 94.6%, recall 94.6%, and f1-score 94.6%. Tests using User Acceptence Testing (UAT) have a success accuracy rate of 87.2% which proves that users can accept quite well.**

***Index Terms*— *Convolutional Neural Network (CNN); EfficientNet B7; Arabic Vocabulary Recognition; Deep Learning*.**

I. INTRODUCTION

Language is a basic need for humans that enables the conveyance of ideas, thoughts, and ideas in oral and written form. Language has a very important role in human life because it facilitates effective communication between individuals. With language, humans can interact, share information, and build better social relationships [1]. In this world there are various kinds of languages, one of which is Arabic. Arabic has a special significance in Islam because it is the language of the Qur'an, the holy book of Muslims. Therefore, understanding Arabic is important for Muslims to understand and practice Islamic teachings, because much Islamic literature is written in Arabic [2].

Vocabulary is a collection of words in a language that is known and understood by humans. Vocabulary is an important part of language and communication skills, because with sufficient vocabulary, a person can more easily convey and understand messages in daily communication. In the context of foreign language learning, such as Arabic, vocabulary mastery is very influential on listening, speaking, reading and writing skills [3]. Therefore, understanding Arabic vocabulary is essential in learning this language, so that students can more easily understand the subject matter and communicate more effectively.

MTs Darun - Najah is one of the Islamic schools that includes Arabic language in its learning curriculum. In this school, in addition to providing Arabic learning materials, students are also trained to speak the language. This aims to make students not only understand the theory of Arabic, but also be able to apply it in everyday life. However, there are challenges in teaching Arabic at MTs Darun - Najah, especially related to student interest. Many students experience difficulties in learning Arabic due to monotonous teaching methods, which can make students quickly get bored and pay less attention to the teacher when teaching [4].

Teacher creativity in delivering material is a key factor in maintaining student interest. Teachers who are able to create a fun and interactive learning atmosphere will be more successful in teaching Arabic. The use of interactive learning media can help the teaching-learning process become more interesting and effective, so that students are more motivated to learn [5]. To overcome the challenges in learning Arabic at MTs Darun - Najah, technology development is a relevant solution.

A website has been developed using Deep Learning technology, specifically Convolutional Neural Network (CNN), to recognise Arabic vocabulary. This technology allows the recognition and classification of Arabic vocabulary from various objects, so that it can help students in learning and understanding Arabic better. This research aims to develop a learning media system for Arabic vocabulary recognition and implement CNN model for Arabic vocabulary classification at MTs Darun - Najah. The benefits of this research include creating Arabic vocabulary learning media for MTs Darun - Najah and knowing the Arabic vocabulary of an object using the Convolutional Neural Network method. With this technology, it is hoped that the Arabic language teaching and learning process at

MTs Darun - Najah can be more effective and fun, and can improve students' ability to understand and use Arabic.

II. THEORY

A. *Arabic Language*

Arabic has privileges over other languages, one of which is Arabic is an eternal language because Arabic is the language of the Qur'an which will last until the Last Day [6]. Arabic is a religious language, because Arabic is always used in various religious activities, such as prayer, sermons, weddings, and so on. In addition, Arabic is a unifying language for Muslims around the world, with the development of the times, Arabic is used as an international language. Arabic is also a pioneer of civilisation because Arabic was used internationally for 8 centuries, so many Arabic words are borrowed in Indonesian [7].

Arabic is one of the oldest and most famous languages in the world due to its large number of mufrodat (vocabulary), lafadz and their meanings and being one of the languages capable of understanding clear meanings. Arabic is also called lughotud dhot, which is a flowing language, whose scope is very broad and its explanation is clear [8]. Arabic is a Semitic language because Arabic has the same origins as other Semitic languages, such as Hebrew, Aramaic, and Akkadian. These Simitic languages are commonly found in the Middle East and North Africa.

B. *Vocabulary*

Vocabulary is the main component before learning foreign languages [9]. In language, vocabulary refers to all the words that can be understood and used by speakers of a language that can be understood and used by speakers of a language. Vocabulary includes nouns, verbs, adjectives, adverbs, pronouns, and so on. Vocabulary also includes vocabulary variations such as formal and informal words, colloquial words and technical language words over terms.

Vocabulary is a set that contains all the new words that a person will later use to construct a new sentence. The amount of vocabulary one has shows the level of education or learning one has mastered. One language expert named Horn said that vocabulary is a collection of words that can later form a language. A person can understand the four language skills if he has mastered vocabulary well [9]. Vocabulary is very important in language skills, because by understanding the right vocabulary, one can convey messages clearly and effectively. In addition, having a rich vocabulary can also help a person to understand readings and conversations more easily and quickly.

C. *Deep Learning*

Deep Learning is a branch of machine learning that adopts algorithms inspired by the structure of the human brain [10]. Deep Learning is used to solve various problems, such as face recognition, voice, and text analysis. Deep learning techniques have been applied in various fields, such as image recognition, voice, data analysis, video games, and robotics. Deep Learning, as a technique in machine learning, enables nonlinear information processing through multiple layers for pattern identification, feature extraction, and classification. This method enables computational models composed of many layers of processing that learn data from different levels of abstraction. Deep Learning algorithms include Stacked Autoencoders, Restricted Boltzmann Machine (RBM), Deep Belief Networks (DBN), and Convolutional Networks. Convolutional Neural Network (CNN) is a type of neural network in deep learning that is often used in computer applications. deep learning that is often used in computer vision applications such as image classification, object detection, and human face recognition. Compared with Artificial Neural Network (ANN) and Recurrent Neural Network (RNN), CNN is particularly used for image data processing by using kernel that extracts features through convolution operations [11].

Deep Learning is a network that consists of several layers. These layers are a collection of nodes. A node is just a place where calculations occur. A node's input is combined with its weight, after which the input and weight are summed and the sum passes through a stage called the node's activation function, to determine how far the signal proceeds further through the network, affecting the final result. The difference between Deep Learning and neural networks is that there are more hidden layers in Deep Learning. If there are more than three layers (including input and output) then it qualifies as “deep” learning. So Deep Learning can be technically defined as machine learning that has more than one hidden layer [12].

D. *Convolutional Neural Network*

CNN (Convolutional Neural Network) is a type of deep learning algorithm used to process data with a certain structure such as images, videos, and other spatial data. CNN works by gradually extracting features from images, starting from simple features such as lines and angles, to more complex features such as shapes, textures, and patterns. This is done by applying a series of convolution operations to the input data, followed by pooling and activation.

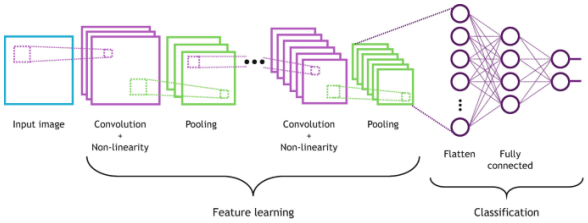


Fig. 1 CNN

CNN is simply an artificial neural network that uses a convolution multiplication matrix in its architecture. The convolution function in CNN is used for feature extraction, and this process will produce certain features that will be processed by a multilayer perceptron to produce an output from the input [13]. Convolutional Neural Networks (CNN) are deep neural network architectures that are highly effective in recognising visual patterns and are widely used in tasks such as image recognition and classification [14]. CNN has been used in various applications, such as face recognition, handwriting recognition, medical image classification, and object recognition in videos. CNNs are also used in other artificial intelligence applications, such as natural language processing and stock price prediction. CNNs became popular due to their ability to process high-dimensional and complex data, as well as the ability to automatically learn feature representations from input data. CNNs continue to be developed and used in various fields, and are one of the leading technologies in artificial intelligence.

E. Confusion Matrix

Confusion matrix testing is one method to measure the performance of a classification model. Confusion matrix is used to evaluate the extent to which the model can classify data correctly. Confusion matrix consists of four main components, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These values are used to calculate the accuracy, precision, recall, and f1-score of a sentiment analysis model.

Accuracy, which is a value that shows the level of accuracy of the system in classifying correctly [15] :

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

Fig. 2 Accuracy

Precision, which is the ratio of the amount of data that may be recognized with the amount of data that is recognized. The following is the precision calculation formula [15] :

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

Fig. 3 Precision

Recall (Sensitivity), which is the ratio of the amount of data that may be recognized with the total all recognized data. Here is the formula recall calculation [15] :

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

Fig. 4 Recall

The F1-score is a metric that assesses the overall performance of a model by computing the harmonic mean of precision and recall. It offers a balanced comparison between precision and recall [15] :

$$F1 - Score = 2 \text{ x } \frac{Precsion\ x\ Recall}{Precision+Recall}$$

Fig. 5 F1-Score

F. User Acceptence Test

User Acceptance Test (UAT) testing is carried out by asking prospective users to test the use of application programs, after conducting an assessment trial through a questionnaire that has been provided [15]. User Acceptance Test (UAT) testing aims to ensure that the application that has been designed is able to fulfill the components in business documents and is acceptable to the user. This test focuses on software functionality and other technical details that will later be used by users. This test is carried out by distributing questionnaires to MTs Darun - Najah students and also interviews with Arabic language teachers at MTs Darun - Najah, this test is carried out to find out the response of users to the system that has been built, whether the system has overcome all the problems contained in the background of this research.

III. METHOD

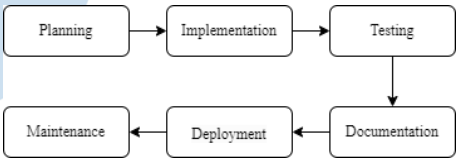


Fig. 6 Agile [16]

The research stages carried out in this study are agile methods. This agile method is designed to help developers work quickly by adjusting their needs [16]. Agile methods allow for customisation of needs and priorities throughout the development process. The team can quickly change plans and focus to adjust to the required changes [17]. Agile methods are an approach to software development that emphasises team collaboration, flexibility, continuous iteration, and rapid response to change. The approach focuses on delivering high-value products through short development cycles, often called sprints, which typically last 1-4 weeks. Key principles of Agile include interaction and collaboration between individuals, prioritisation of working software, active collaboration with customers, and the ability to respond to changing needs even in the late stages of development.

A. Planning

The first stage carried out is planning, which in this stage analyzes to find out and understand user needs for the software to be designed. Starting with conducting a

literature study to find references related to this research, namely by looking for previous journals on object recognition using the Convolutional Neural Network (CNN) method and also architectures that will later be tried. Thus, a comprehensive understanding of the framework that has been applied previously in similar studies can be obtained. Data and information collection methods in this study were also obtained by means of school surveys and interviews with Arabic language teachers at Mts Darun - Najah. This interview stage was conducted with the Arabic language teacher at Mts Darun - Najah. After conducting interviews with MTs Darun - Najah teachers, what is done at this stage is data collection. Data collection is done by taking images from various websites such as kaggle, pinterest, and google. also google.

B. Implementation

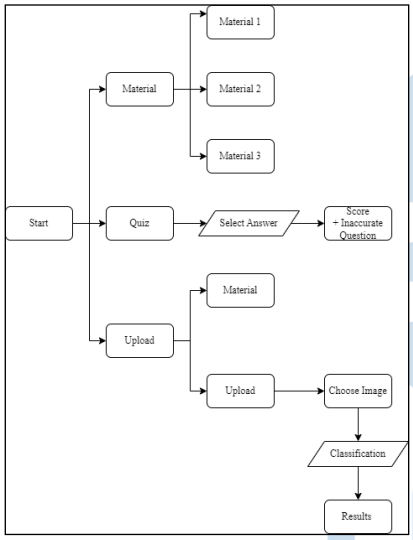


Fig. 7 Design System

First, there is the "Materi" menu which contains several pictures and their Arabic explanations. Second, there is a "Quiz" menu that allows users to work on a series of questions that have been provided. Finally, there is an "Upload" menu that allows users to enter images and find out the Arabic language of the image. The existence of the three main menus is based on the concept of this application as a learning media. It is important that there is a two-way interaction between the user and the application, so that the user can obtain information about Arabic from an object and at the same time involve themselves in the learning process through various activities that have been provided. In the implementation process, researchers also create a website mockup design that will later be applied to the website created. Here is the website development that has been made.

- Homepage

Fig. 8 Homepage

When users first access the learning website, they are greeted with a Homepage designed to provide a clear introduction to the platform.

- Menu Page

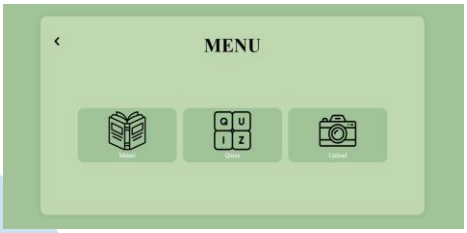


Fig. 9 Menu Page

The Menu Page displays three main options that can be accessed by users of this website.

- Select Material Page

Fig. 10 Select Material Page

The user is given the freedom to choose which chapter the user wants to explore further.

- Material Page

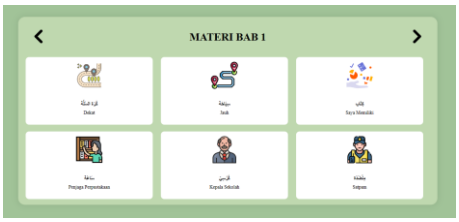


Fig. 11 Material Page

On the Material Page there is a provision of images, and bilingual text, namely Arabic and Indonesian.

• Question Page

Fig. 12 Question Page

Each question provides three answer options that can be selected by the user. The system will automatically change the question after the user selects the correct answer.

• Score Page

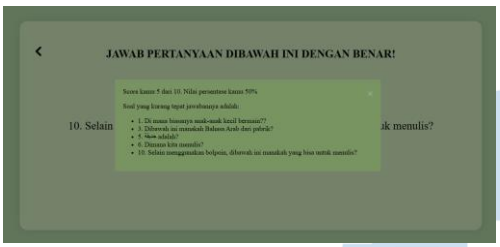


Fig. 13 Score Page

Once the user has completed the set of questions provided, the system will automatically display a Pop-Up containing the score corresponding to their work. In addition to getting a score, users can also see a list of questions that they answered incorrectly.

• Upload Menu Page

Fig. 14 Upload Menu Page

The Upload Menu view, which displays two menu options available to the user. The first is Material Upload, which serves as a repository for various materials relevant to the scanning process. The second is the Upload Menu itself, which is the main focus of using this feature.

• Upload Material

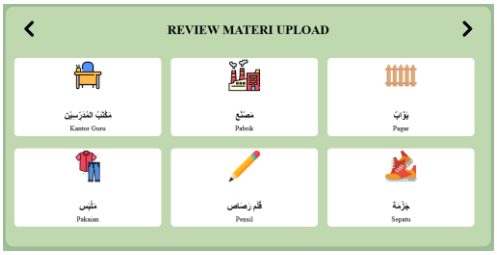


Fig. 15 Upload Material Page

In the Upload Material Menu, each object presented is accompanied by a picture describing the object, a text in Indonesian describing the object, and a translation of the text in Arabic.

• Upload Page

Fig. 16 Upload Page

Fig. 17 Upload Page

Users will upload an image to find out its Arabic equivalent. When the ‘Classification’ button is clicked, the website will display the Arabic translation. In this process, the CNN model with EfficientNet-B7 architecture will process the image uploaded by the user, using 70 epochs to produce optimal classification results.

C. *Testing*

At this stage, after the website is successfully designed, website testing will be carried out using two methods, namely Confusion Matrix for measuring the performance of a model, this test shows how well the model classifies data into the correct class and which errors occur, the confusion matrix consists of four cells namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) and User Acceptance Test (UAT) to users to find out whether the website that has been built can be accepted or not by users.

D. Documentation

In agile methods, the documentation phase focuses on providing relevant and timely information throughout the development cycle. It starts with an application vision and roadmap that identifies the main features of this application. The main feature of the application built is the "Upload" feature which will allow users to know the Arabic language of the uploaded image. Documentation continued with the development of a product backlog containing user stories with descriptions, acceptance criteria, and priorities. Technical documentation such as code and comments are updated to support team understanding, and testing is documented to ensure quality. After the sprint, results and feedback are reviewed and recorded, and release notes and user guides are updated if there are significant changes. Continuous documentation ensures all information remains relevant and supports effective software development and maintenance.

E. Deployment

The deployment stage in agile involves a series of steps that ensure software can be released quickly and reliably after each sprint. The process starts with continuous integration, where newly developed code is routinely merged into the main repository and tested automatically. After that, automated builds produce a version of the software that is ready for further testing. Automated and manual testing is performed to ensure that the software is free of bugs and fulfils the acceptance criteria. When the build passes all tests, preparations for release are made, including compiling release notes and configuring the production environment. Finally, deployment is done automatically or semi-automatically to the production environment, followed by post-release monitoring to ensure everything is on track and address any issues that may arise.

F. Maintenance

Maintenance allows developers to fix errors that occur in the system. This includes fixing errors, improving the implementation of system units, and upgrading and customizing the system as needed.

IV. Result

A. Model Development

1) Resnet – 50

In the development of Resnet - 50, researchers have tried several scenarios, by playing with batch sizes of 32 and 64 with 6 epochs and with the dataset ratio of 80:20. In this experiment using 10 data for manual testing.

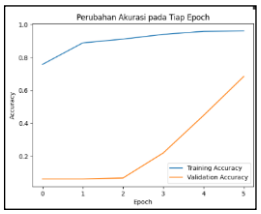


Fig. 18 Accuracy 32

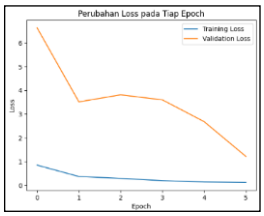


Fig. 19 Loss 32

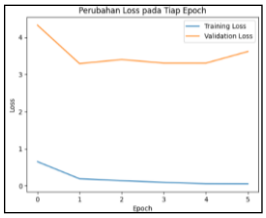


Fig. 20 Accuracy 64

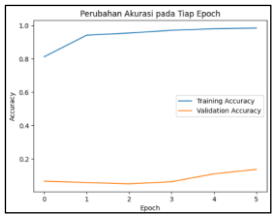


Fig. 21 Loss 64

TABLE I RESNET 50

Batch Size	Accuracy	Manual Testing
Batch Size 32	96%	9 true, 1 false
Batch Size 64	98%	7 true, 3 false

According to scikit-learn.org and medium .com that in both experiments the models experienced overfitting and underfitting, so they did not perform well.

2) Resnet – 18

In the development of Resnet - 18, researchers have tried the 10 epoch scenario with 32 batch sizes and with a dataset ratio of 80:20.

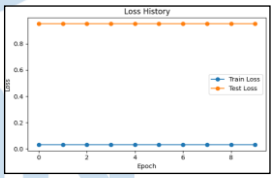


Fig. 22 Loss Resnet 18

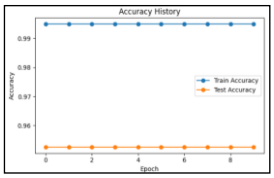


Fig. 23 Accuracy Resnet 18

At the 10th epoch using a batch size of 32, the model achieved an accuracy of 98%. Although the accuracy achieved is quite good, the visualisation results of the model development graph show an unsatisfactory graphical picture.

3) VGG – 16

In the development of VGG -16, researchers have tried several scenarios, namely by playing epochs of 15, and 20, with 32 batch sizes and with a dataset ratio of 80:20. In this experiment using 10 data for manual testing.

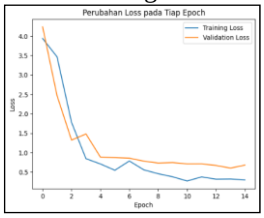


Fig.24 Loss 15

Fig. 25 Accuracy 15

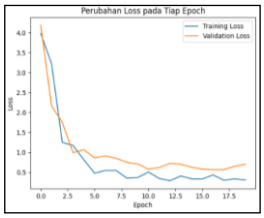


Fig. 26 Loss 20

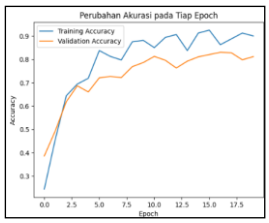


Fig. 27 Accuracy 20

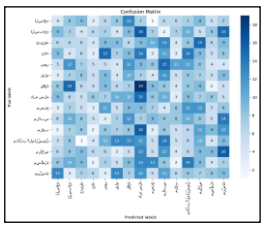


Fig. 28 Matrix 15

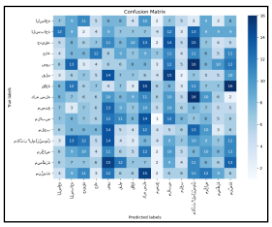


Fig. 29 Matrix 20

TABLE II VGG - 16

Epoch	Accuracy	Manual Testing
Epoch 15	89%	9 true, 1 false
Epoch 20	90%	10 true

According to the websites scikit-learn.org and medium.com both show good graphics as they display some important characteristics of an effective model training process. However, when the confusion matrix testing process is performed, both show that the model has a poor performance in classifying the data. This problem occurs because the model is not able to distinguish between different classes properly. Poor data quality, such as data that is not representative, has a lot of noise, or has wrong labels, can also lead to poor model performance.

4) EfficientNet – B7

In the development of EfficientNetB7, researchers have tried several scenarios, namely by playing epochs of 50, 60, and 70, with 32 batch sizes. In this experiment using 10 data for manual testing.

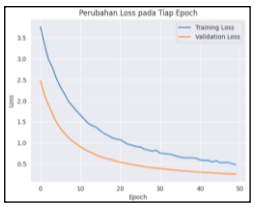


Fig. 30 Loss 50

Fig. 31 Accuracy 50

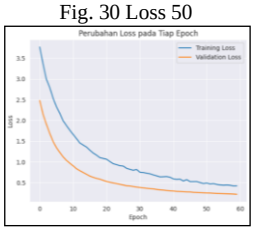


Fig. 32 Loss 60

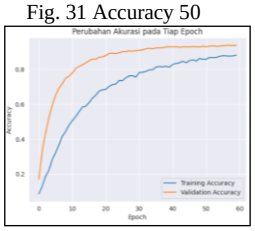


Fig. 33 Accuracy 60

Fig. 34 Accuracy 70

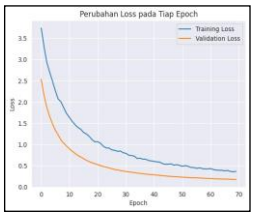


Fig. 35 Loss 70

TABLE III EFFICIENT NET - B7

Epoch	Accuracy	Manual Testing
Epoch 50	86%	10 true
Epoch 60	88%	10 true
Epoch 70	90%	10 true

In all three experiments, according to scikit-learn.org and medium.com, the graphs show a positive trend in the training process of the machine learning model, which means that they are able to provide a good enough model for this study. The consistency of improvement in both curves shows that the model does not experience overfitting, so the model has good predictive ability and is reliable for new data that has never been seen before. From the analysis that has been done, it can be concluded that EfficientNet B7 with epoch 70 and batch size 32 can be used for this study.

B. Testing

1) Confussion Matrix

After finding the optimal architecture, EfficientNet-B7, by running the training process for 70 epochs, the test results showed excellent performance through the resulting confusion matrix. The confusion matrix indicates that this model has a high level of accuracy in classifying the test data.

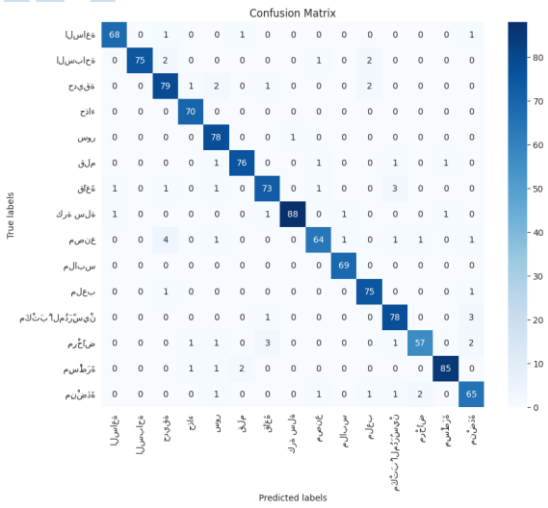


Fig. 36 Confussion Matrix

TABLE IV TP,FN,FP,TN

Class	TP	FN	FP	TN
الساعت	68	3	2	1095
اسباحت	75	5	0	1088
حديقت	79	6	9	1084
حداء	70	0	3	1093
سور	78	1	8	1085
قلم	76	4	3	1087
قعاة	73	7	6	1090
كرتسلة	88	4	1	1075
مسطرة	85	4	2	1078
مَكْتَبْ المُدَرّسِيْن	78	4	7	1085
مصنع	64	9	4	1099
ملبس	69	0	2	1094
ملعب	75	2	5	1088
مرحاض	57	8	3	1106
منضدة	65	6	8	1098

In Tables the concepts of True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) are very important. True Positive occurs when the model correctly identifies positive examples, while False Negative occurs when the model fails to classify examples that are actually positive. Conversely, False Positive occurs when the model incorrectly identifies a negative example as positive, while True Negative occurs when the model correctly classifies a negative example as negative. In this test, the model successfully collected 1163 data from each class. One example is in class 1 which has a total of 71 data that has a TP value of 68, which means there are 3 data that fall into FN, namely data that has an actual value of positive (in class 1) but is predicted as another class or negative. Next is FP which has an actual value negative (other class) but is predicted as the class, in class 1 has 2 data from other classes that are predicted as class 1. In TN data with negative actual value or other classes other than 1 and predicted correctly according to the class is as much as 1095 which is obtained from the total overall data minus the total TP data.

TABLE V PRECISION, RECALL, F1-SCORE

Avg Precision	Avg Recall	Avg F1-Score
94.6%	94.6%	94.6%

From the calculations that have been done, the model shows good performance overall. The average precision of 94.6% indicates the model's ability to accurately predict the positive class, while the average recall of 94.6% indicates the model's ability to find all instances of the positive class. The average F1-Score value of 94.6% indicates a good balance between precision and recall.

2) UAT

UAT testing that researchers do is by distributing Google Form links to collect data from user responses. The questions on the Google Form link are about the functionality of the application that has been built.

TABLE VI UAT

P	SS x 5	S x 4	N x 3	TS x 2	SST x 1	A	%
1.	20	40	0	2	0	62/15 = 4.1	82%
2.	30	32	3	0	0	65/15 = 4.3	86%
3.	25	40	0	0	0	65/15 = 4.3	86%
4.	25	40	0	0	0	65/15 = 4.3	86%
5.	75	0	0	0	0	75/15 = 4.6	92%
6.	75	0	0	0	0	75/15 = 4.6	92%
7.	75	0	0	0	0	75/15 = 4.6	92%
8.	20	40	3	0	0	63 / 15 = 4.2	84%
9.	40	20	6	0	0	66/15 = 4.4	88%
10.	25	36	3	0	0	64/15 = 4.2	84%

In Table VI is the average result and percentage of each question, and it can be concluded that the average percentage is 87.2% which is obtained from the total percentage divided by the number of questions.

V. CONCLUSIONS

The system on the website has run well as evidenced by blackbox testing, UAT testing, testing on students, and has been validated directly by the Arabic teacher MTs Darun - Najah. The system developed by implementing CNN and EfficientNetB7 for Arabic vocabulary recognition in MTs Darun - Najah students has an accuracy of 90% with an average precision of 94.6%, an average recall of 94.6% and an average f1-score of 94.6%.

ACKNOWLEDGMENT

The authors would like to thank the Jember State Polytechnic and the supervisors who have supported this research.

REFERENCES

[1] Mailani, O., Nuraeni, I., Syakila, S. A., Lazuardi, J., & Komunikasi, P. I. (n.d.). Bahasa Sebagai Alat Komunikasi Dalam Kehidupan Manusia (Vol. 1, Issue 2). Online. www.plus62.isha.or.id/index.php/kampret.

[2] Pera Aprizal, A. (n.d.). Jurnal Pendidikan Guru. In *Jurnal Pendidikan Guru* (Vol. 2, Issue 2).

[3] Susilawaty. (2021). Prosiding Seminar Nasional Lingkungan Lahan Basah (Vol. 6).

