POLITEKNIK NEGERI BALI

matriX

JURNAL MANAJEMEN TEKNOLOGI DAN INFORMATIKA

Vol. 11 No. 2 | Page 55 - 118 | July 2021 | ISSN 2580-5630

Editors

Reviewers

PREFACE

We would like to present, with great pleasure, the second issue of Matrix: Jurnal Manajemen Teknologi dan Informatika in Volume 11, No. 2, 2021. This journal is under the management of Scientific Publication, Research and Community Service Center, Politeknik Negeri Bali and is devoted to cover the field of technology and informatics management including managing the rapid changes in information technology, emerging advances in electrical and electronics and new applications, implications of digital convergence and growth of electronics technology, and project management in electrical, mechanical or civil engineering. The scientific articles published in this edition were written by researchers from Universitas Bumigora, Universitas Pendidikan Mandalika, STIKI Malang, Politeknik Negeri Bandung, STMIK STIKOM Indonesia, Universitas Dhyana Pura, STMIK Primakara, Universitas Muria Kudus, dan Universitas Teknologi Yogyakarta. All articles cover topics in the field of Information Management, including KAMI Index as an Evaluation of Academic Information System Security at XYZ University, Designing a Word Recommendation Application Using the Levenshtein Distance Algorithm, Exploratory Data Analysis of Crime Report, DSS for Best E-Commerce Selection Using AHP-WASPAS and AHP-MOORA