[4] Sabaniah, S., Ramdhan, D. F., & Rohmah, S. K. (2021). Peran Guru dalam Pelaksanaan Pembelajaran Jarak Jauh di Tengah Wabah Covid - 19. *Edunesia : Jurnal Ilmiah Pendidikan*, *2*(1), 43–54. https://doi.org/10.51276/edu.v2i1.77.

[5] Jurusan, S., Sosiologi, P., Sultan, U., & Tirtayasa, A. (2019). Media Pembelajaran Dalam Proses Belajar Mengajar. 2(1), 470–477.

[6] Ratih Kusuma Ningtias. (2021). Analisis Kesulitan Belajar Maharoh Kalam Pada Mata Kuliah Bahasa Arab Mahasiswa Jurusan Pendidikan Agama Islam Institut Agama Islam Tarbiyatut Tholabah Lamongan. *Darajat: Jurnal Pendidikan Agama Islam*, *4*(1), 91–100. https://doi.org/10.58518/darajat.v4i1.1751.

[7] Urgensi Bahasa Arab; Bahasa Arab Sebagai Alat Komunikasi Agama Islam Akhiril Pane. (n.d.)..

[8] Suroiyah, E. N., & Zakiyah, D. A. (2021). Perkembangan Bahasa Arab Di Indonesia. 3. https://ejournal.iaiskjmalang.ac.id/index.php/akad/article/view/207.

[9] Mufidah, N., Izha, I., Pendidikan, R., Arab, B., Malik, U. M., & Malang, I. (2020). Pengajaran Kosa Kata Untuk Mahasiswa Kelas Intensif Bahasa Arab (*Vocabulary Teaching For Arabic Intensive Class*). In *Uniqbu Journal Of Social Sciences (UJSS)* (Vol. 1, Issue 1).

[10] Aryanto, R., Alfan Rosid, M., & Busono, S. (2023). Penerapan *Deep Learning* untuk Pengenalan Tulisan Tangan Bahasa Aksara Lota Ende dengan Menggunakan Metode *Convolutional Neural Networks*. Jurnal Informasi Dan Teknologi, 5(1), 258–264. https://doi.org/10.37034/jidt.v5i1.313.

[11] Iskandar Mulyana, D., Ainur Rofik, M., & Ohan Zoharuddin Zakaria, M. (n.d.). Klasifikasi Kendaraan pada Jalan Raya menggunakan Algoritma *Convolutional Neural Network (CNN)*.

[12] Rizki, M., Basuki, S., & Azhar, Y. (2020). Implementasi *Deep Learning* Menggunakan Arsitektur *Long Short Term Memory* Untuk Prediksi Curah Hujan Kota Malang. *REPOSITOR*, *2*(3), 331–338.

[13] Prasetyo, E., Wijaya, N. A., Armanto, H., & Zaman, L. (2020). Klasifikasi Akasara Jawa Dengan Cnn. 13(2), 2620–4770.

[14] Lusiana, Kurniasari, A. A., & Kusumawati, I. F. (2024). Enhancing Accuracy in the Detection of Pneumonia in Adult Patients: An Approach by Using Convolutional Neural Networks BT - Proceedings of the 4th International Conference on Electronics, Biomedical Engineering, and Health Informatics (T. Triwiyanto, A. Rizal, & W. Caesarendra (eds.); pp. 593–609). Springer Nature Singapore.

[15] Sholawati, M., Auliasari, K., & Ariwibisono, F. X. (2022). Pengembangan Aplikasi Pengenalan Bahasa Isyarat Abjad Sibi Menggunakan Metode *Convolutional Neural Network* (CNN). In Jurnal Mahasiswa Teknik Informatika) (Vol. 6, Issue 1).

[16] Pertiwi, T. A., Luchia, N. T., Sinta, P., Aprinastya, R., Dahlia, A., Fachrezi, I. R., & Hamzah, M. L. (2023). Perancangan Dan Implementasi Sistem Informasi Absensi Berbasis Web Menggunakan Metode Agile Software Development. *Jurnal Testing Dan Implementasi Sistem Informasi*, *1*(1), 53–66.

[17] Kurniasari, A. A., Olivia, Z., Lironika Suryana, A., Widiyawati, A., & Maria Rosiana, N. (2023). NutriTalk: Nutrition Intervention by Experts to Reduce the Impact of Stunting Through Mobile-Based Applications Using Agile Method. *Jurnal Teknokes*, *16*(3), 170–177. https://doi.org/10.35882/teknokes.v16i3.632

ISSN 2085-4552

Sentiment Analysis in E-Commerce: Beauty Product Reviews

Gavrila Louise Tumanggor[1], Feliks Victor Parningotan Samosir[2]
[1, 2] Informatics Study Program, Faculty of Computer Science and Information Technology, Indonesia
[1]01082210011@student.uph.edu [2]feliks.parningotan@uph.edu

Accepted 19 August 2024
Approved 08 January 2025

***Abstract*— The increasing popularity of online shopping platforms is fueling the need for automated sentiment analysis for product reviews. This research aims to build an automatic sentiment analysis model in Indonesian for e-commerce product reviews. This model is expected to help consumers make purchasing decisions more quickly. We utilize the IndoBERT model, which has shown to be quite effective for general sentiment analysis, achieving an evaluation accuracy of 66.2% despite a high evaluation loss of 0.8006. The approach used combines Natural Language Processing (NLP) and Machine Learning (ML) techniques. It is hoped that this research will be useful for consumers, shop owners, and researchers in efficiently understanding the sentiment of e-commerce product reviews.**

***Index Terms*— *E-Commerce, Sentiment Analysis, Product Reviews, IndoBERT*.**

I. INTRODUCTION

In this digital era, e-commerce has become an increasingly popular platform for online shopping, with marketplaces like Shopee in Indonesia offering users the convenience of conducting transactions online. Product reviews on e-commerce platforms such as Shopee serve as crucial sources of information for consumers to assist them in making purchasing decisions[1]. These reviews encompass a range of opinions, from positive to negative, and neutral sentiments, providing insights into the quality of a product and significantly impacting both consumers and producers. In today's business environment, customer satisfaction stands as a primary goal for companies, and customer reviews posted on social media can greatly influence the perceptions of potential customers. The increasing prevalence of sentiments and opinions on product review sites underscores the importance of having review data available for analyzing customer opinions [2].

Natural Language Processing (NLP) is a computer science field that focuses on the interaction between computers and human language. NLP techniques can be used to process and analyze text, such as tokenization, stemming, lemmatization, and text classification. Sentiment analysis is a technique used to identify and understand opinions or feelings contained within text [2]. In the context of e- commerce, sentiment analysis of product reviews can help various parties, such as helping consumers choose the right product by understanding other users' opinions and experiences with the product [3], assist store owners in understanding how their products are received by consumers, identify product deficiencies, and improve the quality of their products and services [4], also helps researchers in understanding trends and patterns in consumer opinions towards certain products [5].

Manual sentiment analysis of e-commerce product reviews takes a lot of time and effort, especially if the number of reviews is very large[6]. This is a challenge, especially for consumers who want to quickly find out the general sentiment towards a product. Therefore, an automatic sentiment analysis technique is needed that can process and analyze product reviews quickly and accurately.

Machine learning (ML) is a sector of artificial intelligence dedicated to employing algorithms systematically to uncover the inherent connections within data and information [7]. For instance, ML systems can undergo training with automatic speech recognition systems (like iPhone's Siri) to transform acoustic data into a sequence of speech into a semantic framework represented by a string of words ML algorithms can be used to build sentiment analysis models that can predict the sentiment of a text.

This research aims to develop an automatic sentiment analysis model for e-commerce product reviews in Indonesian. This model is expected to help consumers make purchasing decisions more quickly and easily.

Product reviews play an important role for shop owners and buyers. For store owners, reviews provide direct feedback on the quality and performance of their products. By understanding what customers like or dislike, they can improve their products to meet market needs. Positive reviews also build customer trust and a store's reputation, while negative reviews provide an opportunity to respond quickly to any issues that may arise. For buyers, reviews provide valuable

guidance for making better purchasing decisions. They can learn about other people's experiences with the product before they decide to buy it. Reviews also provide validation of product claims and help buyers avoid waste by uncovering any problems or shortcomings they may encounter[8].

II. THEORY

There are several studies that have carried out e-commerce sentiment analysis. First, in the paper entitled "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT" by Chandradev, et al [9]. The research in this paper aims to analyze the sentiment of hotel reviews using the BERT deep learning method. The model used in this research is SmallBERT, which is a variant of BERT with a smaller size but still maintaining language processing capabilities. The use of the SmallBERT model in this research involved a fine-tuning process on a dataset of 515k hotel reviews. The sentiment analysis process begins by loading the model resulting from the fine-tuning that has been carried out previously. Next, the tokenization stage is carried out using the BERT Tokenizer to convert the review text into a number representation that matches the BERT vocabulary. After that, input packing is carried out to change the data structure into input that is in accordance with the BERT model. The model then performs sentiment classification with binary classification into negative and positive labels. The model achieves low loss values and high accuracy on training and validation data. The model does not suffer from overfitting or underfitting, indicating good ability in analyzing hotel review sentiment. The advantage of this research is the efficient use of the SmallBERT model in analyzing hotel review sentiment with high accuracy. In addition, the BERT deep learning approach makes it possible to capture relationships between words contextually, improving the model's performance on complex sequential tasks in natural language processing A shortcoming of this study may lie in the scale of the dataset used. The results of this research include achieving loss values and accuracy of the SmallBERT model on training and validation data, as well as visualization of sentiment comparison data using bar charts and pie charts.

Further research is in the paper entitled "Implementasi Fine-Tuning BERT untuk Analisis Sentimen terhadap Review Aplikasi PUBG Mobile di Google Play Store" by Alex Sander Prasetya Braja, and Achmad Kodar[10]. In the research discussed in this paper, sentiment analysis was carried out on user reviews of the PUBG Mobile application on the Google Play Store. This research uses the BERT BASE Multilingual and IndoBERT BASE models. Both models are fine-tuned with predetermined hyperparameters. These models were trained using PyTorch and the transformers package from hugging face. The research results show that IndoBERT BASE obtained the best testing accuracy on batch size 32 of 0.93519, while BERT BASE Multilingual obtained the best testing accuracy at 0.71149. Using a fine-tuning approach allows the model to learn from larger data in Indonesian, including formal language and slang. However, the drawback may lie in the unbalanced distribution of sentiment classes in TextBlob-based labeling data.

The next research is regarding "Implementasi-Bert-Untuk-Analisis-Sentimen-Terhadap-Ulasan-Aplikasi-Flip-Berbahasa-Indonesia" [11]. This research aims to apply the Language-Based Transformer (BERT) model in analyzing the sentiment of user reviews of the Flip application in Indonesia. Through natural language processing (NLP) and tokenization techniques, user review data is analyzed using the BERT model to determine positive, negative or neutral sentiment. The research results show that the BERT model can be effective in analyzing sentiment, with the majority of reviews being positive, temporary Negative reviews provide input for application improvement. In this research, it can be seen that BERT implementation has the potential to increase understanding of user sentiment towards the Flip application and provide valuable insights for developers in improving service quality.

The next paper entitled "Analisis Sentimen Pada Ulasan Aplikasi Tokopedia Menggunakan Klasifikasi Naïve Bayes" by Aurelia, et al[12]. The research discussed in this paper conducts sentiment analysis on user reviews of the Tokopedia application using the Naïve Bayes classification method. Data were collected from the Google Play Store and labeled as positive or negative based on user-provided scores. Subsequent processes involved data preprocessing, including the creation of a corpus from reviews and the formation of a sparse document-term matrix. The data were then split into training and testing sets for the classification process. The Naïve Bayes method was chosen as the classification model for its effectiveness in sentiment analysis, particularly in the context of text mining. This model can classify reviews as positive or negative based on the words they contain. The research achieved an accuracy rate of 82.97%, demonstrating success in classifying reviews. The advantage of Naïve Bayes lies in its ability to achieve high accuracy with a relatively small amount of training data. However, its drawback may lie in the naive assumption of feature independence, which may not always hold true in real-world data. Out of 1819 analyzed reviews, 338 were classified as positive and 1481 as negative. For future research, it is suggested to conduct a deeper analysis of neutral reviews (with a score of 3) that were excluded in this study. Additionally, other classification methods could be considered to compare their effectiveness with Naïve Bayes. Thus, this research successfully implements the Naïve Bayes model for sentiment analysis of Tokopedia application reviews with a satisfactory level of accuracy.

The fifth paper is entitled “Analisis Sentimen AicoGPT (Generative Pre-trained Transformer) Menggunakan TF-IDF” by Rahayu, et al [13]. The aim of the study presented in this paper is to conduct sentiment analysis on the reviews of the AicoGPT application from the Google Play Store. The research method involves data collection of AicoGPT comments, data preprocessing, modeling using classification with Logistic Regression (LR) and Support Vector Machine (SVM) utilizing the TF-IDF technique, and evaluation of results using confusion matrix and ROC Curve. The models utilized in this research are Logistic Regression (LR) and Support Vector Machine (SVM) employing the TF-IDF technique. LR is utilized for data classification, while SVM is used to compare model performance. Evaluation results using ROC Curve indicate that the SVM model performs the best with an f1-score of 73.34%. The strength of this study lies in the utilization of advanced methods such as TF-IDF and classification models which can provide valuable insights for AicoGPT application developers. However, its weakness may lie in the limitation of the dataset used and the classification model that does not encompass various possible sentiments that may arise in reviews. The findings of this research indicate that the SVM model outperforms LR in conducting sentiment analysis on AicoGPT application reviews. Thus, this study successfully achieves its goal of providing input for application developers and users. For future research, a more in-depth analysis of specific sentiment types, the utilization of more complex classification models, and the use of broader and more representative datasets to enhance the accuracy and generalization of sentiment analysis results could be considered.

The last paper entitled “Analisis Sentimen pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes” by Abdul, et al[14]. In this study, sentiment analysis was conducted on product reviews from Kinophonecell store on the Shopee e-commerce platform using the Naïve Bayes classification algorithm. The Naïve Bayes model was employed to predict the sentiment category of customer reviews towards the products, classifying the reviews as either positive or negative based on the available dataset. The results of the study indicate that the Naïve Bayes method yielded effective outcomes, with an accuracy of 99.5%, precision of 99.49%, and recall of 100%. Although the accuracy did not reach 100%, these results are considered satisfactory for predicting product review sentiments. One advantage of this research is the utilization of the Naïve Bayes method, which is relatively easy to implement and provides good results in predicting product review sentiments. However, a limitation is the accuracy not reaching 100%, suggesting the need for improving the quality of the training data to enhance prediction accuracy. The study also revealed that positive sentiments outweighed negative sentiments, with a significantly larger number of positive sentiments compared to negative ones. The quantity of training data significantly influences the system's predictions, underscoring the importance of high-quality training data. For future research, it is recommended to enhance the quality of training data by increasing the dataset size and refining ambiguous sentiment labels. Additionally, employing alternative classification methods for comparison with Naïve Bayes results could be beneficial. Conducting deeper analyses on factors influencing product review sentiments and integrating sentiment analysis with other factors such as pricing, promotions, or service quality would provide a more comprehensive understanding.

Based on the literature review, it was decided to use BERT due to its advantages in understanding the context and relationships between words in text, which is crucial for sentiment analysis. Previous studies, such as the research by Chandradev et al. that used SmallBERT for sentiment analysis of hotel reviews and demonstrated high accuracy without overfitting or underfitting, have proven the effectiveness of this model. Additionally, the research by Alex Sander Prasetya Braja and Achmad Kodar, which employed IndoBERT BASE for sentiment analysis of PUBG Mobile application reviews, also showed impressive results with a testing accuracy of 0.93519. Specifically, we will use IndoBERT because our dataset is in Indonesian, and this model has been optimized to understand the nuances and variations of the Indonesian language, including the use of formal and slang expressions. The use of IndoBERT is expected to improve the accuracy of sentiment predictions and provide deeper insights into e- commerce product reviews in the context of the Indonesian language.

III. METHOD

A. Data Preparation

First, data was collected from Hugging Face Hub, a platform that provides open access to a variety of datasets and NLP models that can be used for a variety of tasks, including sentiment analysis. In this research, the review dataset used is from Sekar Mulyani [15] namely beauty products from Shopee taken from the Hugging Face Hub using the API or method provided by Hugging Face.

	Review	Bintang 1	Bintang 2	Bintang 3	Bintang 4	Bintang 5
0	terima kasih shoppe paket ny udah datang denga...	True	False	False	False	False
1	kecewa sekali box nya kurang rapi dan produkny...	False	True	False	False	False
2	it's been days and the itch still haven't gone...	False	True	False	False	False
3	bukanny makin mulus malah tumbuh jerawat gede ...	False	True	False	False	False
4	pengalaman penggunaan:gampang patah	False	True	False	False	False
...	...	...	...	...	...	...
57187	free beauty blender tidak ada. tadi nya saya p...	False	False	False	True	False
57188	kalian wajib liat vid ini.	False	False	True	False	False
57189	pesanan tidak sesuai, tidak tepat waktu, tidak...	False	True	False	False	False
57190	terimakasih, barang sdh di terima dengan baik...	False	False	False	True	False
57191	isi tidak sesuai	True	False	False	False	False

57192 rows × 6 columns

Fig. 1. Initial Dataframe

Based on figure 1, the data used has 57.192 rows and 6 columns, namely review column, 1 star column, 2star column and so on up to 5star column. In this column, if an item gets 1 star, then the 1star column will be true while the other stars will be false.

	Review	Bintang 1	Bintang 2	Bintang 3	Bintang 4	Bintang 5
0	semua bagus, nyampe langsung coba tapi koq bau...	False	False	True	False	False
1	gampang dipake buat ngebentuk mata bahkan bagi...	False	False	False	False	True
2	baru beli udah kering, gada video unboxing di ...	True	False	False	False	False
3	efektivitas:oke tekstur:oke kandungan:oke pesa...	True	False	False	False	False
4	pengiriman ok, hny saja barang tidak di tambah...	False	False	False	True	False
...	...	...	...	...	...	...
2495	parfum nya nyengat dan di kulitku kering bgt, ...	False	True	False	False	False
2496	produk okeh tapi pengirimannya lama banget, ka...	False	True	False	False	False
2497	saya pesen paket yang awet muda, harusnya dape...	True	False	False	False	False
2498	sdh di terima tp belum sempat di coba	False	False	False	True	False
2499	gak bisa pakainya	False	False	False	True	False

2500 rows × 6 columns

Fig. 2. Dataframe used

Figure 2 shows 2500 data points were randomly selected from the original dataset due to limitations in computational resources, such as processing power, memory, and time constraints. This random selection ensures that the subset remains unbiased and representative of the overall data distribution, capturing the essential characteristics and diversity of the original dataset. By working with a smaller, randomly chosen subset, the model can be trained more efficiently, allowing for faster iterations and testing of different parameters, while still maintaining the integrity of the data within the given resource limitations.

Then identify and handle missing values. This activity is carried out to maintain data quality. If this is not done, missing values can result in deviations or distortions that may occur in the analysis results due to certain characteristics of the data used and errors in the analysis, as well as reducing the accuracy of the results. By identifying and handling missing values, the data becomes more complete and can provide more reliable analysis results. The missing values contained in the data used are all 0 so there is no need to clean them.

Sentiment	count
Negative	1066
Positif	886
Neutral	548

dtype: int64

Fig. 3 Data Labeling

Entering data labeling as you can see at Figure 3, sentiment or ratings from product reviews are labeled based on a star scale given by users (for example: 1 star for negative sentiment, 3 stars for neutral sentiment, and 5 stars for positive sentiment). These labels are then used as a basis for training the sentiment analysis model.

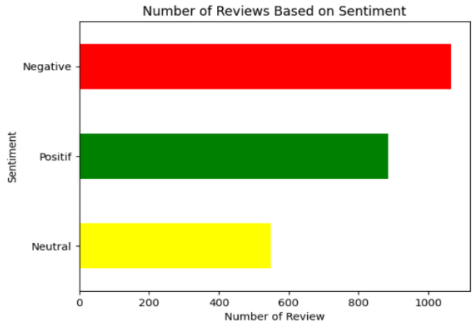


Fig. 3 Results reviews from the new sentiment

has a total of 1066 reviews, positive (4 and 5 stars) which has 886, and neutral which has 548 reviews. Neutral sentiment has the least number of reviews because only 3 stars. We can see the graphics of resuls reviews from the new sentiment in figure 4.

Each product review is divided into tokens or small units (usually words or sub-words) using a tokenizer, in this case, the tokenizer used is nltk Tokenizer. The tokenization process allows the text representation to be input that can be understood by the NLP model.

Entering the data cleaning and data preparation stage, at the data cleaning and preparation stage, text cleaning is carried out which includes removing stop words and normalizing the text to improve the quality of the analysis. The stopwords used are a collection of common words in Indonesian that tend not to have important meaning in the context of text analysis, such as "and", "or", and the like. After removing stopwords, empty tokens are cleaned to ensure data cleanliness. However, there is a problem when some words that are actually irrelevant are still read. This is caused by the existence of abbreviations that are not affected by the stopword remover.

```
30 Kata dengan Frekuensi Tertinggi setelah Pembersihan Kedua:
[('cocok', 1653), ('bagus', 1317), ('barang', 1263), ('beli', 1145), ('tekstur',
1054), ('produk', 1049), ('pengiriman', 1000), ('warna', 749), ('pas', 700),
('kecewa', 698), ('aman', 653), ('dikirim', 647), ('performa', 627), ('kulit',
622), ('kemasan', 598), ('pake', 566), ('cepat', 563), ('coba', 515),
('packing', 509), ('warnanya', 443), ('loudlycryingface', 415), ('pakai', 404),
('harga', 390), ('pesanan', 383), ('ok', 381), ('udh', 380), ('bibir', 378),
('kasih', 352), ('sampe', 351), ('kering', 350)]
```

Fig. 4 Second Cleaning Stage

Figure 5 shows the second cleaning stage. Therefore, the second cleaning stage is carried out by excluding certain predetermined words, such as 'nya', 'yang', 'ga', and others contained in the `exclude_words` list. This step aims to increase the accuracy of the analysis by eliminating words that are irrelevant and do not have significant meaning in the context of text analysis.

After data preparation, visualization is used to understand the distribution of sentiment labels in the dataset. Graphs or plots can be used to display the number of reviews for each sentiment category (negative, neutral, positive), helping in visually understanding the distribution of the data. This data

preparation aims to prepare a product review dataset for sentiment analysis, including steps such as collection, cleaning, normalization, and proper data representation for effective sentiment analysis model training.

B. Model Design

Once the data is ready, the IndoBERT model is configured to be used for training. This model is supplied with a tokenizer that has been pre-trained to optimally process text in Indonesian. The detailed model used in the code above is IndoBERT, which is a variant of the BERT model that has been pretrained for the Indonesian language. IndoBERT uses a Transformer architecture like BERT in general, but has been trained specifically for the Indonesian language. This model has parameters that have been pretrained on common natural language understanding (NLP) tasks such as mapping words into vector representations, so it can be used for various NLP tasks such as text classification. In the code above, the IndoBERT model is used to perform sentiment classification on the beauty product review dataset.

First, the text data consisting of reviews along with corresponding sentiment labels is split into training and validation parts using the `train_test_split` function from the scikit-learn library. This function randomly partitions the dataset into training and validation sets based on a specified ratio, ensuring that both sets represent the original data distribution. In this process, the data is typically split with an 80-20 ratio, where 80% of the data is used for training and 20% is used for validation. This stratified split helps in maintaining the balance of sentiment labels across both subsets, which is crucial for training a model that generalizes well.

Sentiment labels that are initially in text form ("Negative", "Neutral", "Positive") are converted into a numeric format for model training purposes. This conversion is necessary because machine learning models typically require numerical input. For instance, labels might be encoded as 0 for "Negative", 1 for "Neutral", and 2 for "Positive".

Next, the dataset is prepared by implementing the `ReviewDataset` class, which is a derivative of the `Dataset` class from PyTorch. This class is designed to load text and labels from the dataset and prepare them in a format that can be processed by the model. It includes tokenization using the BERT tokenizer, which converts text into a format that BERT can understand.

For training techniques, this model is trained using a fine-tuning approach, where a model that has been pretrained on a general task is re-tuned on a specific task, namely sentiment classification. This training process was carried out using a GPU, with a batch size of 16 and carried out for 3 epochs. The batch size refers to the number of samples processed by the model in each training iteration, while an epoch represents one complete pass through the entire training dataset. A smaller batch size of 16 allows for more frequent updates to the model parameters. The number of epochs determines how many times the model will see the entire dataset during training, with 3 epochs being a common choice for fine-tuning.

In addition, this training technique uses the AdamW optimization algorithm to optimize model parameters. AdamW is a variant of the Adam optimizer that includes weight decay, which helps in preventing overfitting. After the training process is complete, the model is evaluated using a validation dataset to measure the model's performance in performing sentiment classification. This evaluation provides insights into how well the model generalizes to unseen data and helps in tuning hyperparameters if necessary.The training process starts by calling the `trainer.train()` function, which will train the model based on the provided data. During training, the model parameters are adjusted iteratively using the AdamW optimization method, which has been implemented in the Transformers library. GPUs are used to speed up the training process, enabling parallel computing to increase efficiency.

The model training process in this context involves several key concepts: batch size, epoch, and iteration

- Batch size: In deep learning, batch size refers to the number of data samples processed simultaneously in one iteration during model training. Determining the batch size influences how much data is fed into the model at each step of the iteration. Different batch sizes can affect the model's convergence speed, memory usage, and the stability of the training process. Selecting the appropriate batch size can significantly impact the performance and final outcomes of the developed deep learning model [16]. In this context, a batch size of 16 means that 16 beauty product reviews are fed into the model simultaneously at each training step.. In this context, a batch size of 16 means that 16 beauty product reviews are fed into the model simultaneously at each training step

- Epoch: An epoch is a cycle of a machine learning algorithm during which the model learns from the entire training dataset, processed in mini-batches. In the context of using Convolutional Neural Networks (CNN) for classifying fashion and furniture, an epoch represents one iteration of learning for the machine. The more epochs used, the better the accuracy that can be achieved. However, there is a maximum accuracy limit that can be reached within the dataset. Using the appropriate number of epochs is crucial for achieving maximum accuracy in machine

learning [17]. In this case, the model views and learns all the beauty product reviews in the dataset 3 times (a predetermined number of epochs).

- Iteration: Iteration is a process or method used repeatedly to solve a mathematical problem. In the context of numerical methods, iteration is employed to find the numerical solution of an equation or function by repeatedly approaching from initial points until the desired solution is approximated. In programming, iteration also refers to a specific characteristic of an algorithm or computer program where a series of algorithmic steps are performed repeatedly within a program loop to achieve a particular goal[18]

Thus, the training process is carried out iteratively, where the model updates its knowledge about the data at each epoch. By the end of training, the model has better knowledge of how to classify beauty product review sentiment based on what it has learned from the training dataset.

C. Model Traning

TABLE I. TRAINING PROCESS

Running Time	
Num examples	2000
Num Epochs	3
Instantaneous batch size per device	16
Total train batch size (w. Parallel, distributed & accumulation)	16
Gradient Accumulation Steps	1
Total Optimization Steps	375

Table 1 showed the training process of this model. In this context, there are 2000 review examples used to train the model, with a total of 3 epochs. Each epoch includes iterations through the entire training dataset. The instant batch size setting per device is 16, which indicates the number of data instances processed simultaneously at each iteration by the hardware used, such as the GPU. The total training batch size, including parallel processing, distribution, and accumulation, is also 16. The training process includes 375 optimization steps, where the model is gradually fine-tuned to the training data to improve its ability to understand and predict sentiment from reviews. The results of this training are recorded in the training metrics, where the total training time is approximately 531 seconds. The training sample rate is about 11,298 examples per second, while the training steps per second is about 0.706

IV. RESULTS AND ANALYSIS

TABLE II. TRAINING PROCESS

METRIC	VALUE
Num examples	500
Batch size	16
Eval loss	0.8005783557891846
Eval accuracy	0.662
Eval runtime	15.9372 seconds
Samples/s	31.373
Steps/s	2.008
Epoch	3.0

Table 2 shows the evaluation, model evaluation was carried out using 500 test data samples with a batch size of 16. The evaluation results showed that the loss evaluation value (eval_loss) was 0.8005783557891846, while the evaluation accuracy (eval_accuracy) was 0.662. This means the model has an accuracy rate of 66.2% in predicting sentiment from test data. However, a high loss evaluation value indicates that the model still has difficulty in fitting the test data well, which leads to performance degradation.

There are several factors that can cause high evaluation loss and low accuracy. One is the lack of data representation that covers sufficient variation of the target classes in the dataset, especially if there is an imbalance between the number of samples for each class. Additionally, a complex model may suffer from overfitting to the training data, meaning it may learn patterns that are too specific to the training data to generalize well to the test data.

To improve model performance, several steps can be taken. First, expand the dataset by increasing less representative examples or by using data augmentation techniques to create additional variation in the dataset. Second, re-tuning the model architecture or re-adjusting model parameters to reduce overfitting. Additionally, techniques such as regularization or dropout can also be used to prevent overfitting. Furthermore, the evaluation of model performance can also be improved by conducting further research related to the selection of more informative features or the use of more sophisticated natural language processing techniques. By combining these various strategies, it is hoped that the model's performance can be improved in predicting sentiment more accurately on a wider test dataset.

Next, the model will be tested using a testing dataset. The model's performance on the test data will be evaluated through a confusion matrix to assess the

extent of the model's ability to accurately predict all sentiments on the test data. After obtaining the values from the confusion matrix, we can calculate accuracy, precision, recall, and F1-score. The calculation formula can be found in the following equation.

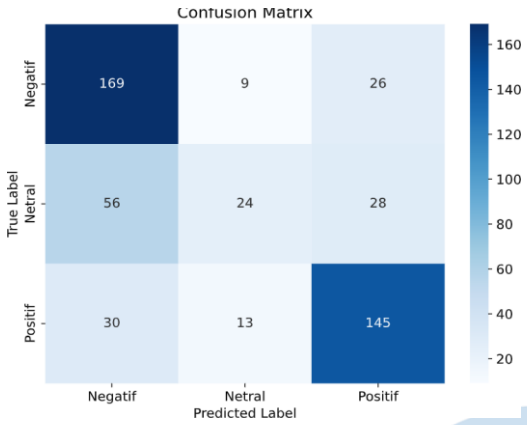


Fig. 5 Confusion Matrix

Figure 6 shows the confusion matrix, it can be seen that the model faces challenges in predicting data with neutral sentiment, but its performance is quite good for positive and negative sentiment.

Based on the results of the confusion matrix presented, it can be seen that the model has shown a relatively good level of effectiveness in conducting general sentiment analysis. From a total of 200 data samples evaluated, the model was able to accurately classify 169 cases of negative sentiment, 26 cases of neutral sentiment, and 145 cases of positive sentiment. Although there are several cases where the model misclassifies sentiment, especially in the classification between neutral and positive sentiment, the proportion of these errors is relatively small compared to the total sample evaluated.

	precision	recall	f1-score	support
Negatif	0.73	0.76	0.75	204
Netral	0.41	0.38	0.39	108
Positif	0.72	0.71	0.72	188
accuracy			0.66	500
macro avg	0.62	0.62	0.62	500
weighted avg	0.66	0.66	0.66	500

Fig. 6 Test data model evaluation

The results of model evaluation on test data as you can see on figure 7 are shown in the table above. Precision measures how accurately the model is in identifying instances that actually fall within a particular class. In this case, for negative sentiment, a precision of 0.73 indicates that of all the instances predicted as negative by the model, 73% of them are actually negative sentiment. Recall measures how well the model captures all instances that actually belong to a particular class. With a recall of 0.76 for negative sentiment, this means the model managed to find 76% of all instances that were actually negative. F1-score is the harmonic average of precision and recall, providing an overall picture of the balance between precision and recall. For negative sentiment, an F1 - score of 0.75 indicates a good balance between precision and recall. Next, support is the number of instances included in each class. Overall, the model achieved an accuracy of 0.66, which is the percentage of instances correctly predicted from the entire test data. In terms of overall evaluation, the macro avg and weighted avg values are the average of the evaluation metrics for each class, by giving the same weight or based on the amount of support for each class.

This evaluation provides an overview of the model's performance in predicting sentiment from test data, highlighting the model's strengths and weaknesses in classifying these sentiments. However, there are several notes that need to be noted. The significant difference in the number of cases misclassified between negative and positive sentiment suggests potential improvements in understanding certain linguistic nuances, such as slang or dialect, that may influence sentiment interpretation. The model may not be fully able to capture certain language variations used in the text, which can lead to errors in classification.

When evaluating a model's accuracy in classifying positive, negative, and neutral sentiment, it is necessary to consider how the model responds to text variations and complex linguistic contexts. The trained model may be able to identify common patterns in sentences that indicate certain sentiments, such as emotional words or signs of satisfaction or dissatisfaction. For example, in a restaurant review, the model may be able to recognize words such as “good” or “bad” that strongly express positive or negative sentiment. However, the model may have difficulty understanding sentences that use irony or innuendo to convey a sentiment that actually contradicts the words used.

For example, if a user writes, "Wow, the service at this restaurant was truly amazing," the model might directly interpret the word "amazing" as an indication of positive sentiment. However, in this context, the author actually uses the word ironically, because the service provided is actually bad. The model is unable to capture nuances of irony or sarcasm in text, and therefore, may produce inaccurate sentiment classifications.

It is important to remember that a model's accuracy in classifying sentiment can be affected by various factors, including language and cultural variations in the text. Languages have many nuances, idioms, and

unique ways of expressing sentiment, which may not always be easy for the model to interpret. Additionally, a model trained on a particular dataset may not be able to handle certain language or context variations well, thereby affecting its accuracy in classifying sentiment. Although the model can provide satisfactory results in sentiment classification tasks, there are limitations in its ability to understand and interpret certain linguistic nuances, such as irony or sarcasm. As a result, careful evaluation of context and language variations is important to understand the performance and limitations of models in identifying sentiment from text.

Thus, although the model has demonstrated a satisfactory level of effectiveness in conducting sentiment analysis in general, it is recommended to make adjustments or improvements to the understanding of certain slang or dialect texts to increase its accuracy and reliability in analyzing sentiment in broader and diverse contexts. Careful evaluation of a confusion matrix like this provides valuable insight into model performance and helps in identifying areas that need improvement to improve the overall quality of sentiment analysis

V. CONCLUSION

The model evaluation results show that even though the model has undergone several epochs (3.0), there are still several problems that need attention. Evaluation shows that the evaluation loss value (eval_loss) is quite high, reaching 0.8006. This shows that the model still has a significant error rate in predicting sentiment on test data. However, the evaluation accuracy (eval_accuracy) of 0.662 shows that the model succeeded in correctly predicting around 66.2% of the total examples in the test data.

The limitations seen in this evaluation may be due to several factors. One of them is the complexity of the data used. Sentiment in texts is often difficult to predict due to variations in language, context, and writing style. In addition, the presence of certain slang or dialect words in the data can also complicate the sentiment classification process. To improve model performance, future research can consider expanding the training dataset by adding more diverse and representative data. More sophisticated text processing can also be applied to address variations in language and writing style.

While there is still room for improvement, the model is still quite effective for general sentiment analysis. However, recommendations for adjustments to slang or dialect texts need to be considered so that the model can be more sensitive to variations in language. Thus, future research can lead to the development of more sophisticated models and be able to better overcome challenges in text sentiment analysis

REFERENCES

[1] Anggraini Nita and Ubaidillah Hasan, "Impact of Online Reviews, Pricing, and Viral Marketing on Shopee Purchases.," *Indonesian Journal of Law and Economics Review*, vol. 19, no. 2598–9928, pp. 10–21070, May 2024, Accessed: Sep. 04, 2024. [Online]. Available: https://ijler.umsida.ac.id/index.php/ijler/article/view/1088

[2] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, pp. 1–135, Apr. 2008, doi: 10.1561/1500000011.

[3] Md. J. Hossain, D. Das Joy, S. Das, and R. Mustafa, "Sentiment Analysis on Reviews of E-commerce Sites Using Machine Learning Algorithms," in *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, 2022, pp. 522–527. doi: 10.1109/ICISET54810.2022.9775846.

[4] B. M. D. Abighail, Fachrifansyah, M. R. Firmanda, M. S. Anggreainy, Harvianto, and Gintoro, "Sentiment Analysis E-commerce Review," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 1039–1045. doi: 10.1016/j.procs.2023.10.613.

[5] S. H. Muhammad *et al.*, "AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages," Feb. 2023, [Online]. Available: http://arxiv.org/abs/2302.08956

[6] B. Liu, "Sentiment Analysis and Opinion Mining," in *Synthesis Lectures on Human Language Technologies*, Apr. 2012. doi: 10.2200/S00416ED1V01Y201204HLT016.

[7] R. Awad Mariette and Khanna, "Machine Learning," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Berkeley, CA: Apress, 2015, pp. 1–18. doi: 10.1007/978-1-4302-5990-9_1.

[8] D. Patil and N. Rane, "Customer experience and satisfaction: Importance of customer reviews and customer value on buying preference," vol. 5, pp. 3437–3447, Apr. 2023, doi: 10.56726/IRJMETS36460.

[9] V. Chandradev, I. Made, A. Dwi Suarjaya, I. Putu, and A. Bayupati, "Chandradev, Analisis Sentimen Review Hotel menggunakan Metode Deep Learning BERT 107 Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *Jurnal Buana Informatika*, vol. 14, no. 2, pp. 107–116, 2023.

[10] A. S. P. Braja and A. Kodar, "Implementasi Fine-Tuning BERT untuk Analisis Sentimen terhadap Review Aplikasi PUBG Mobile di Google Play Store," *J I M P - Jurnal Informatika Merdeka*

Pasuruan, vol. 7, no. 3, p. 120, Sep. 2023, doi: 10.51213/jimp.v7i3.779.

[11] Eza Ananda Putra, "Sentiment Analysis Using Bidirectional Encoder Representations from Transformers (BERT)," Medium. Accessed: May 03, 2024. [Online]. Available: https://medium.com/@eza.a.putra/implementasi-bert-untuk-analisis-sentimen-terhadap-ulasan-aplikasi-flip-berbahasa-indonesia-557d691e0440

[12] N. Aurelia Salsabila, U. Sa, and F. Fauzi, "Analisis Sentimen Pada Ulasan Aplikasi Tokopedia Menggunakan Klasifikasi Naïve Bayes," *PRISMA 2024*, vol. 7, pp. 44–51, 2024, [Online]. Available: https://play.google.com/store/apps/details?id=com.tokopedia.tkpd&hl=en

[13] S. Rahayu, J. Jaya Purnama, A. Hamid, and N. K. Hikmawati, "Analisis Sentimen AicoGPT (Generative Pre-trained Transformer) Menggunakan TF-IDF," *Jurnal Buana Informatika*, vol. 14, pp. 97–106, Nov. 2023.

[14] Abdul Halim Hasugian, M. Fakhriza, and Dinda Zukhoiriyah, "Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes," *Teknologi Sistem Informasi dan Sistem Komputer TGD*, pp. 98–107, 2023, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsk/index

[15] Sekar Mulyani, "Indobertweet Ulasan Beauty Products," Hugging Face.

ISSN 2085-4552

Prostate Cancer Screening for Specific Races Using Bioinformatics and Artificial Intelligence on Genomic Data

David Agustriawan [1], Marlinda Vasty Overbeek [2], Angga Aditya Permana [3]
[1,2,3] Department of Informatics, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Indonesia
[1] david.agustriawan@umn.ac.id, [2] marlinda.vasty@umn.ac.id, [3] angga.permana@umn.ac.id

Accepted 07 August 2024
Approved 14 August 2024

***Abstract*— Prostate cancer is one of a deathly cancer worldwide. The higher incidence and mortality rate shows that it is an urgent call for all of us to fight against it in our own way. This study develops an artificial intelligence system to screening prostate cancer from normal patients in a specific race. Gene expression and its phenotype dataset was downloaded from xenabrowser.net. Data preprocessing and filtering based on a particular race, bioinformatics computational analysis to determine the features and machine learning algorithm such as decision tree and random forest are used to develop AI model. All the procedure and analysis was performed using python programming. The result show that only White and Black African American has a proper number of dataset while Asian and American Indian has a very lack dataset. Differentially expression gene (DEG) analysis was performed to both White and Black African American cancer and normal dataset as a reference. 143 and 1 DEG are found in White and Black African American race respectively. ENSG00000225937.1 (PCA3) is identified as the highest up-regulated gene expression in cancer in both White and Black African American race. The results of DEG analysis then become features to develop Artificial Intelligence (AI) classification system. AI model was developed using decision tree and random forest with GriDSearch parameters optimization and stratified 10-fold cross validation. Both Decision tree and random forest model yield 96% accuracy in training dataset and 93% and 91% accuracy in testing dataset for decision tree and random forest, respectively**

***Index Terms*— *Prostate cancer, Bioinformatics, Artificial Intelligence, Genomic. Machine Learning*.**

I. INTRODUCTION

Prostate cancer is one of deadly disease in a man. Based on cancer statistic 2024 estimated new cancer cases 299,010 and deaths 35,250 in United States [1]. However, the case of prostate cancer is also high in many others countries. Even though a clinical pathway that consist of diagnose, medical treatment and prognostic prediction of prostate cancer is already designed by world health organization (WHO) and health ministry in each country worldwide, the incidence and the mortality is still happened in high number of cases. Prostate cancer risk is 70% higher in black men compared to the white males [2]. Black men living in the United States having the highest risk at 4.2% [3]. Unfortunately, the standard blood test used to screen the prostate cancer, the prostate specific antigen (PSA) test is considered to be in accurate [4]. PSA is failing to identify 8 of every 10 men aged under 60 who later have prostate cancer diagnosed. Therefore, it is really urgent need to do a research to create an accurate prostate cancer test.

Genomic dataset has been used to predict, determine personal treatment and prognosis of a disease. Previously, bioinformatics study design was used to conduct genomic dataset analysis [5]–[8]. Currently, Deep learning and machine learning techniques has been integrated with bioinformatics research pipeline to explore the classification, treatment prediction and regression analysis [9], [10]. Nowadays, the knowledge and technology of genomics, bioinformatics and AI are ready to be used in research and project. However, the research that integrated Genomic, bioinformatics and AI in Health sector is still lack. Therefore, this study will focus utilizing the genomic dataset of prostate cancer to build a classification model utilized bioinformatics and AI study design.

The development of Genomic dataset, Bioinformatics software and tools, advance technology of Artificial Intelligence (AI) can be integrated to create a screening test for prostate cancer. Few studies start to explore a research utilizing genomic dataset, Bioinformatics and Artificial intelligence to develop an AI screening test for prostate cancer. Ramírez-Mena *et al* [11] utilizing 550 samples from TCGA to predict and identify prostate cancer tissue by gene expression with average sensitivity and specificity of 0.90 and 0.8 with AUC of 0.84. Moreover, Yousef et al [12] utilizing machine learning to classify prostate cancer or normal patients using gene - micro RNA (miRNA) pair dataset with the accuracy depends on gene – miRNA group. On

the group 10 with 1 miRNA and 122 genes achieved 0.95 AUC score. Another study by Kaplan and Ertunc [13] was also utilizing gene expression to diagnose prostate cancer that achieved 95.65% accuracy. However, none of those studies consider races as a unique variable in prostate cancer that affect black men incidence and death prostate cancer is higher than other races.

This research addressed racial issue in developing AI system by integrating several AI algorithm and bioinformatics approach to screening prostate cancer using genomic dataset. This study use gene expression and phenotype dataset to separated races and select features for develop AI system using Bioinformatics approach performed differentially expressed gene (DEG) analysis that yields up-regulated and down-regulated gene expression that can be used as features to develop AI system to classify cancer prostate or normal. Moreover, in the future this research design can be elaborated to predict the personal treatment and prognosis. In the future this study will be useful for the precision and personal medicine [14], [15].

II. METHOD

The overall research flowchart for this study is depicted in figure 1. It shows the summary of step by step in conducting the research.

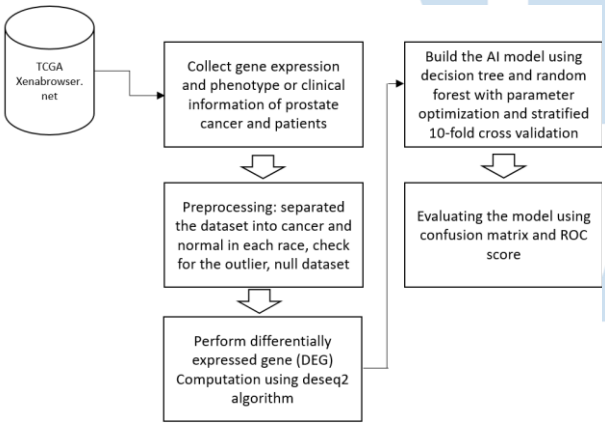


Figure 1. Research flowchart for early detection of prostate cancer using genomic, bioinformatics and AI study design

Data for this study were collected and downloaded from the TCGA project at https://xenabrowser.net/ on April 2024. The selected prostate study is by GDC TCGA using the code "PRAD". Two datasets were downloaded which were Phenotype and gene expression data. All downloaded datasets were saved locally. Data exploration, pre-processing, Bioinformatics computation and AI system development to classify prostate cancer and normal was performed using Python programming.

A. *Dataset Information*

The 551 datasets that were downloaded from the GDC TCGA Prostate Cancer (PRAD) on Xenabrowser.net at https://xenabrowser.net/ on April 2024. This research downloaded HTSeq – Counts gene expression RNAseq and Phenotype dataset. HTSeq – Counts gene expression RNAseq is a genome expression matrix dataset of patients. Moreover, Phenotype is clinical information related to the patients in HTSeq – Counts gene expression RNAseq dataset. Each patient has a unique TCGA ID.

B. *Data Exploration and preprocessing*

The dataset of HTSeq – Counts gene expression RNAseq were normalized. For data exploration and preprocessing, first, this study checks for missing value and outlier dataset. Second, matching TCGA ID between gene expression and phenotype dataset. Third, filtering the dataset based on cancer and normal. Moreover, filtering the data set in each cancer and normal dataset to a specific racial population. All the procedure in data exploration and preprocessing was conducted using Anaconda software and python programming.

C. *Bioinformatics computation Analysis*

This study utilized Bioinformatics methods for the features selection. Since the genome dataset consist of about 60,488 genes therefore this study will not use all the dataset. The differentially expression genes (DEG) dataset will be used for further analysis. DEG computation was performed using pydeseq2 package in python programming. Pydeseq2 is a package that implemented the mathematic procedural of deseq2 algorithm [16]. Commonly DEG calculation package is provided in R algorithm but we found the python version pydeseq2. Furthermore, DEG was performed for white and black race since another race like Asia is lack of the dataset. Top 3 up-regulated and down-regulated genes from DEG computation will be utilized as features to develop AI system.

D. *Develop AI system to screening prostate cancer in a spesific race*

Decision tree and Random forest algorithm was used to build AI system. Since we have imbalanced dataset between the class or label dataset therefore this study uses stratified k-fold cross validation that split the data such that the proportions between classes are the same in each fold as they are in the whole dataset. Stratified sampling is used to overcome the imbalanced issue without losing any inputs [17]. First, the dataset is separated into train and validation dataset. Train dataset will be used to develop the model and the validation dataset is used to test the model. Second, this study builds the model using decision tree and random forest algorithms since three-based algorithms often perform well on imbalanced datasets. Third, to optimize the accuracy, GridSearch parameter optimization was

performed. Lastly, using the best parameter from GridSearch this study build the AI model utilized stratified 10-fold cross validation. All the analysis in developed AI model was performed in Anaconda and python programming.

III. RESULTS AND DISCUSSIONS

A total of 551 patients' prostate cancer and normal was downloaded from xenabrowser.net. It consists of 498, 52, 1 cancer, normal and metastatic patients, respectively. Then, for cancer and normal patients we separated it again based on racial information. This study mapped expression dataset TCGA ID with phenotype dataset TCGA ID to separate the dataset per race. After mapped the TCGA ID this study found a total of 482 cancer patients that consist of 1, 10, 52, 406 patients for American Indian or Alaska native, Asian, Black or African and White race respectively. Moreover, there are 13 patients are not reported for the racial information. Furthermore, in normal population consist of 0, 0, 7, 44 for American Indian or Alaska native, Asian, Black or African and White race respectively and 1 patient is not reported for the racial information. It is shown that the number of genome expression dataset is still lack and still need further effort to make the data available especially for the normal dataset. For further analysis, to compute DEG, this study considers White and Black race. Furthermore, in developing the AI model this study only consider white race since the black race dataset is not proper for the AI model computation.

To compute DEG using pydeseq2, this study combined dataset with labelled information cancer and normal of each white and black race population. In white population the total dataset for DEG computation is 450 patients with 60,488 unique gene ID using ensemble ID. Then we deleted gene with the sum of the expression dataset from all the patients less or equal to 0. It yields only 57,412 genes. Then, baseMean, Log2FoldChange and P-adjusted value was performed to analyze differentially expression genes between cancer and normal population. The same procedure was also performed for the black race population. With the threshold P-adjusted value < 0.05 this study found that there are 143 and 1 genes are differentially expressed in white and black race respectively. DEG in cancer population has two types: (1) up-regulated and (2) down-regulated. Up regulated gene means the expression of a particular gene is higher than normal patients and it can become the potential cause of disease include cancer [18], [19]. In the other hand down regulated gene means the expression of a particular gene lower than the expression in normal that can cause a disease include cancer as well [20], [21] .It is shown that less results found in the black race compare to the white race. It can be caused by the lack of dataset in black race population. Interestingly, the gene that was found in black race ENSG00000225937.1 (PCA3) is also found in the white race with significantly up regulation expression. For further analysis, this study only focus on the white race population since the dataset is still proper for the analysis. Figure 2 is the distribution of the white race population of cancer and normal patients. This study plots all the raw value of input dataset of gene expression. It shows that the separation between patient's cancer and normal can be identified. However, the clustering is not well separated. Therefore, AI model for classification need to be developed.

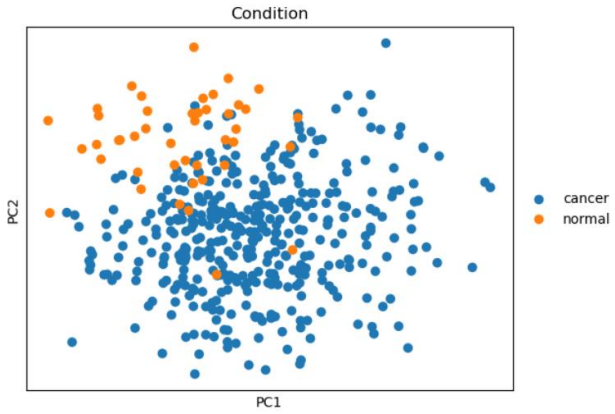


Figure 2. PCA analysis of dataset in white race cancer and normal population using gene expression value

Top 3 up-regulated gene and down-regulated gene resulted from deseq2 algorithm for DEG analysis as shown in table 1 was selected as the features to develop machine learning model for a classification prediction.

Table 1. Deseq2 computation results for white race

Ensembl_ID (Gene symbol)	baseMean	Log2FoldChange	P-adjusted value
ENSG00000225937.1 (PCA3)	12.3	0.6	7.5964E-13
ENSG00000242899.1 (RPL7P16)	11.2	0.4	1.0201E-06
ENSG00000166743.8 (ACSM1)	10.7	0.4	9.4623E-06
ENSG00000196878.11 (LAMB3)	10.6	-0.3	3.6607E-04
ENSG00000244509.3 (APOBEC3C)	10	-0.3	1.1646E-04
ENSG00000101443.16 (WFDC2)	10.2	-0.3	6.9417E-05

Table 1 shows top 3 up-regulated and down-regulated gene in white race prostate cancer compare to the normal population. Top 3 up-regulated gene are PCA3, RPL7P16 and ACSM1 while the top 3 down-regulated genes are LAMB3, APOBEC3C and WFDC2.

This study then evaluates whether those top-3 up-regulated and down-regulated gene can be used to classify the cancer and normal patients in prostate cancer white race population using machine learning analysis. This study use decision tree and random forest classifier. And label information is cancer or normal. This study first developed the model using

decision tree and random forest classifier with stratified 10-fold cross validation without parameters optimization. Stratified 10-fold cross validation was selected since the dataset is imbalanced as shown in figure 3.

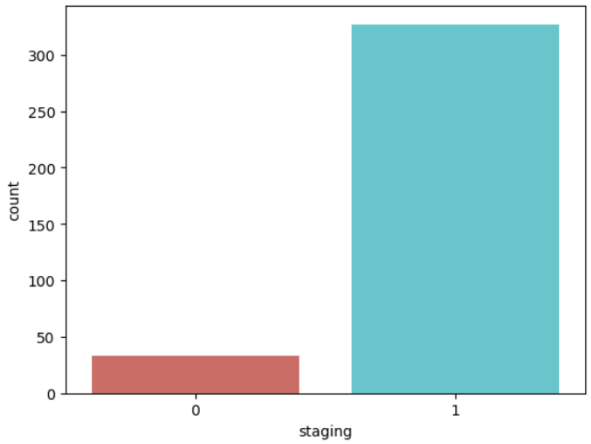


Figure 3. bar plot showing imbalanced dataset

Without GridSearch parameter optimization this study found the accuracy of the training model is 91.4% and 93.9% for decision tree and random forest classification respectively. Furthermore, GridSearch parameter optimization was performed to increase the accuracy of the classification method of decision tree and random forest. For decision tree, this study optimized four parameters criterion, max_depth, max_leaf_nodes and splitter. As the results the best value for each parameter can be seen in table 2. Moreover, for the random forest classification classifier Parameter optimization was also performed for criterion, max_features, n_estimators and min_sample_split and the best value is depicted in table 3.

Table 2. Best parameter optimization from GridSearch calculation for decision tree classifier

Parameter	Best value
Criterion	Entropy
Max_depth	2
Max_leaf_nodes	30
Splitter	Best

Table 3. Best parameter optimization from GridSearch calculation for random forest classifier

Parameter	Best value
Criterion	Entropy
Max_features	Log2
Min_samples_split	20
N_estimators	100

After parameter optimization was performed. The best parameters are used to build AI model using stratified 10-fold cross validation in decision tree and random forest algorithm. In decision tree, this study found there is a slightly increase accuracy score using parameter optimization. The comparison before and after parameter optimization can be seen in table 4.

Table 4. Comparison of evaluation results before and after parameter optimization for decision tree and random forest classifier

Model evaluation parameter	Before parameter optimization	After parameter optimization
Mean accuracy score of decision tree using stratified 10 fold cross validation	0.91	0.96
Mean accuracy score of random forest using stratified 10 fold cross validation	0.94	0.96

Table 4 shows that the training model for decision tree and random forest returned a high accuracy and there is no significant different between decision tree and random forest in building the training dataset AI model.

Test the model was performed using validation dataset. For decision tree for the test model evaluation is shown in table 5 and figure 4

Table 5. Evaluation results from the test dataset of decision tree classifier

Model evaluation type	Score
Accuracy	0.93
Precision	0.96
Recall	0.96
F1	0.96

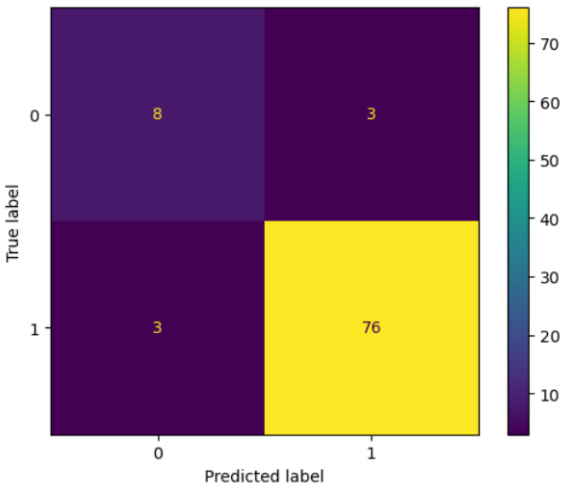


Figure 4. Decision tree confusion matrix from test dataset

Moreover, to evaluate the test result in decision tree scenario, AUC analysis was also performed as shown in figure 5.

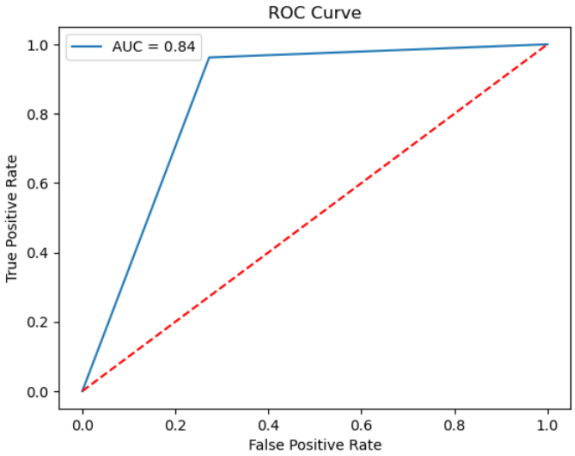


Figure 5. Decision tree ROC curve from test dataset

Figure 5 shows that the AUC score achieved 0.84. it indicates that the algorithm performs pretty well in separating the cancer and normal dataset.

This study was also test the model using random forest classifier and the model evaluation is depicted in table 6 and figure 6

Table 6. Evaluation results from the test dataset of random forest classifier

Model evaluation type	Score
Accuracy	0.91
Precision	0.95
Recall	0.95
F1	0.95

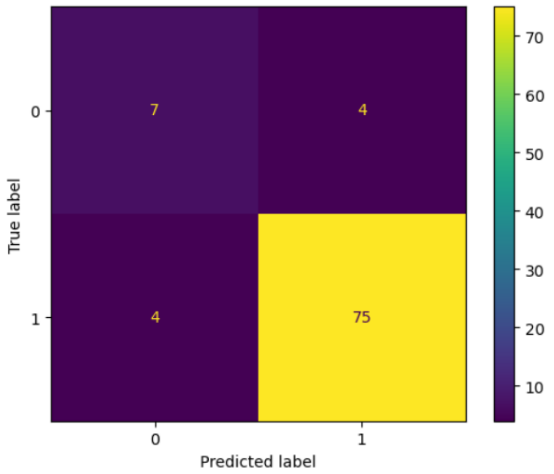


Figure 6. Random forest confusion matrix from test dataset

AUC analysis was also performed for random forest evaluation model for test dataset as shown in figure 7.

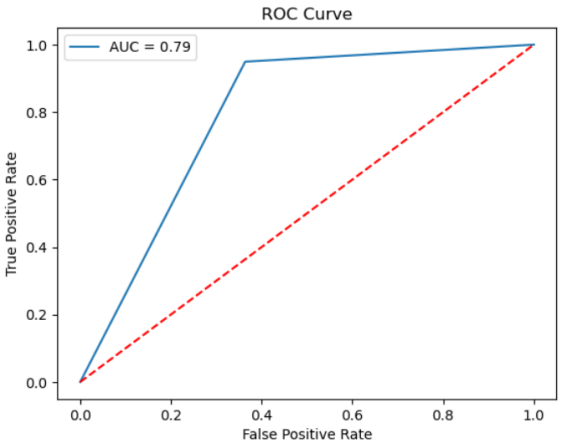


Figure 7. random forest ROC curve from test dataset

Figure 7 shows that the AUC score achieved 0.79. it indicates that score is slightly decreasing compare to the decision tree result.

It is shown that with the new dataset that come from the validation dataset the system still can predict in high accuracy 0.93 and 0.91 using decision tree and random forest algorithm classification classifier. It is shown that our developed model has a very strong dataset and algorithm that can be used as a tool to predict cancer or normal patients in prostate cancer. Through this study as well it is shown that gene expression dataset has a strong or powerful pattern to be used as a features for AI to diagnose a disease

IV. CONCLUSION

This research utilized prostate cancer genomic dataset to classify cancer or normal using bioinformatics and AI approach. The results show that with top 3 up and down regulated genes in specific race are able to screening the prostate cancer from normal with higher accuracy score in training, testing and validation dataset.

Through this study also we can understand that combining bioinformatics approach for feature filtering and selection is a powerful method in selecting the features for AI model development to predict a particular disease.

ACKNOWLEDGEMENT

DA performed major parts of the research and prepared the manuscript. AAP, MVO provide a review on the study methodology and manuscript writing. All authors read and approved the final manuscript. The authors are thankful for Department of Informatics, Multimedia Nusantara University, Serpong Banten, Indonesia, for the support in facility and financial under

internal grant number 0020-RD-LPPM-UMN/P-INT/VI/2024 for the whole research process.

REFERENCES

[1] R. L. Siegel Mph, A. N. Giaquinto, | Ahmedin, J. Dvm, and R. L. Siegel, "Cancer statistics, 2024," *CA. Cancer J. Clin.*, vol. 74, no. 1, pp. 12–49, Jan. 2024, doi: 10.3322/CAAC.21820.

[2] "Press Releases." https://pressroom.cancer.org/ProstateCancerMortality (accessed Aug. 08, 2024).

[3] M. A. Jain, S. W. Leslie, and A. Sapra, "Prostate Cancer Screening," *StatPearls*, Oct. 2023, Accessed: Jul. 20, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK556081/

[4] S. Gottlieb, "Study shows poor reliability of prostate cancer test," *BMJ Br. Med. J.*, vol. 327, no. 7409, p. 249, Aug. 2003, Accessed: Jul. 20, 2024. [Online]. Available: /pmc/articles/PMC1126651/

[5] A. F. Liko, E. Ciputra, N. A. Sanjaya, P. C. Thenaka, and D. Agustriawan, "DNA methylation profiling reveals new potential subtype-specific gene markers for early-stage renal cell carcinoma in Caucasian population," *Quant. Biol.*, vol. 10, no. 1, pp. 79–93, Mar. 2022, doi: 10.15302/J-QB-021-0279.

[6] K. N. Ramanto, K. J. Widianto, S. S. H. Wibowo, and D. Agustriawan, "The regulation of microRNA in each of cancer stage from two different ethnicities as potential biomarker for breast cancer," *Comput. Biol. Chem.*, vol. 93, Aug. 2021, doi: 10.1016/J.COMPBIOLCHEM.2021.107497.

[7] J. Ivan, G. Patricia, and D. Agustriawan, "In silico study of cancer stage-specific DNA methylation pattern in White breast cancer patients based on TCGA dataset," *Comput. Biol. Chem.*, vol. 92, Jun. 2021, doi: 10.1016/J.COMPBIOLCHEM.2021.107498.

[8] M. Gustiananda *et al.*, "Immunoinformatics Identification of the Conserved and Cross-Reactive T-Cell Epitopes of SARS-CoV-2 with Human Common Cold Coronaviruses, SARS-CoV, MERS-CoV and Live Attenuated Vaccines Presented by HLA Alleles of Indonesian Population," *Viruses*, vol. 14, no. 11, Nov. 2022, doi: 10.3390/V14112328.

[9] Z. Zhang *et al.*, "Deep learning in omics: a survey and guideline," *Brief. Funct. Genomics*, vol. 18, no. 1, pp. 41–57, 2019, doi: 10.1093/bfgp/ely030.

[10] Q. Wu *et al.*, "Deep Learning Methods for Predicting Disease Status Using Genomic Data," *J. Biom. Biostat.*, vol. 9, no. 5, 2018, Accessed: Jul. 29, 2024. [Online]. Available: /pmc/articles/PMC6530791/

[11] A. Ramírez-Mena, E. Andrés-León, M. J. Alvarez-Cubero, A. Anguita-Ruiz, L. J. Martinez-Gonzalez, and J. Alcala-Fdez, "Explainable artificial intelligence to predict and identify prostate cancer tissue by gene expression," *Comput. Methods Programs Biomed.*, vol. 240, Oct. 2023, doi: 10.1016/J.CMPB.2023.107719.

[12] M. Yousef, G. Goy, R. Mitra, C. M. Eischen, A. Jabeer, and B. Bakir-Gungor, "Mircornet: Machine learning-based integration of mirna and mrna expression profiles, combined with feature grouping and ranking," *PeerJ*, vol. 9, p. e11458, May 2021, doi: 10.7717/PEERJ.11458/TABLE-12.

[13] K. Kaplan and H. M. Ertunc, "Diagnosis of prostate cancer using gene expression data," *26th IEEE Signal Process. Commun. Appl. Conf. SIU 2018*, pp. 1–4, Jul. 2018, doi: 10.1109/SIU.2018.8404250.

[14] O. Strianese *et al.*, "Precision and Personalized Medicine: How Genomic Approach Improves the Management of Cardiovascular and Neurodegenerative Disease," *Genes (Basel).*, vol. 11, no. 7, pp. 1–24, Jul. 2020, doi: 10.3390/GENES11070747.

[15] S. J. Maceachern and N. D. Forkert, "Machine learning for precision medicine," *Genome*, vol. 64, no. 4, pp. 416–425, 2021, doi: 10.1139/GEN-2020-0131/ASSET/IMAGES/LARGE/GEN-2020-0131F1.JPEG.

[16] D. Rosati *et al.*, "Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review," *Comput. Struct. Biotechnol. J.*, vol. 23, pp. 1154–1168, Dec. 2024, doi: 10.1016/J.CSBJ.2024.02.018.

[17] X. Zhao, J. Liang, and C. Dang, "A stratified sampling based clustering algorithm for large-scale data," *Knowledge-Based Syst.*, vol. 163, pp. 416–428, Jan. 2019, doi: 10.1016/J.KNOSYS.2018.09.007.

[18] E. Scott *et al.*, "Upregulation of GALNT7 in prostate cancer modifies O-glycosylation and promotes tumour growth," *Oncogene*, vol. 42, no. 12, p. 926, Mar. 2023, doi: 10.1038/S41388-023-02604-X.

[19] E. D. Arisan *et al.*, "Upregulated Wnt-11 and miR-21 Expression Trigger Epithelial Mesenchymal Transition in Aggressive Prostate Cancer Cells," *Biol. 2020, Vol. 9, Page 52*, vol. 9, no. 3, p. 52, Mar. 2020, doi: 10.3390/BIOLOGY9030052.

[20] N. C. Whitlock *et al.*, "MEIS1 down-regulation by MYC mediates prostate cancer development through elevated HOXB13 expression and AR activity," *Oncogene*, vol. 39, no. 34, p. 5663, Aug. 2020, doi: 10.1038/S41388-020-01389-7.

[21] V. Nodouzi, M. Nowroozi, M. Hashemi, G. Javadi, and R. Mahdian, "Concurrent Down-Regulation of PTEN and NKX3.1 Expression in Iranian Patients with Prostate Cancer," *Int. braz j urol*, vol. 41, no. 5, pp. 898–905, 2015, doi: 10.1590/S1677-5538.IBJU.2014.0036

ISSN 2085-4552

Improved SVM for Website Phishing Detection Through Recursive Feature Elimination

Feliciano Surya Marcello [1], Farica Perdana Putri [2]
[1,2,3] Department of Informatics, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Indonesia
[1]feliciano.marcello@student.umn.ac.id, [2]farica@umn.ac.id

Accepted 09 January 2025
Approved 16 January 2025

***Abstract*— The website has become one of the most widely used media to obtain information, conduct business, transfer data, and others. The development and use of increasingly sophisticated websites also pose a threat to information security, called cybercrime. Website phishing is one type of cybercrime where the perpetrator creates a fake website that mimics the original one in order to steal sensitive user information. The Support Vector Machine (SVM) algorithm is one way that can be implemented in detecting phishing websites by classifying through checking website features. Selecting relevant features for SVM is important to generalize performance and computational efficiency. Thus, the use of Recursive Feature Elimination (RFE) feature selection is conducted to improve the performance of SVM (SVM-RFE). By eliminating irrelevant features, RFE contributes to the increasing accuracy rate for the SVM by 0.15% with an accuracy of 96.09%. In other experiment, the accuracy results of SVM-RFE significantly improved by approximately 15% to 20%.**

***Index Terms*— *Feature Selection, Recursive Feature Elimination, Support Vector Machine, Website Phishing Detection*.**

I. INTRODUCTION

Behind the rapid development of information technology and the internet, there are threats that can harm users of this technology and internet which is commonly called cybercrime [1]. One of the cybercrimes that often occurs today is phishing. A phishing attack is a cybercrime that uses social engineering to deceive users and steal victims' information data, such as personal identity, important information related to finance, and others. Phishing attacks can be carried out in several ways, such as sending fake messages via email or social media platforms, as well as websites [2], [3], [4]. Users who are not aware of this attack will usually be asked to enter information or download files containing malware that can steal the personal data of the victim's device [5].

Website phishing itself is one of the main problems in website security. Phishers usually send Uniform Resource Locator (URL) to victims via email, SMS, and social media [6]. In the second quarter of 2023, the Anti-Phishing Working Group (APWG) noted that there were 1,286,208 phishing attacks and there were 597,789 phishing website attacks detected. The losses resulting from these attacks are considerable, so improvements need to be made.

A phishing website is a replica of a legitimate website. The entire phishing website is not built for phishing, but only a few pages that are specialized by the perpetrator to provide input or download, so that when data is sent, the data is sent to the attacker. There are several methods of cloning websites, which can be done by using special software or by creating manually [7]. The features used to detect phishing websites fall into three categories, namely URL-based features, content-based features, and external service features. URL-based features are features taken directly from the URL of the website, while content-based features refer to characteristics taken from the content of the web page. The content-based features analyze the items displayed on the website or contained within its HTML code to detect trends or abnormalities that could indicate a phishing attempt, such as hyperlink content and abnormal content. External service features are features obtained by querying third-party services or search engines, such as WHOIS, Alexa, Openpagerank, and Google [1].

Support Vector Machine (SVM) is one of the algorithms in machine learning. In SVM, each data item is mapped as a point in n-dimensional space and this algorithm builds a dividing line for the classification of two classes known as hyperplane [8]. Research conducted in [9] detects phishing websites using the SVM algorithm. The research uses six attributes of the URL, including Long URL, Dots, IP Address, SSL Connection, At (@) Symbol, and Dash (-) symbol which are used as the main features in SVM model training. Study in [10], uses Logistic Regression, k-Nearest Neighbors, Decision Tree, Random Forest, SVM, and Gradient Boosting to detect phishing websites. The results show that the proposed SVM algorithm has the lowest performance. The use of kernels in SVM also cannot handle noisy data and overlapping target classes. Therefore, it can be

concluded that the SVM algorithm still dealing with high-dimensionality in the feature space [10]. Selecting relevant features for SVM classifiers improves generalization performance, computational efficiency, and feature interpretability [11], [12]. To address the issue of high-dimensional feature space, feature selection methods play a crucial role.

Recursive Feature Elimination (RFE) is one of the feature selection methods that performs model selection based on the learnt model and classification accuracy. RFE sequentially removes unimportant features that can cause a decrease in classification accuracy so that after obtaining the best features, this technique rebuilds a new classification model. The model is trained with the training dataset, feature weights that reflect the importance of each feature are obtained. The features that have been sorted by the highest weight, will be reclassified so that the use of the RFE method based on feature importance can be obtained [13]. This study aims to enhance the performance of SVM algorithm by utilizing RFE feature selection for detecting the web phishing.

II. THEORY

A. Recursive Feature Elimination

Recursive Feature Elimination (RFE) is a wrapper method in feature selection. It is a method that works by removing redundant and weak features whose removal affects the training error the least and keeping independent and strong features to improve the generalization performance of the model. RFE initially builds a model on the entire feature set and ranks the features based on their importance. After ranking, the lowest ranked feature is removed and the model is rebuilt and re-ranked for the most important feature [14]. Here are the steps in performing the RFE process, namely:

1. Train the model using all features with 10-fold cross-validation.
2. Calculate the mode performance to determine the value (accuracy, precision, and recall) Determine the confusion matrix to use
3. Comparing features with the highest to lowest weights
4. Discard the feature with the lowest weight
5. Perform iterations to calculate the performance of the model until it is finalized:
   (a) Train and test the model on the most recent features
   (b) Recalculate model performance
   (c) Comparing the features with the highest to the lowest weights
   (d) Discard the feature with the lowest weight
6. Use the selected optimal model

B. Support Vector Machine

Support Vector Machine (SVM) is one of the powerful algorithms in machine learning. In SVM, each data item is mapped as a point in an n-dimensional space and the algorithm constructs a dividing line for classification of two classes known as a hyperplane [8]. A hyperplane is a dividing line between classes by maximizing the margin between them. The margin is the distance between the hyperplane and the nearest point in n-dimensional space in each class. The SVM algorithm is included in the ensemble learning method. This is because the learning system of this model uses a hypothesis space in the form of functions from a high-dimensional feature space. In this space, there are many boundaries that can be used to separate the classes, but there is only one boundary that maximizes the margin [15]. The kernel function plays a crucial role in SVM [16] by transforming the input data into a higher-dimensional space where the data becomes linearly separable, including Linear, Radial Basic Function (RBF), and Polynomial kernel.

Linear kernel is the simplest kernel, without using the gamma value (γ) as in (1). x_i is the value of the training data, x_j is the value of the test data, and $k(x_i.x_j)$ is the kernel value.

$$k(x_i.x_j) = x_i^T x_j \quad (1)$$

RBF is a non-linear kernel, using the gamma value parameter (γ>0) as a determinant of the flexibility of this kernel, can be described by (2). This kernel is suitable for data that cannot be solved linearly with a high level of accuracy and precision [17].

$$k(x_i.x_j) = exp\left(-\gamma.\left\|x_i.x_j\right\|^2\right) \quad (2)$$

Polynomial kernel is a non-linear kernel, using the parameter value of the gamma value (γ > 0) and the value of *d* as the coefficient of the penalty degree for flexibility as described in (3).

$$k(x_i.x_j) = \gamma\left(x_i^T.x_j + r\right)^d \quad (3)$$

A large gamma value will also calculate training data that is far from the decision boundary but will cause a small accuracy value, and the value of *C* is a free parameter, the value of *r* is a bias, and the gamma and *r* parameters have a strong relationship [18].

III. METHOD

The methodology of this study consists of several steps and is presented in this section. The following steps are described in detail: data collection, data preprocessing, and the proposed method.

A. Data Collection

The dataset used is a dataset from Mendeley data entitled Web page phishing detection from [1]. This dataset contains 11,430 data with 87 attributes. This data has been labelled as phishing and legitimate. All feature values used in this study are integer data types. Only the label is a string data type. Three feature categories included 56 URL-based features, 24 content-based features, and 7 external service features. The

labels are equally distributed. Data is divided into 70% testing data and 30% training data.

B. Data Preprocessing

The preprocessing process is divided into three parts, namely data cleaning, standardization, and label encoding. The data cleaning process checks data errors, incorrect data types, duplicates, and empty data. If those exists, the data will be deleted. After data cleaning, the standardization process is carried out to handle outlier data so that it is still in a good scale distribution. Standardization improved model convergence and performance by ensuring features were scaled uniformly. Standardization using the Z-score method as in (4).

$$Z = \frac{x-\mu}{\sigma} \quad (4)$$

where x is a value of data point, μ is mean of data points, and σ is the standard deviation of all data. Z-score standardization transforms features to have a mean of 0 and a standard deviation of 1. This ensures all features contribute equally to the distance-based calculations, particularly for algorithms like SVM.

Labels on imported datasets have a string data type. It is necessary to do a label encoding process to convert category label data into numeric label data since SVM model requires numeric labels to be able to train the model. The LabelEncoder class from scikit-learn is used to transform the status data (label) to change the value of legitimate website to 1 and phishing website to 0.

C. Proposed Method

The process iterates the training process for each feature using SVM and RFE models. We use 10 cross-validation and calculate the feature ranking based on the accuracy score. After the features are sorted, then remove the features that have the lowest weight. The iteration continues until the features reach convergence. Convergence is a condition when the selected features have found the most optimal optimization and the criteria have been met. In this study, the criteria are the number of features that have been determined.

The SVM model uses parameters previously obtained through the gridsearchCV process. The data training process is carried out by checking convergence. Convergence in SVM refers to SVM training to achieve the most optimal solution or close to the optimal solution. If the convergence has not reached the optimal point, then the training iterates with different points and the point vector will be updated. However, if it has converged, then the accuracy of the hyperparameter will be checked. The prediction results will later be used for system evaluation and determine the performance of the RFE and SVM algorithms

IV. Results and Discussions

In this section, the experiment results are described and analyzed. We also explain the model parameters used in RFE and SVM training, which then become the basis for tuning hyperparameters in the improved-SVM model using RFE, which in this article will be referred to as SVM-RFE.

A. Experimental Setup

The RFECV class of scikit-learn is used with parameters in it, namely, the estimator is SVM as a model, CV is cross validation worth 10, and the weight seen is accuracy. In SVM, the parameters used are C, gamma, and kernel. The values of C are 0.01, 0.1, 1. Gamma values are 0.01, 0.1, 1. Kernels used are linear, polynomial, and RBF. Hyperparameters like the penalty parameter C, kernel type, and kernel-specific parameters control the trade-off between model complexity and accuracy or how the decision boundary is shaped. The hyperparameters values are determined based on the previous studies. Training is carried out on the dataset used, and the results sought are the best accuracy of each parameter.

B. Evaluation Measures

Precision, recall, accuracy, and F1-score are used to evaluate the proposed model. Then the evaluation results are visualized using a heatmap, to see the number of correct and incorrect predictions from the test data using the improved SVM model with RFE. The metrics precision, recall, F1-score, and accuracy are calculated as in (5), (6), (7), and (8), respectively.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1-score = \frac{2 \times (Recall \times Precision)}{Recall+Precision} \quad (7)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

C. Results

The results of finding the best hyperparameters using *gridsearchcv* get results that are C = 1.0, gamma = 0.01, and kernel = 'RBF'. The accuracy of the SVM is 0.96092.

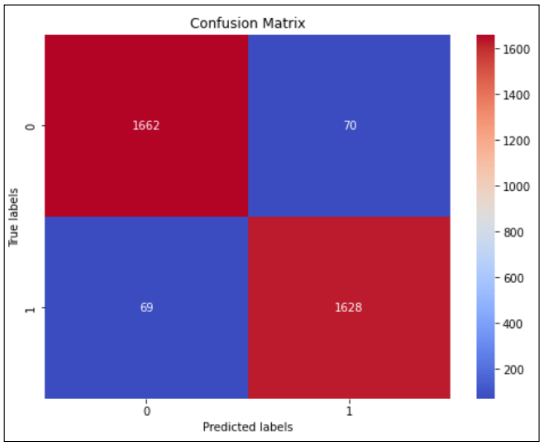


Fig. 1. Confusion matrix without RFE

The evaluation results of the SVM model without RFE can be seen in Figure 1. There are 1662 data for true negative (TN), 69 for false negative (FN), 70 for false positive (FP), and 1628 true positive (TP) data. The evaluation results of the SVM model with RFE are shown in Figure 2. There are 1667 true negative data, 69 false negative data, 65 false positive data, and 1628 true positive data.

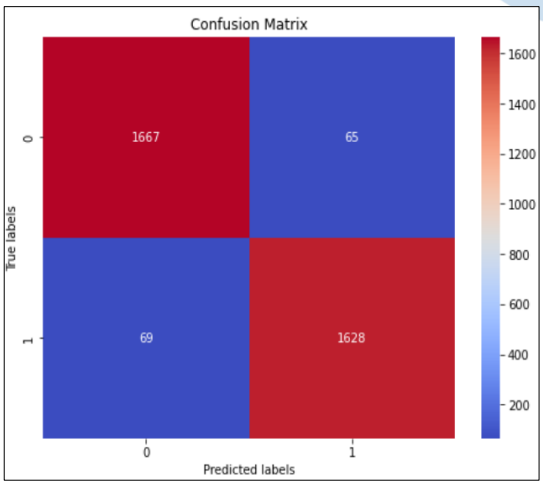


Fig. 2. Confusion matrix with RFE

TABLE I. COMPARISONS OF SVM WITH AND WITHOUT RFE

Method	Precision	Recall	F1-score	Accuracy
Without RFE	0.96	0.96	0.96	0.9594
With RFE	0.96	0.96	0.96	0.9609

In false positives, there is a difference in the number, namely RFE = 70 and without RFE = 65. In Table 1, the precision, recall, and F1-score values have the same results in RFE and without RFE, there is a difference in the accuracy value, namely with RFE 96.09% and without RFE 95.94%. From the results of this analysis, it can be concluded that the use of RFE on the dataset used in this study has a good impact because it has increased accuracy by 0.15%. The significance of a 0.15% improvement in accuracy may appear insignificant on the surface, but it can have a significant real-world impact, particularly in critical applications such as phishing website detection. Phishing website detection systems are deployed on a huge scale, screening millions or billions of webpages each day. A 0.15% increase in accuracy means properly identifying thousands (or possibly more) of extra phishing websites every day, stopping multiple phishing attacks.

TABLE II. FEATURE SELECTION COMPARISONS WITH [1]

Feature Selection	Accuracy
U + C	81.74%
U + E	76.50%
C + E	80.20%
RFE	**96.04%**

To test the effectiveness of the proposed method, namely SVM-RFE, we also conducted a comparison with the research conducted by Abdelhakim Hannousse [1]. In [1], a combination of features was performed, with the following features: URL-based features (U in Table 2), content-based features (C), and external-based features (E). In Table 2, the accuracy results in this current study which uses RFE have significantly improved compared to previous studies with an increase in accuracy of approximately 15% to 20%.

This paper also uses a different dataset to analyse the proposed SVM-RFE algorithm. In research conducted by Muhammad Hasan [9], the research used a different dataset taken from kaggle which has 32 features, and 11,054 data samples. With the same division of training and testing data: 70% training data and 30% testing data. Study in [9] does not use any selection features and uses the full features in the dataset. The results obtained from the study are also not very good accuracy which is 56.05%. Experiments were also conducted for the dataset used in [9] by comparing with SVM-RFE.

The features selected and removed from the new dataset are nine features and the remaining 22 selected features used for classification. The SVM model was performed using gridsearchCV using a polynomial kernel, gamma of 0.1, and C of 1.0 which are the best parameters. SVM-RFE model obtained an accuracy of

95.23%, while [9] only obtained an accuracy result of 56.05%. This result proves that RFE can provide a great improvement to the performance of SVM classification in detecting phishing websites.

V. CONCLUSION

Phishing websites are common and cause harm to ordinary people. SVM is one of the machine learning algorithms used to detect phishing websites. However, SVM has disadvantages if the data used has many features. Therefore, feature selection is used to remove irrelevant and redundant features. The use of RFE feature selection coupled with evaluation using 10 cross-validation, makes feature elimination more accurate and better. The use of standardization is also important for the dataset used so that there is no outlier data so that the accuracy generated when predicting training data with the SVM algorithm becomes more accurate and higher.

In this study, the prediction results have an accuracy of 96.09%. In experiments with different datasets with SVM algorithm and RFE feature selection, there is a significant improvement. Without RFE, different datasets only get 56.05% accuracy, while using RFE increases accuracy to 95.23%. Thus, from several trials conducted, it can be concluded that the use of feature selection Recursive Feature Elimination has a good impact in handling many features and helping the Support Vector Machine algorithm in optimizing accuracy for phishing website classification. RFE feature selection is very large due to the repetition of feature checking that is done, necessitating further research on techniques to improve computation complexity.

ACKNOWLEDGMENT

We would like to thank Universitas Multimedia Nusantara (UMN) for supporting this research

REFERENCES

[1] A. Hannousse and S. Yahiouche, "Towards benchmark datasets for machine learning based website phishing detection: An experimental study," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104347, Sep. 2021, doi: 10.1016/j.engappai.2021.104347.

[2] M. Zouina and B. Outtaj, "A novel lightweight URL phishing detection system using SVM and similarity index," *Hum. Cent. Comput. Inf. Sci.*, vol. 7, no. 1, p. 17, Dec. 2017, doi: 10.1186/s13673-017-0098-1.

[3] A. Safi and S. Singh, "A systematic literature review on phishing website detection techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 590–611, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.004.

[4] M. H. Alkawaz, S. J. Steven, A. I. Hajamydeen, and R. Ramli, "A Comprehensive Survey on Identification and Analysis of Phishing Website based on Machine Learning Methods," in *2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, Malaysia: IEEE, Apr. 2021, pp. 82–87. doi: 10.1109/ISCAIE51753.2021.9431794.

[5] S. Asiri, Y. Xiao, S. Alzahrani, S. Li, and T. Li, "A Survey of Intelligent Detection Designs of HTML URL Phishing Attacks," *IEEE Access*, vol. 11, pp. 6421–6443, 2023, doi: 10.1109/ACCESS.2023.3237798.

[6] H. Abusaimeh, "Detecting the Phishing Website with the Highest Accuracy," *TEM Journal*, pp. 947–953, May 2021, doi: 10.18421/TEM102-58.

[7] T. Chin, K. Xiong, and C. Hu, "Phishlimiter: A Phishing Detection and Mitigation Approach Using Software-Defined Networking," *IEEE Access*, vol. 6, pp. 42516–42531, 2018, doi: 10.1109/ACCESS.2018.2837889.

[8] S. Anupam and A. K. Kar, "Phishing website detection using support vector machines and nature-inspired optimization algorithms," *Telecommun Syst*, vol. 76, no. 1, pp. 17–32, Jan. 2021, doi: 10.1007/s11235-020-00739-w.

[9] A. Abdulwakil, M. A. Aydin, and D. Aksu, "Detecting phishing websites using support vector machine algorithm," *Pressacademia*, vol. 5, no. 1, pp. 139–142, Jun. 2017, doi: 10.17261/Pressacademia.2017.582.

[10] S. M. Mahamudul Hasan, N. M. Jakilim, Md. Forhad Rabbi, and R. M. S. Rahman Pir, "Determining the Most Effective Machine Learning Techniques for Detecting Phishing Websites," in *Applications of Artificial Intelligence and Machine Learning*, vol. 925, B. Unhelker, H. M. Pandey, and G. Raj, Eds., in Lecture Notes in Electrical Engineering, vol. 925. , Singapore: Springer Nature Singapore, 2022, pp. 593–603. doi: 10.1007/978-981-19-4831-2_48.

[11] M. H. Nguyen and F. De La Torre, "Optimal feature selection for support vector machines," *Pattern Recognition*, vol. 43, no. 3, pp. 584–591, Mar. 2010, doi: 10.1016/j.patcog.2009.09.003.

[12] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J Big Data*, vol. 7, no. 1, p. 52, Dec. 2020, doi: 10.1186/s40537-020-00327-4.

[13] H. Jeon and S. Oh, "Hybrid-Recursive Feature Elimination for Efficient Feature Selection," *Applied Sciences*, vol. 10, no. 9, p. 3211, May 2020, doi: 10.3390/app10093211.

[14] T. E. Mathew, "A Logistic Regression with Recursive Feature Elimination Model for Breast Cancer Diagnosis," 2019.

[15] R. I. Arumnisaa and A. W. Wijayanto, "Comparison of Ensemble Learning Method: Random Forest, Support Vector Machine, AdaBoost for Classification Human Development Index (HDI)," *SISTEMASI*, vol. 12, no. 1, p. 206, Jan. 2023, doi: 10.32520/stmsi.v12i1.2501.

[16] Jonathan Leander and Arya Wicaksana, "Optimizing a Personalized Movie Recommendation System with Support Vector Machine and Content-Based Filtering," *JSMS*, vol. 14, no. 1, Dec. 2023, doi: 10.33168/JSMS.2024.0128.

[17] D. K. Choubey, S. Tripathi, P. Kumar, V. Shukla, and V. K. Dhandhania, "Classification of Diabetes by Kernel Based SVM with PSO," *RACSC*, vol. 14, no. 4, pp. 1242–1255, Jul. 2021, doi: 10.2174/2213275912666190716094836.

[18] K. V. Kamran, B. Feizizadeh, B. Khorrami, and Y. Ebadi, "A comparative approach of support vector machine kernel functions for GIS-based landslide susceptibility mapping," *Appl Geomat*, vol. 13, no. 4, pp. 837–851, Dec. 2021, doi: 10.1007/s12518-021-00393-0

ISSN 2085-4552

Implementation of Deep Learning Model for Identification of Skin Diseases by Utilizing Convolutional Neural Network

Lysa Apriani [1], Muhamad Bahrul Ulum [2]
[1,2]Informatics, Esa Unggul University, Jakarta, Indonesia
lysapriani@gmail.com[1], m.bahrul_ulum@esaunggul.ac.id[2]

Accepted 29 October 2024
Approved 15 November 2024

***Abstract*— Skin diseases are health problems that affect many individuals worldwide. Rapid and accurate diagnosis of skin diseases is essential for effective treatment. In an effort to improve diagnosis, information technology and artificial intelligence have taken on increasingly significant roles. This study focuses on the implementation of deep learning models for skin disease identification using CNN architectures EfficientNetB0, Xception and VGG16. The models were trained and tested on a dataset of 1800 images with 5 dermatitis classes and 1 normal class. Confusion matrices were used to assess the performance of the three deep learning models on the components of accuracy, recall, precision, and F1-score. The results of the deep learning model that can classify dermatitis skin diseases with a performance of more than 90% for each evaluation matrix are deep learning models utilizing EfficientNetB0 transfer learning with an accuracy of 93%. In contrast, the Xception model indicates overfitting with a training accuracy of 99.96% and a validation accuracy of 86.38%. The VGG16 model indicates underfitting with a training accuracy of 69.71% and a validation accuracy of 46.79%.**

Index Terms*— Skin Disease Classification; CNN; Transfer Learning; EfficientNetB0; Xception; VGG16.

I. INTRODUCTION

The integumentary system serves as the exterior barrier that safeguards the human body against external environmental factors. The skin, being the outermost epithelial organ in the human body, is very vulnerable to diseases due to its direct exposure to the external environment, which harbors a significant amount of pollutants, germs, and viruses. Epidermal illness, often known as dermatosis, is a pathological disorder that impacts the human skin. These skin diseases can arise due to a multitude of circumstances, encompassing infections, allergies, autoimmune disorders, genetic predisposition, environmental influences, and other contributing variables. Contact dermatitis, actinic curatory, neoplasms, dermatophytosis, acne and granulomas are various skin diseases that surrender certain workers to almost 70-95% [1]. Dermatitis with a prevalence of 10% of cases in the world is an inflammation or skin disease that causes clinical abnormalities in the form of polymorphic collation and itching complaints [2]. Dermatitis, particularly contact dermatitis, constitutes a substantial proportion of dermatologist consultations, accounting for around 4-7% of cases. The prevalence of hand dermatitis is high, with a reported incidence rate of 2% among the general population. Indeed, a 20% prevalence rate of hand dermatitis is observed among women, occurring at least once during their lifetime. Furthermore, the findings from the patch test indicated that 30% of children diagnosed with dermatitis exhibited possible allergens[3]. The data obtained shows a significant increase in dermatitis cases in Indonesia from year to year. In 2019, the percentage of dermatitis incidence reached 60.79%[4].

People without a medical background tend to ignore the early manifestations of skin diseases because of the difficulty in distinguishing one type of skin disease from another. The provision of information that is common on the internet is often inadequate in providing adequate guidance for patients trying to understand the condition of their skin disease. Even though this type of skin disease can be cured today, these diseases have indeed brought problems to the patient's life[5]. Indonesia's Health Profile 2022 shows that restrictions on community activities due to the COVID-19 pandemic in 2022 have hampered efforts to detect early cases of skin diseases [6]. Proper identification of skin diseases is important in order to provide appropriate treatment. At this time, technological developments have taken a big role in the health sector, including the use of deep learning in classifying skin disease types based solely on images.

Previous research comparing CNN, RFC, SVM and KNN algorithms resulted in the CNN algorithm having the highest accuracy of 97.89%, followed by RFC with 87.43% accuracy, SVM with 78.61% accuracy and KNN with 76.96% accuracy[7], conducting research on the development of a CNN model in detecting kulti diseases without utilizing transfer learning resulting in an accuracy of 73%[8], research on large class classifications will cause the model to have many class options so that to correctly classify the class is smaller[9], comparing EfficientNet, ResNet, and VGG

in classifying skin diseases with the result that EfficientNet is the model with the best performance reaching an accuracy of 87.31%[10], using the transfer learning models ResNet50, Inception-V3, Inception-ResNet, DenseNet, MobileNet, and Xception to classify skin diseases with Xception results showing the best accuracy at 97% followed by MobileNet (96%), Inception-ResNet (5%), Inception-V3 (95%), DenseNet (93%) and ResNet50 (87%)[11].

Research reports that the accuracy of skin disease detection using image processing and CNN ranges from 70% to 95% [12]. This study aims to identify skin diseases using a convolutional neural network algorithm by utilizing CNN architectures, namely EfficientNet, Xception and VGG16. This study will compare the performance of each CNN architecture using components in the confusion matrix so that it can provide the best performance architectural results. By incorporating deep learning technology, it is hoped that the results of this study can help in making more accurate and efficient identification of skin diseases, as well as making an important contribution to better diagnosis efforts and more efficient management of skin diseases overall

II. METHOD

The research method in this study consists of dataset collection, data pre-processing, CNN modeling, training, and performance evaluation. The flow of the research methodology can be seen in figure 1.

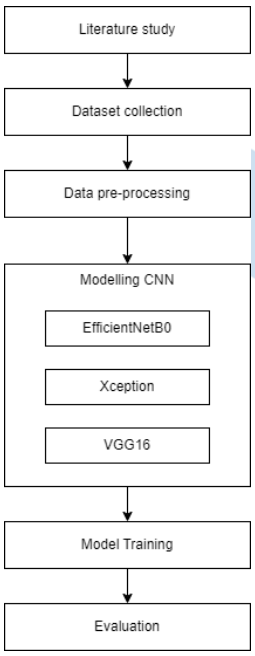


Fig. 1. Research Methods

A. Literature Studies

The literature study stage includes a comprehensive review of relevant scientific literature obtained from websites, articles and journals with the science of skin disease classification.

B. Dataset Collection

At this stage, data collection from other sources was carried out, datasets of allergic contact dermatitis, atopic dermatitis, perioral dermatitis, seborrheic dermatitis, and neurodermatitis used from [13] and the normal class dataset used comes from[14].

The dataset used has the format of *Joint Photographic Experts Group (jpeg/jpg).* This dataset consists of 1800 images with 300 images each for each class. Table I shows the distribution of datasets and classes that will be used in this study.

Table 1. Dataset Distribution

No	Class	Sum
1	Allergic Contact Dermatitis	300
2	Atopic Dermatitis	300
3	Seborrheic Dermatitis	300
4	Dermatitis Perioral	300
5	Neurodermatitis	300
6	Normal	300

C. Data Pre-processing

In the processing process, several things will be done, including changing the size of the image, augmenting data, and distributing the dataset. The image resizing is carried out because the dataset has a different size, so the image size is changed before the process so that the data has the same size so that the model can receive appropriate input. In this study, the image size was changed to 224 x 224 pixels.

Fig. 2 Data augmentation

Data augmentation is a way to multiply the number of images[15]. Augmentation is a technique that aims to enhance the quantity and variety of training data by strategically applying specific alterations to pre-existing data. To enhance the generalization of the model, several augmentations will be implemented. These include rotation with a maximum value of 0.7 for both right and left rotation, flip right and left with a value of 0.5, flip up and down with a value of 0.5, zoom random with a value of 0.5, crop random with a value of 0.5, random brightness with a value of 0.5, and random contrast with a value of 0.5.

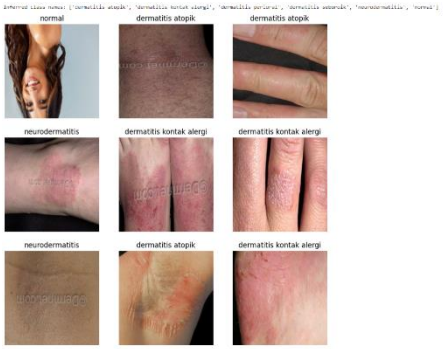


Fig. 3 Examples of images that have been augmented

The process of distributing the training, validation, and test datasets involves a method designed to carefully divide the dataset into distinct segments for training, testing, and validation, while ensuring that the proportion of each class is consistently maintained across all sections. This approach is crucial for preserving the integrity and balance of the dataset, allowing the model to learn effectively from the training data and be accurately evaluated on the test and validation sets. In the context of this study, the dataset has been allocated in a manner where 80% of the data is used for training the model, 10% is reserved for testing its performance, and the remaining 10% is dedicated to validating the model's accuracy and generalization capabilities. This proportional distribution is intended to optimize the model's learning process and ensure robust and reliable results.

D. Modelling CNN

Following the pre-processing of the dataset, the subsequent step involves the development of a high-performance deep learning model. The modeling procedure involves conducting multiple experimental iterations using various hyperparameters in order to identify the model that yields the highest level of accuracy. By iteratively optimizing accuracy through trial and error to identify the hyperparameters that yield optimal outcomes.

Figure 4 illustrates the CNN architecture employed in the present investigation. The employed architectural design has five distinct layers, specifically the layer base model, batch normalizing step, dense layer, dropout layer, and fully linked layer.

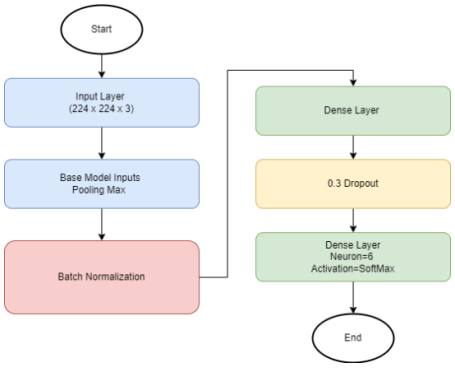


Fig. 4 Applied CNN architecture

The hyperparameters used in this study are shown in table II. EfficientNetB0, VGG16 and Xception were chosen as CNN architectures for detecting dermatitis skin disease classes. The performance of the three architectures will be compared to find the best architecture. The data is split into 80% for training, 10% for validation, and 10% for testing. With a batch size of 64, this means that if there are 4800 training samples, the CNN algorithm will take 64 data samples from the total 4800 available. These samples will then be trained by the neural network until complete, after which it will take the next 64 samples, continuing this process until all data is processed. While 5 layers means the architecture consists of 5 layers that have been explained in figure 4.1. Epoch 25 means the training process is carried out as many as 25 iterations or repetitions which are then applied earlystopping callbacks that can stop training if according to the defined metrics, then there is adamax optimizer which is used to iteratively improve weights based on training data, and the use of learning rate of 0.001 where the smaller the learning rate, the model will learn the training data in more detail, the learning rate also applies reduce_lr callbacks which will reduce the learning rate if the defined metrics are met.

Table 2. Hyperparameters

No	Hyperparameter	
1	CNN Architecture	EfficientNetB0, Xception, VGG16
2	Batch Size	64
3	Epoch	25
4	Optimizer	Adamax
5	Learning rate	0.001

E. Training

At this stage, the training stage is carried out with the dataset owned on the selected architecture. The training process is used using training data and updating the model using optimization algorithms. In the training model, callbacks such as

'ReduceLROnPlateau', 'ModelCheckpoint' and 'EarlyStopping' are applied which are used to maximize the model training process. ReduceLROnPlateau is used to reduce the learning rate when the monitored metric, namely 'val_loss', does not improve. ModelCheckpoint is used to store the best model based on the best 'val_accuracy' metric or validation accuracy, ensuring the most optimal model is not lost during training. EarlyStopping is used to stop training or training if the monitored metric 'val_accuracy' stops increasing, this aims to prevent overfitting and save computing resources. By implementing these callbacks, the training process will become more efficient and adaptive to validation performance.

```
from tensorflow.keras.callbacks import ReduceLROnPlateau, EarlyStopping, ModelCheckpoint

reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=3, min_lr=1e-5)
model_checkpoint = ModelCheckpoint('model-Model-New.h5', monitor='val_accuracy', mode='max', verbose=1, save_best_only=True)
early_stopping = EarlyStopping(monitor='val_accuracy', patience=5, restore_best_weights=True)

history = model.fit(train_ds, validation_data=val_ds, epochs=initial_epochs, callbacks=[model_checkpoint, reduce_lr])
```

Fig. 5 Application of callbacks in model training

F. Evaluation

This study will analyze the work of the classification model by using the confusiation matrix as the core evaluation instrument. The confusion matrix is an evaluation tool that provides a detailed overview of the performance of a classification model that shows how correct a model is in grouping data [16]. The confusion matrix is shown in table 3.

Table 3 Confusion Matrix

Data Classes	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual negative	False Positive (FP)	True Negative (TN)

True Positive (TP) indicates a correct prediction when the model accurately classifies data as positive, and the data is actually positive. True Negative (TN) represents a correct prediction when the model correctly identifies data as negative, and the data is indeed negative. On the other hand, False Positive (FP) occurs when the model incorrectly classifies negative data as positive, while False Negative (FN) happens when the model incorrectly labels positive data as negative. These components of the confusion matrix are essential for calculating key metrics such as accuracy, precision, recall, and F1-Score, which are used to assess and compare the performance of the model.

Equation (1) shows the formula for calculating accuracy.

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

Equation (2) shows the precision calculation formula.

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

Equation (3) shows the recall calculation formula.

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

Equation (4) shows the formula for calculating the F1-score.

$$F1 - Score = 2\,x\,\frac{Precision\ x\ Recall}{Precision+Recall} \quad (4)$$

III. Results and Discussions

In this section, it will be explained in detail how the results have been obtained, starting from the results of model training, and the results of the confusion matrix test. Table 4 shows the results of the training process of each archivist.

Table 4 Model Training Results

No	Architecture	Acc	Val Acc	Loss	Val Loss
1	EfficientNetB0	0.99	0.93	0.03	0.24
2	Xception	0.99	0.86	0.54	0.84
3	VGG16	0.69	0.46	0.86	1.31

EfficietNetB0 performs very well in testing and training. With a validation accuracy of 0.9987 and a training accuracy of 0.9343, this model shows that it can effectively learn patterns from training data and generalize them to new data (data validation).

Xception shows signs of overfitting. Given the very high validation accuracy (0.9996) and lower training accuracy (0.8638), it seems that this model may be overly dependent on the training set and unsuitable for generalization. The much larger loss value in the validation data (0.8412) compared to the training data (0.5400) further supports this.

VGG16 shows overfitting and gives substandard results. The model's poor training accuracy (0.4679) and low validation accuracy (0.6971) indicate that the model has difficulty generalizing previously unknown data and learning patterns from the training set. It can be seen from the high loss value in the validation data (1.3071) and training data (0.8697) that this model is not able to maximize learning.

Macro averages consider precision, recall, and F1-score for each class individually, without applying any weights based on the class distribution. On the other hand, the weighted average also takes into account precision, recall, and F1-score for each class but incorporates weights that correspond to the size or importance of each class. This means that in a weighted average, classes with more samples or higher relevance have a greater influence on the overall metric.[17]. Table 5 presents the confusion matrix outcomes derived from the calculations of three models.

Table 5 Confusion matrix calculation results

No	Architecture	Acc	Macro Avg			Weighted Avg		
			Precision	Recall	F1-score	Precision	Recall	F1-Score
1	EfficientNetB0	0.93	0.93	0.93	0.92	0.93	0.93	0.92
2	Xception	0.86	0.84	0.84	0.83	0.84	0.85	0.84
3	VGG16	0.46	0.62	0.49	0.47	0.49	0.47	0.47

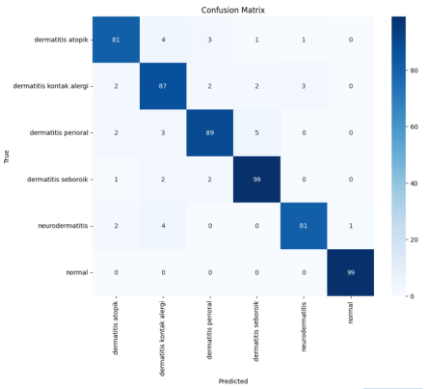


Fig. 6 EfficientNetB0 confusion matrix results

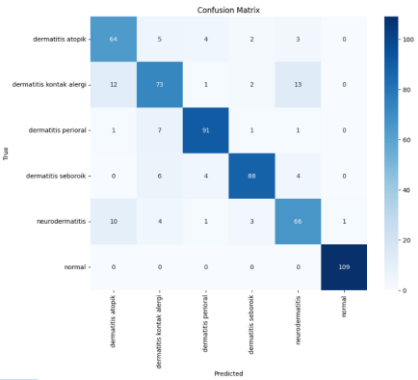


Fig. 7 Xception confusion matrix results

Figure 6 show the confusion matrix of EfficientNetB0 model. EfficientNetB0 model performs very well in both testing and training. This model is not overfitting, as evidenced by the low loss values on training and validation data. The precision, recall, and F1-Score values for Macro Average are 0.93, 0.93, and 0.92 respectively, then for the precision, recall, and F1-Score values for Weighted Average are 0.93, 0.93, and 0.92 respectively, which are consistently the same between the two. This approach not only concentrates on the majority class but also effectively manages the minority class, as seen from the high macro and weighted average performance that is constant across all classes. This shows that the model is reliable and has strong generalization across various types of data in the dataset.

Figure 7 show the confusion matrix of Xception model. The Xception model shows signs of overfitting. The macro average values for precision, recall and F1-score are 0.84, 0.84, and 0.83 respectively and the precision, recall and f1-score values for the weighted average are 0.84, 0.85, and 0.84.

The decrease in the weighted average and macro values indicates that the performance of this model is not evenly distributed across all classes. The reason why the performance of this model is inconsistent across datasets may be that this model over-learns the majority class and ignores the minority class.

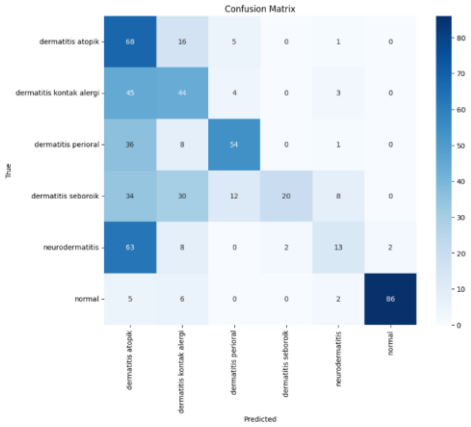


Fig. 8 VGG16 confusion matrix results

Figure 8 show the confusion matrix of VGG16 model. The VGG16 model shows overfitting and produces subpar results. The precision, recall, and F1-score values for the macro average are 0.62, 0.49, and 0.47, respectively, which are lower than the weighted average values of 0.49, 0.47, and 0.47, respectively. The very low weighted average and macro values indicate that the model's performance is not balanced across all classes and cannot manage classes that are in the majority or not.

IV. Conclusion

Based on the findings from the study on the implementation of Deep Learning models for skin disease identification using Convolutional Neural Networks (CNNs), it can be concluded that the process involves several key stages. These stages include data collection, image pre-processing, integrating transfer learning models, batch normalization, adding dense layers, incorporating dropout layers, and final dense layers. The study applied three transfer learning models—EfficientNetB0, Xception, and VGG16—utilizing 1800 images, which were augmented to expand the training dataset.

The performance of the skin disease classification model was evaluated using a confusion matrix, where the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values played a crucial role in calculating metrics such as recall, precision, and F1-Score. These metrics were essential in determining the model's accuracy. Among the various experiments conducted (A to C), the best result was achieved in experiment A using the EfficientNetB0 model, which obtained a training accuracy of 93%. Validation using the confusion matrix confirmed this performance, yielding an accuracy of 93%, precision of 93%, recall of 93%, and an F1-Score of 92%.

For further development, it is imperative to conduct research employing diverse transfer learning architectures. Additionally, expanding the dataset size to encompass a higher number of instances can yield improved accuracy outcomes. Furthermore, the utilization of convolutional neural network (CNN) architectures with varying hyperparameters can effectively mitigate the issues of overfitting and underfitting

References

[1] J. S. Park, E. K. Park, H. K. Kim, and G. S. Choi, "Prevalence and risk factors of occupational skin disease in Korean workers from the 2014 Korean working conditions survey," *Yonsei Med. J.*, vol. 61, no. 1, pp. 64–72, 2020, doi: 10.3349/ymj.2020.61.1.64.

[2] A. S. Maudani, M. Ikhtiar, and A. Baharuddin, "Analisis Spasial Penyakit Dermatitis di Puskesmas Labakkang Kabupaten Pangkep," *Ikesma*, vol. 16, no. 1, p. 51, 2020, doi: 10.19184/ikesma.v16i1.16998.

[3] R. Apriliani, S. Suherman, E. Ernyasih, N. Romdhona, and M. Fauziah, "Hubungan Personal Hygiene Dengan Kejadian Dermatitis Kontak Iritan Pada Pemulung Di Tpa Bantargebang," *Environ. Occup. Heal. Saf. J.*, vol. 2, no. 2, p. 221, 2022, doi: 10.24853/eohjs.2.2.221-234.

[4] G. Soegiarto, M. S. Abdullah, L. A. Damayanti, A. Suseno, and C. Effendi, "The prevalence of allergic diseases in school children of metropolitan city in Indonesia shows a similar pattern to that of developed countries," *Asia Pac. Allergy*, vol. 9, no. 2, p. e17, 2019, doi: 10.5415/apallergy.2019.9.e17.

[5] L. S. Wei, Q. Gan, and T. Ji, "Skin Disease Recognition Method Based on Image Color and Texture Features," *Comput. Math. Methods Med.*, vol. 2018, 2018, doi: 10.1155/2018/8145713.

[6] Kemenkes RI, *Profil Kesehatan Indo-nesia*. 2022. [Online]. Available: https://www.kemkes.go.id/downloads/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-2021.pdf

[7] T. A. Rimi, N. Sultana, and M. F. Ahmed Foysal, "Derm-NN: Skin Diseases Detection Using Convolutional Neural Network," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 1205–1209, 2020, doi: 10.1109/ICICCS48265.2020.9120925.

[8] J. Boman and A. Volminger, "Evaluating a deep convolutional neural network for classification of skin cancer Evaluating a deep convolutional neural network for classification of skin cancer," 2018.

[9] A. Rafay and W. Hussain, "EfficientSkinDis: An EfficientNet-based classification model for a large manually curated dataset of 31 skin diseases," *Biomed. Signal Process. Control*, vol. 85, no. March, p. 104869, 2023, doi: 10.1016/j.bspc.2023.104869.

[10] R. Sadik, A. Majumder, A. A. Biswas, B. Ahammad, and M. M. Rahman, "An in-depth analysis of Convolutional Neural Network architectures with transfer learning for skin disease diagnosis," *Healthc. Anal.*, vol. 3, no. January, 2023, doi: 10.1016/j.health.2023.100143.

[11] Bergi Veeresha Gowda, Likith Kumar S, Gadilingappa G V, K. Likhith, and G Sai Ranga, "Skin Disease Detection using Image Processing and CNN," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 286–295, 2023, doi: 10.48175/ijarsct-9751.

[12] M. Groh *et al.*, "Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 1820–1828, 2021, doi: 10.1109/CVPRW53098.2021.00201.

[13] D. Bala *et al.*, "MonkeyNet: A robust deep convolutional neural network for monkeypox disease detection and classification," *Neural Networks*, vol. 161, pp. 757–775, 2023, doi: 10.1016/j.neunet.2023.02.022.

[14] D. Putri Ayuni, Jasril, M. Irsyad, F. Yanto, and S. Sanjaya,

"Augmentasi Data Pada Implementasi Convolutional Neural Network Arsitektur Efficientnet-B3 Untuk Klasifikasi Penyakit Daun Padi," *Zo. J. Sist. Inf.*, vol. 5, no. 2, pp. 239–249, 2023, doi: 10.31849/zn.v5i2.13874.

[15] L. Qadrini, A. Sepperwali, and A. Aina, "Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial," *J. Inov. Penelit.*, vol. 2, no. 7, pp. 1959–1966, 2021.

[16] D. Morales, E. Talavera, and B. Remeseiro, "Playing to distraction: towards a robust training of CNN classifiers through visual explanation techniques," *Neural Comput. Appl.*, vol. 33, no. 24, pp. 16937–16949, 2021, doi: 10.1007/s00521-021-06282-2.

ISSN 2085-4552

Comparison of Multilinear Regression and AdaBoost Regression Algorithms in Predicting Corrosion Inhibition Efficiency Using Pyridazine Compounds

Yudha Mulyana[1], Muhamad Akrom[2], Gustina Alfa Trisnapradika[3], Nabila Putri Setiawan[4]
[1,2,3,4]Informatics Engineering Program, Dian Nuswantoro University, Semarang, Indonesia
[1]yudha.myn@gmail.com,[2] m.akrom@dsn.dinus.ac.id, [3] gustina.alfa@gmail.com,[4] nabilasei@gmail.com

Accepted 25 October 2024
Approved 08 January 2025

***Abstract*— Corrosion is a severe problem in various industries, leading to increased production costs, maintenance, and decreased equipment efficiency. Using organic compounds as corrosion inhibitors has become an increasingly desirable solution due to their effectiveness and environmental friendliness. This study compares the performance of two machine learning algorithms, Multilinear Regression (MLR) and AdaBoost Regression (ABR), in predicting the corrosion inhibition efficiency (CIE) of pyridazine-derived compounds. The dataset consists of molecular properties as independent variables and CIE values as targets. A k-fold cross-validation using the k=10 process was used to measure the model's performance, where the dataset was divided into equal subsets. Each iteration uses one subset as validation data and the other as training data. Results show that the AdaBoost Regression model achieves higher accuracy (99%) than Multilinear Regression (98%) in predicting CIE. Important feature analysis showed that Total Energy (TE) and Dipole Moment (μ) were the most influential variables in the ABR model, highlighting their important role in inhibitor effectiveness. Model evaluation was performed with R2 and RMSE metrics, where nonlinear models such as ABR were shown to be superior in predicting corrosion inhibition efficiency. These findings support using nonlinear methods to improve the effectiveness of protecting industrial equipment from corrosion.**

***Index Terms*— Pyridazine; Corrosion inhibitors; AdaBoost Regression; Multilinear Regression; Corrosion inhibition efficiency**.

I. INTRODUCTION

Corrosion is a significant phenomenon in industry that can result in large economic losses, increased production costs, and decreased efficiency and equipment life [1]. The main corrosion causes include air humidity, metal contamination with acidic compounds, sulfur, and other corrosive gases. According to the definition of the International Union of Pure and Applied Chemistry (IUPAC), corrosion is a natural process that occurs in materials such as metals, ceramics, and polymers [2]. In this process, the material reacts with its surrounding environment, resulting in the degradation or deterioration of the material. Significant impacts of corrosion include increased maintenance costs, decreased equipment efficiency, and losses in production [1],[3].

To address the problem of corrosion, the use of corrosion inhibitors has become a major focus to reduce its adverse effects [4]. To tackle corrosion problems, much recent research has led to the use of corrosion inhibitors made from organic compounds for industrial metals [5]. These compounds are chosen because they are considered environmentally friendly, cost-effective, and effective in preventing corrosion [6]. These organic compounds can create a molecular layer that effectively protects metal surfaces from direct attack by corrosive substances, which is one of the main reasons for their use [7].

Pyridazine-derived compounds attract attention because they have diverse molecular structures, including functional groups such as sulfur, nitrogen, and oxygen that affect the adsorption ability of inhibitors on steel surfaces [3]. These compounds have been widely studied in the context of corrosion inhibition through various experimental methods such as gravimetry, potentiodynamic, and impedance spectroscopy [8].

Although experimental research provides valuable insights, it often requires significant time, cost, and resources [9]. As an alternative, Machine Learning (ML) approaches are becoming increasingly popular for evaluating corrosion inhibitor compounds [10]. ML facilitates the development of predictive models based on quantitative relationships between molecular structures and their properties and activities [11], [12]. Predictive models that have been developed allow prediction of the level of corrosion inhibition effectiveness of pyridazine-derived compounds [3].

Multilinear Regression (MLR) is a popular method in regression analysis that is effective for describing the linear relationship between input (X) and output (Y)

variables [13]. MLR works by determining the best line that represents the relationship between variables X and Y [14]. The main advantage of MLR is its ease of interpretation, which helps us understand the significant influence of input variables on output variables [3].

Adaptive Boosting Regression (ABR) often referred to as AdaBoost, is an ensemble learning method that improves model performance by combining several weak models into one strong model [1], [15]. ABR works by giving weight to each model used, where models that have low performance will be given a higher weight to be improved in the next iteration [16]. The advantage of ABR is its ability to handle overfitting and improve prediction accuracy with different model combinations [17], [15].

Many studies on the effectiveness of pyridazine compounds as corrosion inhibitors have utilized experimental methods, including gravimetry, potentiodynamics, and impedance spectroscopy [8], [18]. However, these experimental methods often require a lot of cost, time, and resources [19], [20]. As a solution, Machine Learning (ML) approaches offer a more efficient way of evaluating corrosion inhibitor performance. Using ML, it is possible to predict the effectiveness of inhibitors by identifying quantitative relationships between a compound's structure and its molecular properties [21], [14], [22]. Research by [14] shows that the artificial neural network (ANN) model has better performance than multilinear regression (MLR), with RMSE, MSE, and MAPE values of 111.5910, 10.5637, and 10.2362, respectively. Another study by Haikal et al. showed that the Decision Tree Regression (DTR) model is more effective than MLR, with R^2, MAE, and RMSE values of 0.908, 2.059, and 2.704, respectively [3].

However, our model achieved even lower error metrics in this study, making it more accurate and effective. Specifically, our model reached R^2, MSE, RMSE, and MAE values of 0.990, 0.004, 0.057, and 0.040, respectively, significantly outperforming the results from [14] and [3]. This improvement in model accuracy indicates the effectiveness of our approach in predicting the corrosion inhibition efficiency of pyridazine-derived compounds, highlighting the potential of more refined ML models in this field [23]. Therefore, this study aims to assess and compare the performance of Multilinear Regression (MLR) and AdaBoost Regression (ABR) models in forecasting the corrosion inhibition efficiency of pyridazine-derived compounds. This study implemented data normalization techniques in the pre-processing phase and used k-fold cross-validation using k=10 in the Machine Learning (ML) model-building process [24], [13]. the results of this study can provide a more accurate model and can be an important contribution to the development of more effective corrosion inhibitors in the future [6].

II. METHOD

Fig 1 illustrates the process of developing this study's Machine Learning (ML) model. The initial stage was the selection of the pyridazine dataset, followed by data preprocessing to address scale differences and sensitivity to outliers [3]. After that, the ML algorithm was selected to model the relationship between input and output variables that can accurately predict the corrosion inhibition efficiency of the pyridazine dataset. Furthermore, the model was trained using the k-fold cross-validation technique, which helps avoid overfitting and obtain a more generalized model. Evaluation of model performance was done using several metrics that are often used in regression model evaluation, namely Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). These steps not only helped select the optimal ML model but also ensured that the model could produce accurate predictions of the corrosion inhibition efficiency of pyridazine in this study.

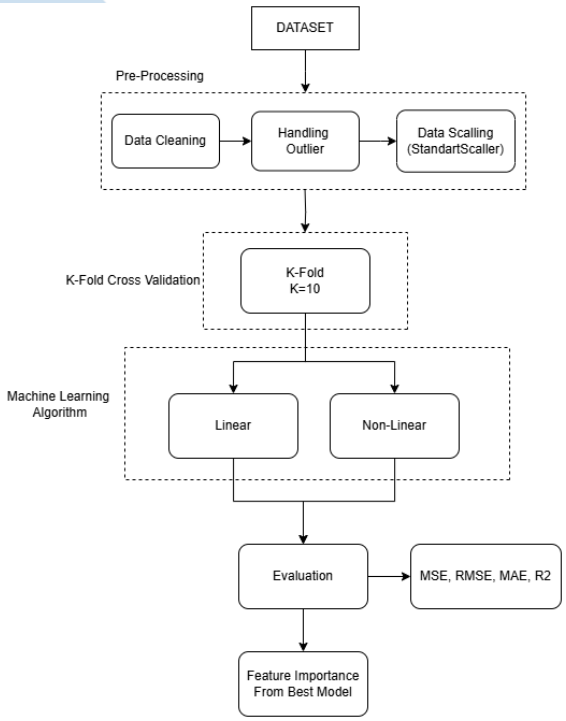


Fig 1. Development of Machine Learning Model

A. Dataset

The dataset used in this study is a collection of pyridazine-derived compound data published by Quadri et al. (2022) [14]. This data collection comprises 20 pyridazine compounds with various quantum molecular properties, used as independent variables and CIE values as dependent variables. The features included in this analysis are total energy (TE), HOMO (Highest Occupied Molecular Orbital), LUMO (Lowest Unoccupied Molecular Orbital), energy gap (ΔE), dipole moment (μ), ionization potential (I), electron

affinity (A), electronegativity (χ), global hardness (η), global softness (σ), and fraction of transferred electrons (ΔN).

B. Data Preprocessing

The initial stage of developing a Machine Learning (ML) model is to perform data normalization. This normalization process is fundamental in the pre-processing stage, as it ensures that all features are on a uniform scale. This helps to improve the performance of ML algorithms, especially if the features in the dataset have a wide range of values. In this study, the StandardScaler technique is used for normalization, transforming the data to have a mean of 0 and a standard deviation of 1. This approach effectively minimizes the impact of outliers, allowing the algorithm to perform more optimally [25], [13]. In addition, researchers applied the Interquartile Range (IQR) method to remove outlier data [18]. This step was taken to cleanse the dataset of unusual or unrepresentative data that could interfere with interpreting the analysis results. By calculating the IQR, researchers can identify and eliminate extreme data that may hurt model performance [6], [19].

C. K-Fold Cross-Validation

Researchers applied the k-fold cross-validation method with k = 10 to split the data into 10 equal parts [6]. At each iteration, the model was trained with 9 subsets and tested with the remaining subsets, alternating through all subsets [1]. This process aims to identify the model with the least error rate, thus ensuring model robustness and reliability [24]. The selection of k = 10 is done to maximize the use of data and minimize bias and variance [23]. The selection of k=10 was done to maximize the use of data and minimize bias and variance [23]. The selection of the k-fold value is tailored to the dataset's characteristics, although k=5 or k=10 values are commonly used in machine learning. This study evaluated the model's performance across various fold values from 1 to 10. After testing, it was determined that k=7 provided the best accuracy for this dataset. Therefore, we selected k=7 for the final model, reported in the results section. This explanation is included here to clarify the choice of k and its rationale. These steps were carefully designed to ensure the dataset's quality, with data preprocessing as a foundational step.

The ultimate goal is to create an accurate and consistent machine-learning model for predicting the corrosion inhibition efficiency of pyridazine derivatives [6]. Therefore, data preprocessing is an important basis in building a reliable and relevant model for this study.

D. Machine Learning Algorithm

In this research, various linear and non-linear regression algorithms have been applied to project CIE values [3]. This study involves a comparison between linear algorithms, which typically utilize linear relationships between input and output variables, and non-linear algorithms that can manage more complex and non-linear interactions between the two variable [6],[10]. Linear algorithms assessed include multilinear regression (MLR), ridge, lasso, Elastic-Net (EN), Support Vector Regression (SVR), and Generalized Linear Model (GLM), all prioritizing a linear relationship between input and output variables [6]. On the other hand, non-linear algorithms explored include random forest (RF), k-nearest neighbors (KNN), nu-support vector regressor (NuSVR), decision tree regressor (DT), gradient boosting regressor (GBR), orthogonal matching pursuit (OMP), kernel ridge (KR), partial least square (PLS), AdaBoost regressor (ABR), and bagging regressor (BR), all of which are designed to address more complex and non-linear relationships between input and output variables [1], [3], [6].

Through these experiments, the main objective was to identify the most effective algorithm to estimate how effective corrosion inhibition by pyridazine-derived compounds is [26]. The results show that the best model for predicting the efficiency of corrosion inhibitors is the AdaBoost Regressor (ABR). ABR is a boosting algorithm used to improve the accuracy of predictive models.[10]. The prediction of the AdaBoost Regressor is based on a combination of several weak learners [27]. Each weak learner contributes to the final prediction through a weighted vote, where the weight reflects the accuracy of the weak learner [28].

Multilinear Regression is a statistical technique for modeling linear relationships between input (X) and output (Y) variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

where β is the regression coefficient that shows the effect of each independent variable X on the dependent variable$Y$, and? is the error term [29]. The MLR model was implemented using the scikit-learn library in Python.

AdaBoost Regression is an ensemble learning technique that aims to improve prediction accuracy by combining several weak models into one strong model [28],[6]. ABR works by assigning weights to each observation, and models that have high prediction errors get a more considerable weight in the next iteration [30]. The ABR algorithm is implemented using the scikit-learn library in Python.

The final prediction of H(x) from the AdaBoost Regressor is calculated as follows:

$$H(x) = \sum_{m=1}^{M} \alpha_m h_m(x)$$

where H(x) is the final prediction of the ensemble model,h_m is the weight of the mth weak learner, h_m is the prediction of the mth weak learner for input x, and M is the total number of weak learners in the ensemble [6]. The weighta_m is calculated based on the accuracy of each weak learner.

E. Model Evaluation

Model performance is evaluated using three main metrics, namely the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), which are evaluation metrics used to assess model performance [12], [31], [32]. The coefficient of determination (R^2) measures the effectiveness of independent variables in explaining changes in dependent variables, with values between 0 and 1, where values closer to 1 indicate that the model is more accurate [27], [33]. The MAE gives the average absolute error between the predicted and true values, providing a more intuitive understanding of the error rate [34]. RMSE measures the average prediction error but provides a larger penalty for more significant errors [10]. RMSE values closer to 0 indicate a more accurate model [30], [35]. By implementing k-fold cross-validation and utilizing these evaluation metrics, this study aims to provide a deep insight into the effectiveness of Multilinear Regression and AdaBoost Regression models in predicting the corrosion inhibition efficiency of pyridazine-derived compounds [3], [19].

F. Important Features

The analysis of important features in this study is a crucial step that requires an in-depth understanding of the relationship between the molecular and physicochemical characteristics of pyridazine-derived compounds and their corrosion inhibition efficiency [33]This method not only helps identify the most influential features in the prediction but also provides deeper insights into the corrosion-inhibiting mechanism of action of such compounds [33].[3]. Thus, the important feature analysis results provide a solid basis for selecting the most relevant features to construct accurate and effective Machine Learning models [27]. The information obtained from salient feature analysis also has important implications in the context of industrial applications and further research [6]. The discovery of the most significant features in demonstrating corrosion inhibition effectiveness can be used to direct the development of more cost-effective and environmentally friendly corrosion-inhibiting materials [36]. In addition, a deeper understanding of the relationship between the molecular properties of pyridazine-derived compounds and their performance as corrosion inhibitors provides opportunities for further research in the optimization of inhibitor formulations that can be widely applied in the corrosion industry [37]. Therefore, the analysis of important features is not only an important part of this research but also results in important impacts on the development of knowledge and technology in the effort to control corrosion [3], [6], [33].

III. Results and Discussions

The results of this study, including a comparison of the performance of Multilinear Regression (MLR) and AdaBoost Regression (ABR) models in estimating the corrosion inhibition effectiveness (CIE) of pyridazine-derived compounds, are illustrated in Table 1. This table presents the results of both models based on evaluation metrics such as the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

TABLE I. Model Performance

Model	MLR and ABR model evaluation			
	MSE	***RMSE***	***MAE***	***R_2***
MLR	0.006	0.078	0.063	0.982
ABR	0.004	0.057	0.040	0.990

The ABR model shows an R^2 value closer to 1 and smaller MAE and RMSE values compared to the MLR model, proving that ABR has better prediction performance.

Figures 2, 3, 4, and 5 present the prediction distribution visualization and residual error plots for each model. These two types of visualizations are used to conduct a more detailed analysis of each model's accuracy.

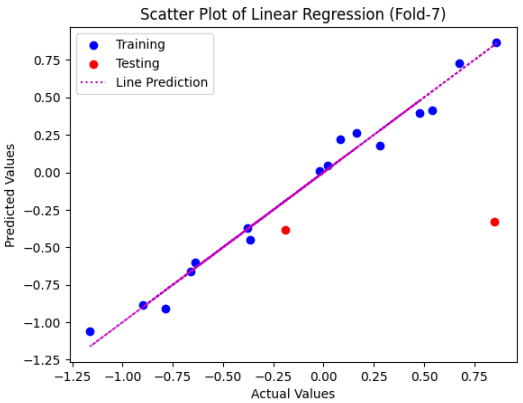


Fig 2. Scatter plot of MLR

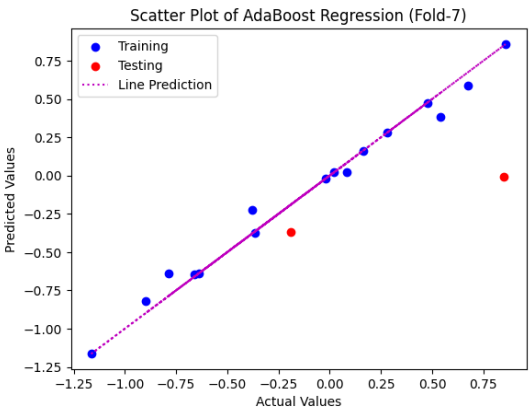


Fig 3. Scatter plot of ABR

In Figures 2 and 3, from the observed results, the distribution of prediction data for the AdaBoost Regressor (ABR) model looks closer to the line representing the true value (fitting line) compared to Multilinear Regression (MLR). This indicates that the ABR model has better prediction performance in processing the pyridazine derivative compound dataset. This means that the ABR model can produce more accurate predictions and closer to the original value than MLR. In addition, Figures 4 and 5 display the residual error, which is the difference between the actual and predicted values, showing that the distribution of prediction errors for the ABR model is closer to the 0 line than the MLR model. This shows that ABR produces smaller forecast errors, indicating higher accuracy in this model.

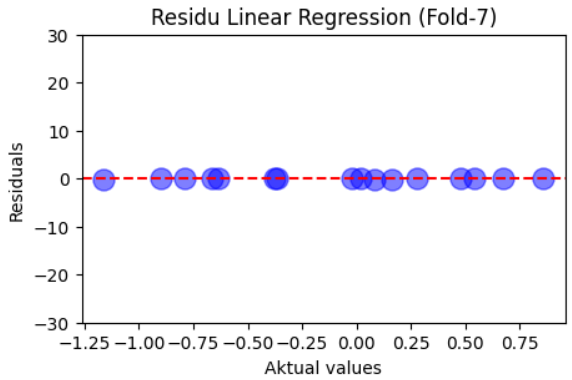


Fig 4. Residual Error on MLR model

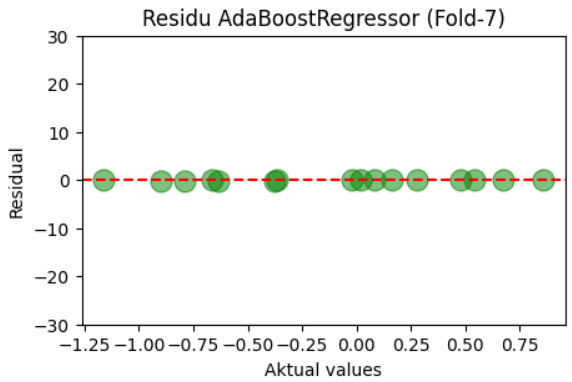


Fig 5. Residual Error on ABR model

Fig 6 shows the performance comparison between Multilinear Regression (MLR) and AdaBoost Regressor (ABR) based on R2 values. From the results shown, it can be seen that the ABR model consistently achieves R2 values close to 0.99 across all folds evaluated, demonstrating its ability to consistently explain around 95% of the variability in the data. In contrast, the MLR model showed more significant variation in R2 values, ranging from 0.82 to 0.98. This analysis suggests that the ABR model has a higher and more consistent accuracy in predicting pyridazine efficiency than the MLR model. The ability of the ABR model to consistently achieve R2 values close to 1 indicates that it can represent the relationship between input and output variables with a high degree of accuracy. Therefore, based on this evaluation, the ABR model is considered a more suitable choice to predict pyridazine derivatives' corrosion inhibition efficiency. This is supported by the ABR model's ability to provide accurate and consistent predictions and explain the variations in the evaluation data.

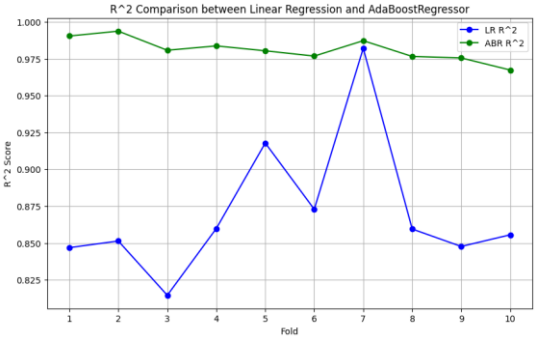


Fig 6. Comparison of the best linear (MLR) and nonlinear (ABR) algorithms for 10-fold trials.

The selection of (ABR) as the superior model is reinforced by the results of the feature importance analysis, which identifies key features that influence the prediction of (CIE). The important features found were total energy (TE), LUMO, dipole moment (μ), global softness (σ), HOMO, and ionization potential (I). Further analysis in Figure 12 shows that total energy (TE) and LUMO are the two most significant features in the (ABR) model. This shows that (ABR), with its ability to capture complex patterns, is better than other models such as (MLR) in predicting (CIE), making it a more appropriate choice for corrosion inhibition prediction.

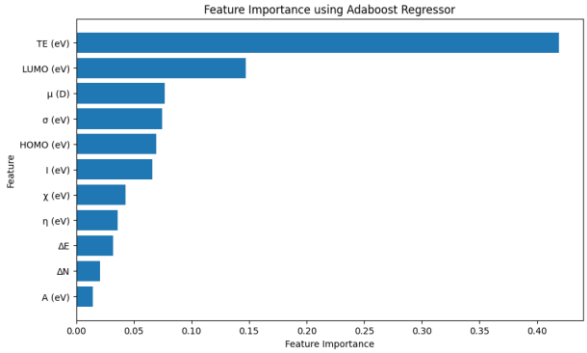


Fig 7. Feature Importance of ABR

IV. Conclusion

This study compares multilinear regression (MLR) and AdaBoost regression (ABR) algorithms to predict the corrosion inhibition effectiveness (CIE) of pyridazine derivatives. The results show that ABR is superior to MLR, with higher R^2 values (0.990 vs. 0.982) and smaller MAE and RMSE values. ABR also showed superior predictive performance and more effectively captured the complex relationship between molecular features and CIEs.

Feature analysis identified total energy (TE) and dipole moment (μ) as significant factors influencing CIE. The research methodology, including k-fold cross-validation and evaluation metrics, ensured robust and reliable results. These findings demonstrate the potential of ensemble models such as ABR in improving the prediction and efficiency of corrosion inhibitors, making an important contribution to developing more effective and environmentally friendly inhibitors.

ACKNOWLEDGMENT

The researchers would like to thank Allah SWT for His abundance of grace, making it easy to complete this research. We would also like to thank the Supervisor, Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Dian Nuswantoro University, for the support and facilities provided during this research process. We also want to express our deepest gratitude to our parents for their unending prayers, support, and motivation during this research process. Hopefully, the results of this research will be useful for the advancement of science and the welfare of mankind

REFERENCES

[1] Adryan Putra Sumarjono, M. Akrom, and G. Alfa Trisnapradika, "Comparison of the Best Machine Learning Model to Predict Corrosion Inhibition Capability of Benzimidazole Compounds," 2023.

[2] Ramadhan and T. L. Pratiwi, "SUPPRESSION OF CORROSION RATE OF ZINK IN HCl SOLUTION Utilization of Plants as Organic Inhibitors for Devotion Of Corrosion Rate Of Zinc in Hydrochloric Acid Utilization of Plants As Organic Inhibitors ... (Aditia Ramadhan, et al.)," vol. 14, no. 1, pp. 1-8.

[3] M. Haikal, M. Akrom, and G. A. Trisnapradika, "Comparison of Multilinear Regression and Decision Tree Regressor Algorithms in Predicting Pyridazine Corrosion Inhibition Efficiency," *Edumatic J. Educ. Inform.*, vol. 7, no. 2, pp. 307-315, Dec. 2023, doi: 10.29408/edumatic.v7i2.22127.

[4] Budi, M. Akrom, G. A. Trisnapradika, and T. Sutojo, "Optimization of Polynomial Functions on the NuSVR Algorithm Based on Machine Learning: Case Studies on Regression Datasets," vol. 10, no. 2, pp. 151-158, 2023, doi: 10.15294/sji.v10i2.43929.

[5] G. Against and L. Corrosion, "Effect of water hyacinth extract organic inhibitor addition on corrosion rate," vol. 2, no. 2, 2018.

[6] Y Mulyana, M. Akrom, and G. A. Trisnapradika, "Comparison of Linear and Non-Linear Machine Learning Algorithms for Predicting the Effectiveness of Plant Extracts as Corrosion Inhibitors," vol. 11, no. 1, 2024.

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

[8] E. H. R. Salim, M. M. B. Hammouti, and D. S. C. S. M. Almutairi, "Electrochemical Studies on New Pyridazinium Derivatives as Corrosion Inhibitors of Carbon Steel in Acidic Medium,"*J. Bio- Tribo-Corrosion*, vol. 0, no. 0, p. 0, 2019, doi: 10.1007/s40735-018-0195-3.

[9] 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(1), 2023.

[10] V. Putranto, M. Akrom, and G. A. Trinapradika, "Implementation of Polynomial Function in Gradient Boosting Regressor Algorithm: Regression Study on Expired Medicines Dataset as Anticorrosion Material,"*J. Technol. and Manaj. Inform.*, vol. 9, no. 2, pp. 172-182, 2023, doi: 10.26905/jtmi.v9i2.11192.

[11] S. Budi, M. Akrom, H. Al Azies, U. Sudibyo, T. Sutojo, G. A. Trisnapradika, A. N. Safitri, A. Pertiwi, and S. Rustad, Implementation of polynomial functions to improve the accuracy of machine learning models in predicting the corrosion inhibition efficiency of pyridine-quinoline compounds as corrosion inhibitors, KnE Eng., pp. 78-87, 2024.

[12] Leni, "Selection of Optimal Machine Learning Algorithm for Prediction of Mechanical Properties of Aluminum,"*J. Energy Engines, Manufacturing, and Mater.*, vol. 7, no. 1, p. 35, 2023, doi: 10.30588/jeemm.v7i1.1490.

[13] Ambarwari, Q. Jafar Adrian, and Y. Herdiyeni, "Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification,"*J. RESTI (Engineering Systems and Technol. Information)*, vol. 4, no. 1, pp. 117-122, 2020, doi: 10.29207/resti.v4i1.1517.

[14] W. Quadri*et al.*, "Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors," *Mater. Today Commun.*, vol. 30, no. October 2021, p. 103163, 2022, doi: 10.1016/j.mtcomm.2022.103163.

[15] Juni, "Comparison of Linear and Non- Linear Machine Learning Algortima for Predicting the Effectiveness of Plant Extracts as Corrosion Inhibitors," 2024.

[16] Irma Prianti, "Comparison of K-Nearest Neighbor and Adaptive Boosting Methods in Multi-class Classification Cases,"*J Stat. J. Ilm. Theor. and Appl. Stat.*, vol. 13, no. 1, pp. 39-47, 2020, doi: 10.36456/jstat.vol13.no1.a3269.

[17] B. Aji, "Faculty of Engineering - Muria Kudus University153," *Pros. 4th SNATIF 2017*, no. 2015, pp. 153-160, 2017.

[18] Luo*et al.*, "A new pyridazine derivative synthesized as an efficient corrosion inhibitor for copper in sulfuric acid medium: Experimental and theoretical calculation studies," *J. Mol. Liq.*, vol. 341, p. 117370, 2021, doi: 10.1016/j.molliq.2021.117370.

[19] Akrom and T. Sutojo, "Investigation of QSPR-Based Machine Learning Models in Pyrimidine Corrosion Inhibitors," vol. 20, no. 2, pp. 107-111, 2023.

[20] T. Ser, Ž. Petar, and M. W. Wong, "Applied Surface Science Prediction of corrosion inhibition e ffi ciency of pyridines and quinolines on an iron surface using machine learning-powered quantitativestructure- property relationships," vol. 512, no. February, 2020, doi: 10.1016/j.apsusc.2020.145612.

[21] Hassan*et al.*, "Heliyon Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium," vol. 6, no. April, 2020, doi: 10.1016/j.heliyon.2020.e05067.

[22] Sutojo, "OPEN A machine learning approach for corrosion small datasets," 2023, doi: 10.1038/s41529-023-00336-7.

[23] antoni wibowo, "10 Fold-Cross Validation," vol.

(accessed, 2003.

[24] Khoirunnisa and N. G. Ramadhan, "Improving malaria prediction with ensemble learning and robust scaler: An integrated approach for enhanced accuracy," *J. Infotel*, vol. 15, no. 4, pp. 326-334, 2023, doi: 10.20895/infotel.v15i4.1056.

[25] Subhi*et al.*, "Implementation of Scaling Techniques in a Website-Based Server Balancing Management System," *J. Comput. and Appl.*, vol. 09, no. 02, pp. 316-326, 2021.

[26] R. Rosyid, L. Mawaddah, and M. Akrom, "Investigation of Regression Machine Learning Models on Medicinal Compounds as Corrosion Inhibitors," no. Ml, 2024, doi: 10.33364/algoritma/v.21-1.1598.

[27] Anjum, K. Khan, W. Ahmad, A. Ahmad, M. N. Amin, and A. Nafees, "Application of Ensemble Machine Learning Methods to Estimate the Compressive Strength of Fiber-Reinforced Nano-Silica Modified Concrete,"*Polymers (Basel).*, vol. 14, no. 18, 2022, doi: 10.3390/polym14183906.

[28] P. Solomatine and D. L. Shrestha, "AdaBoost.RT: A boosting algorithm for regression problems,"*IEEE Int. Conf. Neural Networks - Conf. Proc.*, vol. 2, no. August 2004, pp. 1163-1168, 2004, doi: 10.1109/ijcnn.2004.1380102.

[29] Wulandari, A. Perdana Windarto, H. Qurniawan, and G. Artikel, "Application of Multiple Regression in Estimating the Amount of Population Growth in Siantar District Article Info ABSTRACT,"*JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 4, pp. 2828-9099, 2022, doi: 10.55123/jomlai.v1i4.1677.

[30] A. Mortara, M. Permatasari, A. Desiani, Y. Andriani, and M. Arhami, "Comparison of C4.5 and Adaptive Boosting Algorithms in Alzheimer's Disease Classification," *J. Technol. and Inf.*, vol. 13, no. 2, pp. 196-207, 2023, doi: 10.34010/jati.v13i2.10525.

[31] Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,"*PeerJ. Comput. Sci.*, vol. 7, p. e623, 2021, doi: 10.7717/peerj-cs.623.

[32] O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," no. 2, pp. 5481-5487, 2022.

[33] Leng*et al.*, "Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass," *Biochar*, vol. 4, no. 1, 2022, doi: 10.1007/s42773-022-00183-w.

[34] A. Suryanto, "Application of Mean Absolute Error (Mea) Method in Linear Regression Algorithm for Rice Production Prediction,"*Saintekbu*, vol. 11, no. 1, pp. 78-83, 2019, doi: 10.32764/saintekbu.v11i1.298.

[35] Dzisofi Amelia, S. Wahyuni, and D. Harisuseno, "Evaluation of Satellite Data Suitability as an Alternative to Evaporation Data Availability in Wonorejo Reservoir,"*J. Tech. Irrigation*, vol. 12, no. 2, pp. 127-138, 2021, doi: 10.21776/ub.pengairan.2021.012.02.05.

[36] Akrom and C. Science, "Investigation Of Natural Extracts As Green Corrosion Inhibitors In Steel Using Density Functional Theory Extracts of Natural Products as Green Corrosion Inhibitors on Steel Nowadays, the development of," vol. 10, no. 01, pp. 89-102, 2022.

[37] Nayem, A. C. Mohammad, and K. Mohamed, "An overview of green corrosion inhibitors for sustainable and environment friendly industrial development An overview of green corrosion inhibitors for sustainable and environment ... An overview of green corrosion inhibitors for sustainable and environment friendly industrial Nayem Hossain, Mohammad Asaduzzaman Chowdhury & Mohamed,"*J. Adhes. Sci. Technol.*, vol. 0, no. 0, pp. 1-18, 2020, doi: 10.1080/01694243.2020.1816793

ISSN 2085-4552

Implementation of Gamification Method and Fisher-Yates Shuffle Algorithm for Design and Development Django Learning Application

Ade Kiswara[1], Fenina Adline Twince Tobing[2]*,
Cian Ramadhona Hassolthine[3], Muhammad Ikhwani Saputra[4]
[1,2] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[3,4] Informatics Study Program, Universitas Siber Asia, Jakarta, Indonesia
ade.kiswara@student.umn.ac.id[1], fenina.tobing@umn.ac.id[2*],
cianhassolthine@lecturer.unsia.ac.id[3], muhammadikhwani@lecturer.unsia.ac.id[4]

Accepted 06 January 2025
Approved 16 January 2025

***Abstract*— The web framework emerges as a solution to enhance web development efficiency. Django, an open-source web framework written in the Python programming language, is one of the popular frameworks. Currently, there are not many programming learning platforms that provide specific programming learning materials for Django, implementing a method to boost user interest in using the platform. This research aims to design and build a web-based Django learning application using gamification methods designed based on the octalysis framework to enhance user learning interest. It also incorporates the Fisher-Yates shuffle algorithm to randomize questions for more variety. The application was tested by several users by filling out a questionnaire prepared using the Hedonic Motivation System Adoption Model (HMSAM). The evaluation results of the application obtained an average percentage of 84,15% in the aspect of behavioral intention to use, which means users strongly agree that the djangoing application generates a desire to use it again in the future. Furthermore, the results in the aspect of immersion were 81,44%, which means users agree that the djangoing application creates an immersive learning experience for the Django framework.**

***Index Terms*— Django, Fisher-Yates Shuffle, Gamification, Learning Application, Website**.

I. INTRODUCTION

Technological developments in this digital era have had a big impact in making human life easier. One of the technological developments that we have experienced is in the education sector, where technology has contributed significantly to the teaching and learning process through website development. The use of websites in the teaching and learning process will help speed up access to information, enable interaction between teachers and students, and make it easier to share learning materials. [1].

A website is a page that contains information that can be accessed from all over the world via the internet as long as there is a connection to the internet network. To create a website, several elements such as JavaScript, HTML, and CSS are required. This is very important so that a web address can operate well and display a quality website [2]. However, website development without using a web framework or using vanilla methods has disadvantages, including a lack of efficiency and time effectiveness [3].

With advances in technology, the emergence of web frameworks has become a solution to overcome the weaknesses in vanilla methods. Web frameworks are basically tools that help developers to build websites so that they avoid bugs and can save time in developing websites [4].

Web frameworks come in various types, developers can choose the framework they want to use according to their individual needs [4]. An example of a framework that allows it to be used in website development is Django. Django is a web framework that is open source and written in the Python programming language. Due to its rapid development features, Django is highly sought after by developers today as it takes very little time to build any application [5].

Currently, there are not many programming learning platforms that provide discussions about programming learning materials, especially Django, which implement a method to increase users' interest in learning when using the platform. Increasing user motivation in learning can be done by implementing a method called gamification. This is proven by previous research which discusses the implementation of gamification methods in learning the React Native framework [6]. From this research, results were obtained in the form of implementing gamification

methods in learning which can increase focus and interest in using applications in the future.

During the learning process, quizzes will be provided based on the material that has been studied. The mechanism used in a quiz is randomization of the questions provided, so an algorithm is used to randomize the quizzes irregularly using Fisher-Yates Shuffle on the quizzes provided. There is previous research that discusses the implementation of the Fisher-Yates Shuffle algorithm to randomize exam questions [7]. From this research, results were obtained in the form of applying the Fisher-Yates Shuffle algorithm which can randomize questions optimally and efficiently. There are also previous researchers who discussed the comparison between the performance of the Fisher Yates Shuffle algorithm and the Comparison of Linear [8]. From this research, the results obtained were that the performance of the Fisher Yates Shuffle algorithm was faster than the Comparison of Linear, namely 11.76% of the total of five data tests.

Based on the background of the existing problems and referring to previous research, a django learning platform will be designed and built using website-based gamification methods. It is hoped that by creating this learning application it can attract anyone who has an interest in learning and developing websites using the Django framework through the material provided.

II. THEORY

A. *Django*

Django is a web framework used to develop a project. Django emphasizes its advantages in fast and practical development, and has a simple design. Django was built by experienced developers to solve many problem in web development. By using HTML, CSS, and JavaScript the frontend appearance of the project is made more user friendly. Meanwhile, the Python programming language is used for the backend [9]. Compared to other web frameworks that use the MVC (Model View Controller) design structure, Django uses another design structure, namely MVT (Model View Template) [10].

B. *Gamifikasi*

Gamification is the application of game mechanisms and activities to contexts where the primary goal is not gaming [11]. In more detail, gamification can be defined as an idea that combines game-based mechanics, aesthetic elements, and game thinking to engage others, motivate action, encourage learning, and solve problems [12]. Gamification is now increasingly commonly used in non-gaming related applications to improve user experience and engagement [13].

C. *Octalysis Framework*

Octalysis Framework is the result of research from Yu-Kai Chou in the form of establishing a gamification framework designed using eight core drives that focus on human behavior such as epic meaning, accomplishment, empowerment, ownership, social influence, scarcity, unpredictability, and avoidance [14].

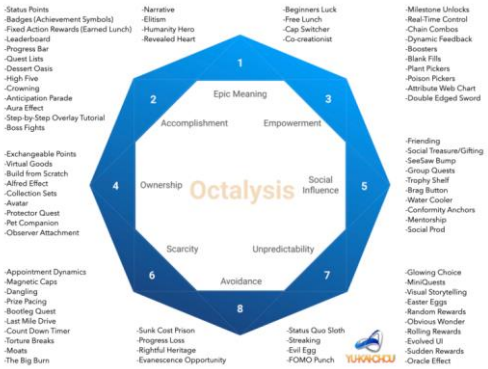


Fig. 1. Core Drive Octalysis Framework

D. *Algoritma Fisher-Yates Shuffle*

The Fisher-Yates algorithm (taken from the names of Ronal Fisher and Frank Yates) or also known as the Knuth Shuffle (taken from the name of Donald Knuth), is an algorithm that produces random permutations of a finite set, in other words randomizing a finite set. arrange. If implemented correctly, the results of this algorithm will not be biased so that each permutation has the same probability [15]. The way to generate random permutations from questions number 1 to N is as follows [16].

1. Write down all the questions from number 1 to N (total index number).
2. Next, determine non-sequentially a number K (one index) up to the number of numbers that have not been crossed out.
3. After that, look for the question number K that has not been crossed out from the bottom, then write down the number in a different place.
4. Repeat steps 2 and 3 until all question numbers are crossed out.
5. Thus, the sequence of question numbers recorded in step 3 will form a random permutation of the original question number sequence.

E. *Blackbox Testing*

Black box testing is a software testing technique used to determine the functionality of an application. The main focus of black box testing is the input available to an application and the expected output for each input value. This blackbox testing method is based on software requirements and specifications [17].

F. *Hedonic Motivation System Adoption Model (HMSAM)*

HMSAM is a model that has been adapted from the Hedonic Motivation System (HMS) as a model that

functions to measure hedonic motivation in a system [18].

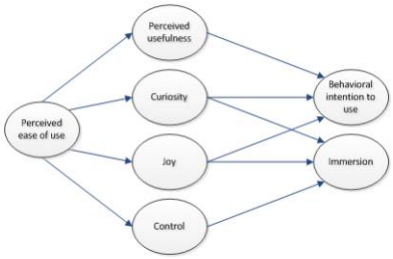


Fig. 2. Model HMSAM.

In Figure 2 there are several factors for measurement in HMSAM, namely as follows.

1. Perceived ease of use is a measure of the ease of using a system.
2. Perceived usefulness is a measure of the usage performance of a system.
3. Curiosity is a measure of the user's level of curiosity about using a system.
4. Joy is a measure of the feeling of pleasure obtained when interacting with the system.
5. Control is a measure of user perception of interaction with the system.
6. Behavioral intention to use is a measure of the user's desire to use the application.
7. Focused immersion is a measure of how focused the user is in using the application.

HMSAM has the final result in the form of behavioral intention to use and also focused immersion. Behavioral intention to use is determined by two aspects, namely perceived usefulness and curiosity. Meanwhile, focused immersion is determined by two other aspects such as joy and control.

G. Likert Scale

The Likert scale is a type of psychometric scale that is often used in questionnaires, and is a common choice in research studies such as surveys [19]. The Likert scale is used to assess the views, opinions and perceptions of individuals or groups regarding social phenomena [20]. The Likert scale consists of five categories with respective points, ranging from strongly agree to strongly disagree.

TABLE 1. LIKERT SCALE CATEGORY INDICATORS

Category	*Point*
Strongly Agree (SS)	5
Agree (S)	4
Neutral (N)	3
Disagree (TS)	2
Strongly Disagree (STS)	1

When the testing process on respondents is carried out, the HMSAM questionnaire will be used. The results of the questionnaire scores can be calculated to get a percentage using formula 1 for the HMSAM criteria with a positive connotation, while formula 2 is used for the HMSAM criteria with a negative connotation.

$$Percentage = \frac{(SS * 5) + (S * 4) + (N * 3) + (TS * 2) + (STS * 1)}{5 * Total\ Respondent}\ x\ 100\%$$

$$Percentage = \frac{(STS * 5) + (TS * 4) + (N * 3) + (S * 2) + (SS * 1)}{5 * Total\ Respondent}\ x\ 100\%$$

From the results of percentage calculations, the respondents' assessment levels were obtained which were categorized based on the percentage intervals listed in Table 2.

TABLE 2. PERCENTAGE INTERVAL FOR RESPONDENT ASSESSMENT CATEGORIES

Percentage Interval	*Category*
80% - 100%	Strongly Agree (SS)
60% - 79,99%	Agree (S)
40% - 59,99%	Neutral (N)
20% - 39,99%	Disagree (TS)
0% - 19,99%	Strongly Disagree (STS)

III. METHOD

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

A. Literatur 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. The theories collected in this stage are about Django, Gamification, Octalysis Framework, Fisher-Yates Shuffle algorithm, HMSAM, and Likert scale.

B. Application Design

The application design starts from gamification design using the Octalysis method, followed by designing the application model, flowchart, and designing the interface.

C. Application Programming

Application programming is carried out as a form of implementation of the application design that has been created previously. The technology used is the Next.js framework, React library, and Firebase as a database.

D. Application Testing

Application testing is carried out to ensure that the application created can run well according to the design that has been made. Testing will be carried out using the

Microsoft Edge browser. Next, the application will be distributed to a number of respondents who are interested in learning and developing websites using Django for testing. Users who have tried will answer several questions based on the HMSAM model and Likert scale measurements through the online questionnaire that has been distributed. Questionnaires are used to measure the level of behavioral intention to use and immersion of the applications that have been created.

E. Evaluation

Evaluation is carried out by processing data obtained from the results of answers to online questionnaires that have been filled in by users who have carried out experiments on the applications created. Measurement of the level of behavioral intention to use and immersion was carried out based on Likert scale interpretation of online questionnaire answer data.

F. Documentation

Making a research report is carried out after the evaluation results have been analyzed, with the aim of documenting research activities in writing and then displaying the results and conclusions which can be used as a reference for further research.

IV. Results and Discussions

Application implementation is carried out based on application model design, flowcharts, database design, interface design, and asset selection.

G. Implementation

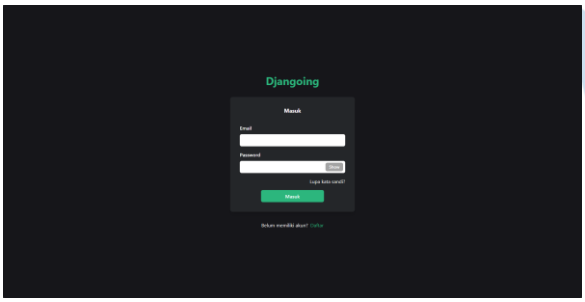


Fig. 1. Login Page

Figure 3 is the result of the implementation of the login page. On the login page, there are 2 columns that must be filled in, namely email and password. If the user has filled in the required data, the user can press the login button so they can move to the home page if the login process is successful. If the user forgets the account password that has been created, the user can press the link. If the user doesn't have an account, there is a registration link that will direct the user to the registration page.

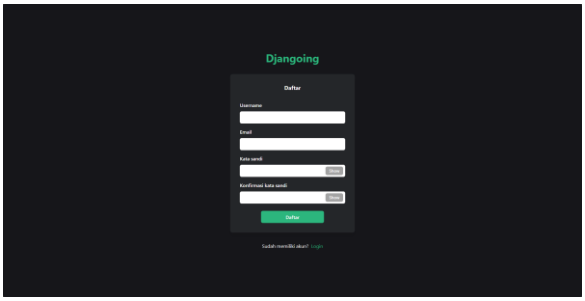


Fig. 1. Halaman Register

Figure 4 is the implementation result of the register page. On the registration page, there are 4 columns that must be filled in, namely username, email, password and also password confirmation. If the user has filled in the required data, the user can press the register button to register a new account if the registration process is successful. If the user already has an account, the user can press the login link and will be directed to the login page.

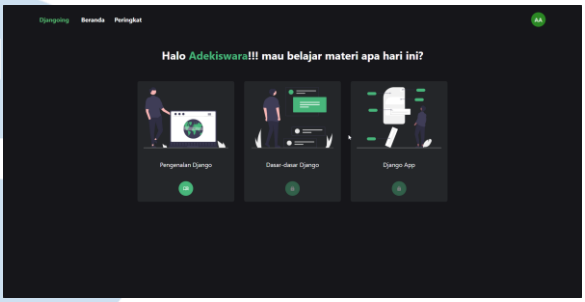


Fig. 2. Home Page

Figure 5 is the implementation result of the home page. On the home page, there is a greeting along with learning modules. Learning modules that are still locked will open if the user has completed or passed the previous module or learning material. If the user presses the button with the book icon, the user will be directed to the material list page.

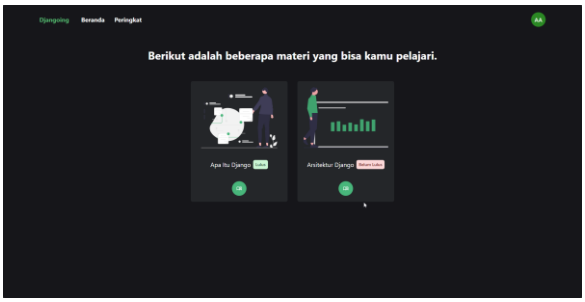


Fig. 3. List Materi Page

Figure 6 is the result of the implementation of the material list page. On the material list page, users can see several materials according to the previously selected module. Learning material will open if the user has completed or passed the previous material. If the user presses the button with the book icon, the user will be directed to the material page.

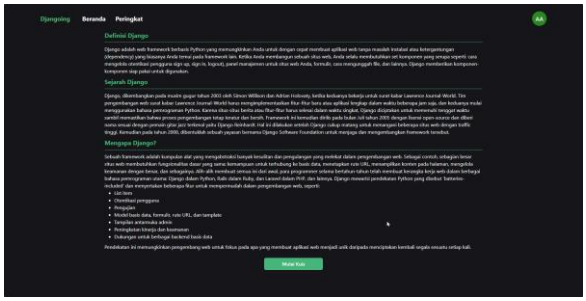


Fig. 4. Lesson Page

Figure 7 is the result of the implementation of the material page. On the material page, users can view and study the material according to the previously selected material title. If the user has finished understanding the material, the user can press the quiz button to be directed to the quiz page.

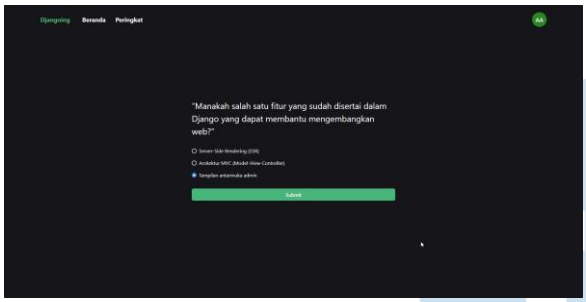


Fig. 5. Quiz Page

Figure 8 is the implementation result of the quiz page. On the quiz page, users can answer questions according to previously selected material. The quiz only consists of one type of answer, namely a choice answer with the contents of 3 questions which will be displayed randomly in each quiz. Users can press the submit button if they feel they have chosen or answered correctly and will be directed to the next question. After reaching the last question, the correct answer will be counted and returned to the home page.

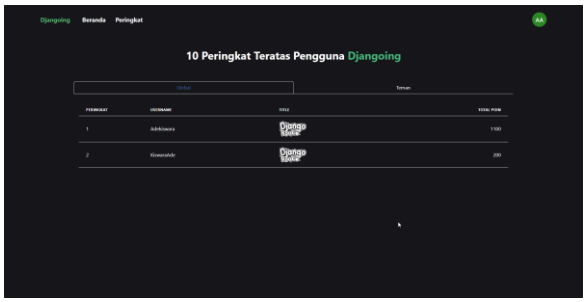


Fig. 6. Leaderboard Page

Figure 9 is the implementation result of the ranking page. On the ranking page, users can see a ranking list of other users with only 10 users. Users can choose to display the ranking list globally or friends. If the user presses another user's name, the user will be redirected to the other user's profile page.

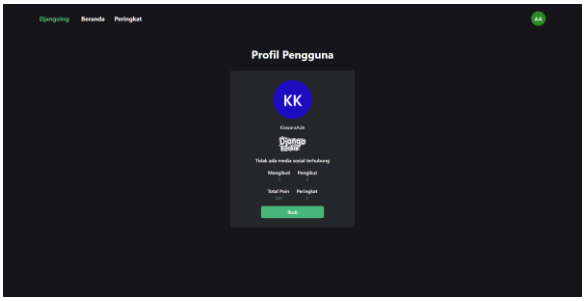


Fig. 7. Other User Profile Page

Figure 10 is the result of the implementation of another user's profile page which can be accessed via the ranking page. Users can press the follow button to add other users as friends.

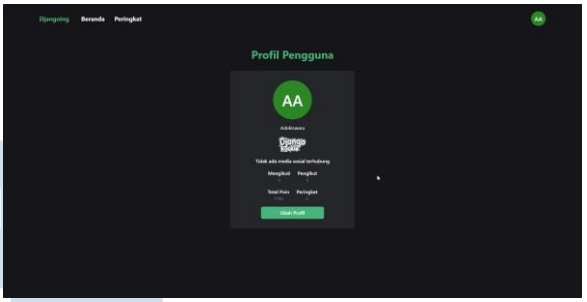


Fig. 9. User Profile Page

Figure 9 is the result of the implementation of the profile page. On the profile page, users can see the avatar, title, number of followers, number of followers, total points and ranking. If the user presses the change profile button, the user will be directed to the edit profile page.

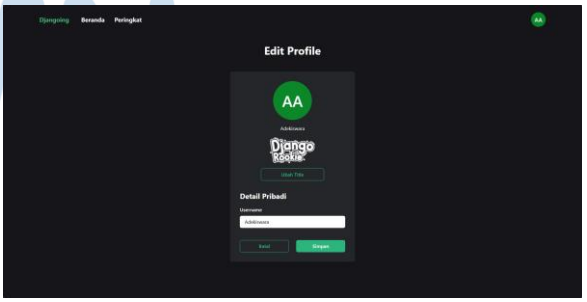


Fig. 10. Edit Profile Page

Figure 10 is the result of the implementation of the edit profile page, the user can change the title via the change title button, apart from that the user can also change the username and it will be saved to the database after pressing the save button.

H. Evaluation

Application trials are carried out by sharing the access link to the application via social media for testing by a number of users who have the criteria, namely that they are studying or working in the IT field and want to learn or get to know the Django framework in basic terms. Users will be asked to fill out an online questionnaire prepared using the HMSAM model to

measure the level of user acceptance of the application based on the level of application behavioral intention to use and immersion.

From the results of previous application testing, the questionnaire calculation results were obtained using a Likert scale in each aspect of HMSAM which can be seen in Table 3.

TABLE 3. EVALUATION OF TEST RESULTS

HMSAM Aspects	*Result*
Perceived Ease of Use	88,93%
Perceived Usefulness	87,39%
Curiosity	85,70%
Control	82,76%
Joy	82,19%
Behavioral Intention to Use	84,15%
Immersion	81,44%

Table 3 it can be seen that the highest calculation result is perceived ease of use, namely 88.93%. These results were obtained because users felt that the Djangoing application was easy to use. Apart from that, the immersion aspect has the lowest calculation at 81.44%. These results were obtained because users felt that the Djangoing application did not give them a feeling of being carried away when learning the Django framework. In the results of calculating all aspects of the HMSAM, the final average percentage was obtained at 84.15% which fell into the Strongly Agree (SS) category.

V. CONSLUSION

Based on the results of research that has been carried out, the Django framework learning application using the gamification method and Fisher-Yates Shuffle has been successfully implemented in the form of a website-based application that has been successfully designed and built. The Djangoing application is used using the NextJS framework, React library, Javascript programming language, and Chakra UI. The game techniques in the Octalysis Framework that have been successfully implemented are Easter eggs, leaderboard, milestone unlock, virtual goods, friending, prize pacing, free lunch and progress loss.

The Djangoing application has been tested by 30 users by filling out a questionnaire prepared using the HMSAM model. The evaluation results of the Djangoing application testing were calculated using a Likert scale with results obtained in the behavioral intention to use aspect of 84.15%, which means that users strongly agree that the Djangoing application creates a desire to use it again in the future, apart from that the results in the immersion aspect were 81 .44% which means users agree that the djangoing application creates a feeling of being carried away when learning the django framework

REFERENCES

[1] B. Hu and J. Zhang, "The application of internet platform in college English teaching," *Applied Mechanics and Materials*, vol. 321–324, pp. 3061–3064, 2013, doi: 10.4028/WWW.SCIENTIFIC.NET/AMM.321-324.3061.

[2] S. K. Lala, A. Kumar, and T. Subbulakshmi, "Secure web development using OWASP guidelines," *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, pp. 323–332, May 2021, doi: 10.1109/ICICCS51141.2021.9432179.

[3] V. Rachapudi, "Performance Comparison of applications with and without Web Frameworks," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 2, pp. 1020–1028, Apr. 2020, doi: 10.30534/IJATCSE/2020/19922020.

[4] S. Qu *et al.*, "Analysis on Web Frameworks," *J Phys Conf Ser*, vol. 1362, no. 1, p. 012114, Nov. 2019, doi: 10.1088/1742-6596/1362/1/012114.

[5] M. Rajeswari and K. R. Kamalam, "DEVELOPING A DJANGO WEB FRAMEWORK," *JETIRFN06036 Journal of Emerging Technologies and Innovative Research*, vol. 9, 2022, Accessed: Nov. 03, 2023. [Online]. Available: http://www.python.org/download/

[6] A. J. Irawan, F. A. T. Tobing, and E. E. Surbakti, "Implementation of Gamification Octalysis Method at Design and Build a React Native Framework Learning Application," *Proceedings of 2021 6th International Conference on New Media Studies, CONMEDIA 2021*, pp. 118–123, 2021, doi: 10.1109/CONMEDIA53104.2021.9617171.

[7] Y. Nurhayati, S. Maesyaroh, S. G. Supartman, E. Darmawan, and E. Herlina, "Implementation of the fisher yates shuffle algorithm in the randomization of department recommendation examinations at PMB FKOM UNIKU," *J Phys Conf Ser*, vol. 1933, no. 1, p. 012008, Jun. 2021, doi: 10.1088/1742-6596/1933/1/012008.

[8] F. Panca Juniawan, H. Arie Pradana, Laurentinus, and D. Yuny Sylfania, "Performance Comparison of Linear Congruent Method and Fisher-Yates Shuffle for Data Randomization," *J Phys Conf Ser*, vol. 1196, no. 1, p. 012035, Mar. 2019, doi: 10.1088/1742-6596/1196/1/012035.

[9] "Django , The Python Web Framework - IJSREM." Accessed: Nov. 03, 2023. [Online]. Available: https://ijsrem.com/download/django-the-python-web-framework/

[10] R. Kumar Singh, H. Gore, A. Singh, and A. Pratap Singh, "Django Web Development Simple & Fast," 2021, Accessed: Nov. 03, 2023. [Online]. Available: www.ijcrt.org

[11] D. Maryono, Budiyono, Sajidan, and M. Akhyar, "Implementation of Gamification in Programming Learning: Literature Review," *International Journal of Information and Education Technology*, vol. 12, no. 12, pp. 1448–1457, Dec. 2022, doi: 10.18178/IJIET.2022.12.12.1771.

[12] "Penggunaan Gamifikasi dalam Proses Pembelajaran Heni Jusuf Perangkingan Usability Website menggunakan Metode Multiple Criteria Decision Analisys Riska Hanifah Pengaruh Adopsi ASTRA Dealer Management System menggunakan

Technology Acceptance Model Terhadap Kinerja Layanan Purna Jual Kendaraan Daihatsu Wildan Wiguna, Dwiza Riana Security Chatting Berbasis Desktop dengan Enkripsi Caesar Cipher Key Random."

[13] A. Andoko, K. Karyono, and E. Ellianto, "Dampak Gamifikasi Tangga Interaktif untuk Mengubah Kebiasaan Manusia Studi Kasus Universitas Multimedia Nusantara," *Ultima Computing : Jurnal Sistem Komputer*, vol. 10, no. 2, pp. 75–80, Mar. 2018, doi: 10.31937/SK.V10I2.1011.

[14] "The Octalysis Framework for Gamification & Behavioral Design." Accessed: Nov. 03, 2023. [Online]. Available: https://yukaichou.com/gamification-examples/octalysis-complete-gamification-framework/

[15] Yusfrizal, D. Adhar, U. Indriani, E. Panggabean, A. Sabir, and H. Kurniawan, "Application of the Fisher-Yates Shuffle Algorithm in the Game Matching the World Monument Picture," *2020 2nd International Conference on Cybernetics and Intelligent System, ICORIS 2020*, Oct. 2020, doi: 10.1109/ICORIS50180.2020.9320766.

[16] I. Maryono, W. B. Zulfikar, and R. Kariadinata, "The implementation of fisher yates shuffle on aljabar learning media based on hybrid application," *MATEC Web of Conferences*, vol. 197, p. 01006, Sep. 2018, doi: 10.1051/MATECCONF/201819701006.

[17] A. Verma, A. Khatana, and S. Chaudhary, "A Comparative Study of Black Box Testing and White Box Testing," *International Journal of Computer Sciences and Engineering*, vol. 5, no. 12, pp. 301–304, Dec. 2017, doi: 10.26438/IJCSE/V5I12.301304.

[18] P. B. Lowry *et al.*, "Taking 'Fun and Games' Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM)," *J Assoc Inf Syst*, vol. 14, no. 11, p. 2, Nov. 2013, doi: 10.17705/1jais.00347.

[19] S. Syofian, T. Setiyaningsih, and N. Syamsiah, "OTOMATISASI METODE PENELITIAN SKALA LIKERT BERBASIS WEB," *Prosiding Semnastek*, vol. 0, no. 0, Nov. 2015, Accessed: Nov. 03, 2023. [Online]. Available: https://jurnal.umj.ac.id/index.php/semnastek/article/view/540

[20] Sugiyono, "Metode Penelitian kuantitatif, kualitatif dan R & D / Sugiyono." Accessed: Nov. 03, 2023. [Online]. Available: https://opac.perpusnas.go.id/DetailOpac.aspx?id=911046#

ISSN 2085-4552

Leveraging Content-Based Filtering for Personalized Game Recommendations: A Flutter-Based Mobile Application Development

Bryan Rezki Nugraha[1], Alexander Waworuntu[2]
[1,2] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]bryan.nugraha@student.umn.ac.id, [2]alex.wawo@umn.ac.id

Accepted 06 January 2025
Approved 16 January 2025

***Abstract*— The background of this study stems from the need for a recommendation system to assist users in finding games that match their interests. With the rapid growth of the gaming market, an increasing number of people engage in gaming activities. In 2022, the personal computer (PC) gaming market accounted for 37.9% of all gamers worldwide. One of the largest PC gaming platforms is Steam, developed by Valve Corporation, which boasts over 184 million active users. However, the overwhelming number of options can lead users to lose interest in purchasing games. Therefore, a recommendation system is required to help users find games that align with their preferences. The methods/theories employed in this study include data from the Steam Web API, SteamSpy API, and local JSON files. The Content-Based Filtering method, using the Cosine Similarity algorithm, was implemented to determine the similarity index between games and user preferences. Flutter was used for application development and to display the recommendation results to users. The results of this study show that the application was successfully developed, and the Content-Based Filtering method provided recommendations that met expectations. The highest cosine similarity factor achieved was 0.6454972244, indicating a fairly good level of accuracy. Application evaluation using the Technology Acceptance Model revealed positive reception, with a "Perceived Usefulness" score of 82.6% and a "Perceived Ease of Use" score of 86.2%, indicating that users found the application both useful and easy to use.**

***Index Terms*— *Terms*—Content-Based Filtering; Cosine Similarity; Flutter; similarity; Steam; SteamSpy API; Steam Web API**.

I. INTRODUCTION

One of the most popular activities in the digital era is gaming. With services like Steam, users can easily access a variety of games. Video games are a form of digital media-based activity where players aim to achieve predetermined objectives within the game [1]. According to Statista, the total revenue from the video game market is projected to reach 625.64 trillion USD by 2028 [2]. Additionally, data from 2022 reveals that the market for personal computer (PC) games is the second largest, following mobile games, accounting for 37.9% of all video game players worldwide [3].

One of the largest marketplaces for PC gaming is Steam, a game distribution platform developed by Valve Corporation, a U.S.-based company. Steam offers over 8,000 available games and boasts more than 184 million active users [4]. However, the abundance of game options presents a significant challenge for users, as too many choices can lead to decision fatigue. Research by Chernev, Böckenholt, and Goodman demonstrates that an excessive number of options can reduce consumer interest in making purchases [5]. This phenomenon highlights the need for an effective recommendation system to assist users in navigating Steam's extensive library and identifying games that align with their preferences.

Steam was chosen as the research object for several compelling reasons. As one of the largest and most influential game distribution platforms globally, its significant user base and vast library of games make it a prime candidate for studying recommendation systems. Additionally, Steam's robust Steamworks Web API and supplementary services like SteamSpy API provide access to valuable data on user activity and game information, enabling the development and testing of advanced algorithms. By addressing the issue of decision fatigue on a platform as prominent as Steam, the findings of this study have practical relevance and the potential to improve user satisfaction and engagement while supporting game developers in reaching their target audience.

Previously, several recommendation systems have been proposed to tackle similar challenges, including those based on Deep Learning [6], the K-Nearest Neighbor (KNN) algorithm [7], and matrix factorization techniques [8]. Each method has its strengths and limitations: Deep Learning achieves high accuracy but requires extensive user data for implementation, the KNN algorithm performs well with existing input, and matrix factorization can

identify personal preferences for approximately 33% of Steam users.

Among the various approaches, Content-Based Filtering has emerged as a widely used method for building recommendation systems. This technique has been successfully applied in other domains, such as movie recommendation systems based on genres [9] and property recommendation systems [10]. In the context of game recommendation systems, Content-Based Filtering utilizes user activity to generate personalized suggestions. A comparative study by the State University of Feira de Santana [11] indicates that this method is particularly effective with large and dense datasets, such as Steam's extensive game library and user base.

For these reasons, this study adopts Content-Based Filtering to develop a recommendation system that mitigates decision fatigue and assists users in finding games that match their preferences. By leveraging this approach, the study aims to address the challenges posed by Steam's vast game selection and improve the overall user experience on the platform.

II. THEORY

A. *Flutter*

Flutter is a cross-platform framework used to develop high-performance mobile applications. It was launched by Google in 2016 [12]. Flutter enables the creation of high-performance applications akin to native apps, thanks to its high-performance rendering engine. In Flutter's architecture, C/C++ code is compiled using the NDK on Android and LLVM on iOS, while Dart code is compiled Ahead Of Time (AOT) [13].

B. *Dart*

Dart is a programming language designed with principles of ease of use, familiarity for most programmers, and scalability. Dart was created to provide tools specifically tailored to meet the needs of modern software and hardware [14]. It is an Object-Oriented Programming (OOP) language developed and maintained by Google. Dart has also been utilized to develop large-scale web applications [15].

C. *Content-Based Filtering*

The Content-Based Filtering (CBF) algorithm is one of the most successful recommendation algorithms, using correlations between content as its foundation. CBF relies on an item's information, represented by attributes, which are compared with other items to calculate similarity [16]. An example of its implementation is a recommendation system that compares a user profile with the content of each document in a collection. The content of a document can be represented by several keywords that reflect the user's profile [17]. One of the methods to calculate similarity between textual data is through Cosine Similarity.

D. *Cosine Similarity*

Cosine Similarity is a commonly used metric to measure the degree of similarity between two vectors, calculated based on the cosine of the angle between them [18]. This method is also useful for measuring the similarity between two documents based on matching terms [19]. The formula for Cosine Similarity is as follows:

$$\text{cosine\_similarity}\ (\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \qquad (1)$$

Explanation:

- a: Represents vector a
- b: Represents vector b
- Cosine Similarity (a,b): The similarity value between vectors a and b, calculated based on the cosine of the angle between them.

E. Technology Acceptance Model

With the growing technological demands in the 1970s and increasing system adoption failures across various organizations, predicting system usability became a popular field among researchers. In 1985, Fred Davis proposed the Technology Acceptance Model (TAM), based on the Theory of Reasoned Action (TRA), to explain individual behavior in adopting technology [20].

The Technology Acceptance Model suggests that Perceived Ease of Use and Perceived Usefulness are significant predictors of application usage, which play a major role in evaluating the effectiveness and usability of a technological system [21].

III. METHOD

A. *Problem Identification*

Before designing and developing the application, the primary step is identifying existing problems. This process was conducted through market research and by gathering information from previous studies on game recommendation systems.

B. *Literature Review*

The literature review involved collecting and analyzing related studies from various written sources, such as journals, articles, and research reports that address similar topics.

C. *Application UI Prototyping*

The initial stage of mobile application development involved creating a user interface (UI) prototype. The prototype was designed using Figma and served as the foundation for the UI of the developed application.

D. Integration of Steamworks API

In this phase, the Steamworks Web API was integrated into the application. This included implementing the login system via Steam using the OpenID service and extracting user data through the API's GET function. The Uri.parse() method was used to retrieve specific user data, which was utilized for the recommendation system.

E. Development of Core Application Functions

At this stage, the core functions designed during the prototyping phase were developed and implemented into the application. These functions were built using the tools and components provided by the Flutter framework.

F. Integration of Content-Based Filtering

The Content-Based Filtering algorithm was implemented to generate personalized recommendations by comparing the similarity between games previously played and other available games on the Steam platform. The tag data from the Steamworks Web API was utilized for this purpose. The integration process followed these steps (Figure 1):

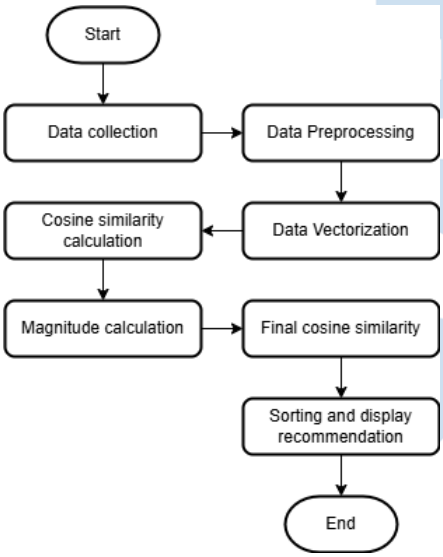


Fig. 1. Content-Based Filtering Process

1) Data Collection: The application retrieved data from Steam Web API, SteamSpy API, and a JSON dataset containing information on 55,000 Steam games.
2) Data Preprocessing: Symbols were removed from the data, and all tags were converted to lowercase for uniformity.
3) Vectorization: The preprocessed tags were vectorized into binary vectors, which were used for cosine similarity calculations.
4) Cosine Similarity Calculation: The dot product was calculated between Vector A: Representing the user profile, containing tags from the games played by the user, and Vector B: Representing tags from a specific game being compared.
5) Magnitude Calculation: The magnitudes of both vectors were calculated by summing the squared values of their components.
6) Final Cosine Similarity: The cosine similarity value was computed by dividing the dot product by the square root of the magnitudes of both vectors.
7) Sorting and Displaying Results: The system iterated through steps 4 to 6 for all games in the dataset, sorted the results based on the cosine similarity values, and displayed the top 10 games with the highest similarity scores.

G. Testing and Debuging

Once development was completed, the application underwent comprehensive testing to ensure its functionality met expectations. Testing focused on evaluating functionality, reliability, performance, and security. Any identified bugs were addressed through debugging processes to ensure smooth application performance.

H. Evaluation using Technology Acceptance Model (TAM)

The developed application was distributed to research participants, who evaluated it using a **survey** based on the Technology Acceptance Model (TAM). The survey included questions addressing two key factors, which are Perceived Ease of Use, and Perceived Usefulness. The TAM framework, proposed by **Fred Davis** and **Richard Bagozzi** [20], was used to measure technology acceptance. The collected results were analyzed to determine the application's acceptance level among users

IV. Results and Discussions

I. Application Interface

Figure 2 shows the application's login interface, where users log in using Steam to extract their gaming data.

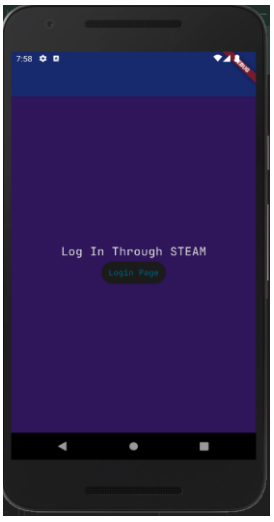


Fig. 2. Login Page Interface

Figure 3 shows the main page of the application, where users can choose to request game recommendations or view their profiles.

Fig. 3. Home Page Interface

Figure 4 illustrates the recommendation page, where users can request recommendations based on games they own or have played in the past two weeks.

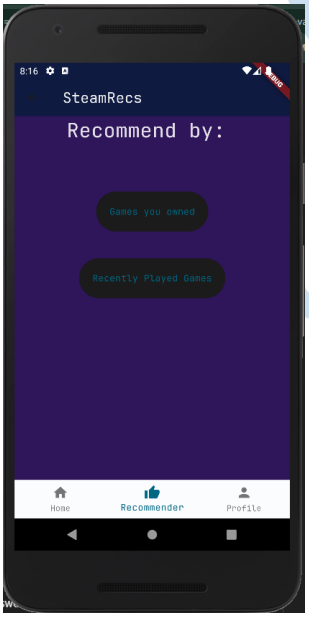


Fig. 4. Recommender Page Interface

Figure 5 displays the results page, where the system shows recommended games along with their **cosine** similarity scores.

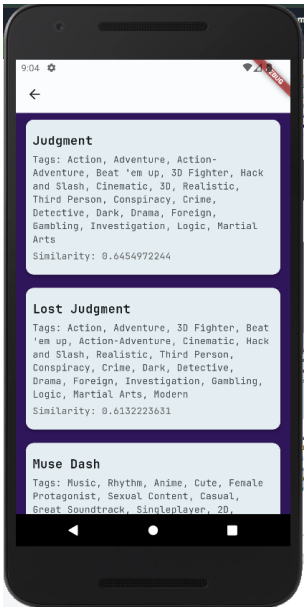


Fig. 5. Recommendation Result Interface

J. *Implementation of Steam Web API*

The Steam Web API was implemented to support the login system and retrieve user data. The login system utilized the WebView plugin, which redirected users to the Steam login page via OpenID. Once the Steam account was linked, users were directed to the main application interface.

The Steam Web API was also used to extract critical user data, including UserID, the user's Steam Library, and general game data from Steam. This data was obtained through the GET function, returning a JSON file.

K. *Data Acquisition*

Before providing game recommendations, necessary data was collected:

1) App ID Data: Retrieved via the Steamworks Web API for detailed game information.
2) Alternative Data Source: Due to limitations of the Steam Web API, the SteamSpy API was used to supplement the required game data.
3) Game Dataset: A JSON file containing data for 55,000 Steam games sourced from Kaggle was utilized to overcome API rate limitations.

L. *Content-Based Filtering Implementation*

After gathering the necessary data, the recommendation process commenced. The first step was pre-processing, which involved filtering games based on playtime, removing symbols and special characters, converting tags to lowercase, and preventing duplicates by converting the data into **Sets**. This ensured uniformity and efficiency in subsequent processes.

Next, the tags were vectorized into binary vectors, enabling the system to perform similarity calculations. The Cosine Similarity metric was then used to measure the similarity between the User Profile Vector (A**)**, representing tags derived from the user's game library, and the Game Vector (B), which contained tags for each game in the dataset.

After calculating the cosine similarity for all games in the dataset, the system sorted the results based on the highest similarity values. The top 10 games with the highest cosine similarity scores were then displayed to the user as recommendations.

The system was tested using the researcher's Steam account, where the highest cosine similarity score of 0.6454972244 was obtained for the game *"Judgment"*. Tags associated with *Judgment*, such as *Action, Adventure, Beat'em Up, Hack and Slash*, closely aligned with the tags in the user profile, demonstrating the accuracy of the recommendation system.

M. Application Evaluation

The application evaluation was conducted using the Technology Acceptance Model (TAM)**.** Upon completion of the application, it was distributed to participants along with a survey created using Google Forms. The survey included questions designed in accordance with TAM principles (Table 1).

TABLE I. SURVEY QUESTIONS BASED ON TECHNOLOGY ACCEPTANCE MODEL (TAM)

No.	Question	Response Scale
1	Name (Initials allowed)	Text input
2	Is the application useful to you?	1 (Not Useful) - 5 (Very Useful)
3	Does the application simplify finding games you want to play?	1 (Strongly Disagree) - 5 (Strongly Agree)
4	Can you easily obtain useful information from the application?	1 (Strongly Disagree) - 5 (Strongly Agree)
5	Does the application help you better understand recommendation systems?	1 (Strongly Disagree) - 5 (Strongly Agree)
6	Does the app interface facilitate your interaction with the recommendation system?	1 (Strongly Disagree) - 5 (Strongly Agree)
7	Does interacting with the app make using recommendation systems easier?	1 (Strongly Disagree) - 5 (Strongly Agree)

Questions 2 to 4 focused on the perceived usefulness of the application, assessing how beneficial the application was for users. Meanwhile, questions 5 to 7 measured the perceived ease of use, evaluating how easy it was for users to interact with and utilize the application. The survey received responses from 30 participants, adhering to the sampling method suggested by Sugiyono [22].

TABLE II. PERCEIVED USEFULNESS

Perceived Usefulness	1	2	3	4	5
Question 2	0	0	3	15	12
Question 3	0	0	8	12	10
Question 4	0	1	6	14	9

The results of the survey are presented in Table II for the "perceived usefulness" aspect and Table III for the "perceived ease of use" aspect. These tables summarize participants' responses on a scale of 1 (Strongly Disagree/Not Useful) to 5 (Strongly Agree/Very Useful).

TABLE III. PERCEIVED EASE OF USE

Perceived Usefulness	1	2	3	4	5
Question 5	0	0	5	11	14
Question 6	0	0	5	12	13
Question 7	0	0	3	13	14

The survey results provided insights into the participants' opinions on the application. To calculate the percentage of perceived usefulness and perceived ease of use, the following formula was applied:

$$\frac{\sum_{i=1}^{p}(x_i \times y_i)}{(p \times s \times j)} \times 100\% \quad (2)$$

Explanation:

- i: Scale value of the question.
- p: Total number of scale points in the question.
- x_i: Specific scale value for a given question.
- y_i: Total number of responses corresponding to the scale value x_i
- s: Total number of survey participants (sample size).
- j: Total number of questions related to the evaluation factor being measured.

Based on the results, Formula 3 was used to calculate the percentage for perceived usefulness, while Formula 4 was applied to compute the percentage for perceived ease of use. The evaluation demonstrated the application's effectiveness and usability from the perspective of the participants.

$$\text{perceived usefulness} = \frac{(0\times1)+(1\times2)+(17\times3)+(41\times4)+(31\times5)}{5\times30\times3} \times 100\% = 82.6\% \quad (3)$$

$$\text{perceived ease of use} = \frac{(0\times1)+(1\times2)+(17\times3)+(41\times4)+(31\times5)}{5\times30\times3} \times 100\% = 86.2\% \quad (4).$$

V. CONSLUSION

The study concluded that the design and development of the application were successfully completed, with the application running smoothly on nearly all devices used during testing. The recommendation system, which employed the Content-Based Filtering method, delivered satisfactory results, achieving the highest Cosine Similarity score of 0.6454972244. Additionally, the evaluation using

the Technology Acceptance Model (TAM) demonstrated positive reception from users, with a perceived usefulness score of 82.6% and a perceived ease of use score of 86.2%**,** indicating that the majority of participants found the application effective and user-friendly.

For future research, it is recommended to explore alternative methods to access dynamic data without being constrained by API call limitations, enabling the use of more accurate and comprehensive data from Steam services. Additionally, the recommendation process could be expedited by optimizing the algorithm further or adopting alternative methods that can deliver equivalent or superior results with reduced processing time.

REFERENCES

[1] R. Ramadan and Y. Widyani, "Game development life cycle guidelines," in *2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE, 2013, pp. 95–100.

[2] J. Clement, "Video game market revenue worldwide from 2018 to 2028," Dec. 2023. [Online]. Available: https://www.statista.com/statistics/1344668/revenue-video-game-worldwide/

[3] ——, "Leading devices used to play games worldwide 2022," Jan. 2023. [Online]. Available: https://www.statista.com/statistics/533047/leading-devices-play-games/

[4] G. Sergey, "All the data and stats about steam games," 2016. [Online]. Available: http://steamspy.com/

[5] A. Chernev, U. Bockenholt, and J. Goodman, "Choice overload: A conceptual review and meta-analysis," *Journal of Consumer Psychology*, vol. 25, no. 2, pp. 333–358, 2015.

[6] D. Wang, M. Moh, and T.-S. Moh, "Using deep learning and steam user data for better video game recommendations: Proceedings of the 2020 ACM southeast conference," Apr. 2020. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3374135.3385283

[7] G. Cheuque, I. P. U. Catolica Santiago, J. Guzmán, D. Parra, G. Tech, and O. Metrics, "Recommender systems for online video game platforms: The case of steam: Companion proceedings of the 2019 world wide web conference," May 2019. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3308560.3316457

[8] S. Bayram, "Game recommendation system for steam platform," Jan. 1970. [Online]. Available: https://openaccess.mef.edu.tr/xmlui/handle/20.500.11779/1721

[9] S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," Jan. 1970. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-13-1927-3_42

[10] T. Badriyah, S. Azvy, W. Yuwono, and I. Syarif, "Recommendation system for property search using content-based filtering method," Mar. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8350801

[11] R. Glauber and A. Loula, "Collaborative filtering vs. content-based filtering: Differences and similarities," Dec. 2019. [Online]. Available: https://arxiv.org/abs/1912.08932

[12] A. Tashildar, N. Shah, R. Gala, T. Giri, and P. Chavhan, "Application development using flutter," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 2, no. 8, pp. 1262–1266, 2020.

[13] W. Wu, "React native vs flutter, cross-platforms mobile application frameworks," 2018.

[14] A. Hassan, "Java and dart programming languages: Conceptual comparison," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 2, pp. 845–849, 2020.

[15] G. I. Arb and K. Al-Majdi, "A freights status management system based on dart and flutter programming language," in *Journal of Physics: Conference Series*, vol. 1530, no. 1. IOP Publishing, 2020, p. 012020.

[16] M. Al-Shamri, B. Amiri, A. Biswas, W. Carrer-Neto, F. Colace, S. Choi, M. Everett, L. Freeman, O. Kwon, Y. Li, et al., "Content-based filtering for recommendation systems using multiattribute networks," Aug. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0957417417305468

[17] R. van Meteren and M. van Someren, "Using content-based filtering for ... - ics-forth." [Online]. Available: http://users.ics.forth.gr/~potamias/mlnia/paper_6.pdf

[18] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," Sep. 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7577578/

[19] D. Gunawan, C. A. Sembiring, and M. A. Budiman, "The implementation of cosine similarity to calculate text relevance between two documents," Mar. 2018. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/978/1/012120/meta

[20] M. Chuttur, "Overview of the technology acceptance model: Origins, developments and future directions," Jun. 2013. [Online]. Available: https://aisel.aisnet.org/sprouts_all/290

[21] M. Masrom, "Technology acceptance model and E-learning," May 2007. [Online]. Available: https://www.researchgate.net/publication/228851659_Technology_acceptance_model_and_E-learning

[22] Sugiyono, *Metode penelitian pendidikan: pendekatan kuantitatif, kualitatif, dan RD*, revised edition ed. Bandung: Alfabeta, 2016.

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.
- For December 2021 publication onwards, Ultimatics : Jurnal Teknik Informatika will be receiving and publishing manuscripts written in English only.

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: **ultimatics@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 ULTIMATICS 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 ULTIMATICS. 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

ULTIMATICS_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		

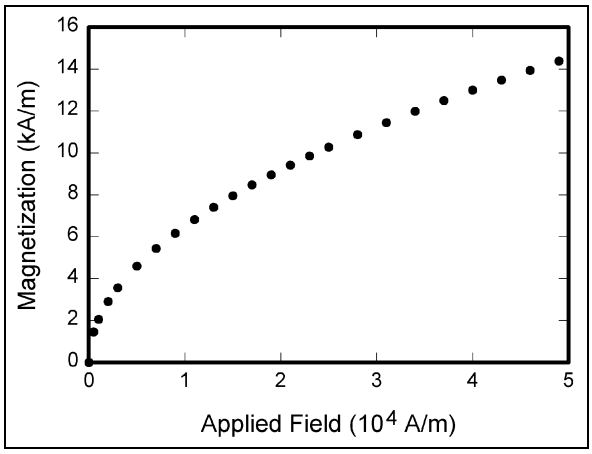


Fig. 1. Example of a figure caption

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 2085-4552

Universitas Multimedia Nusantara
Scientia Garden Jl. Boulevard Gading Serpong, Tangerang
Telp. (021) 5422 0808 | Fax. (021) 5422 0800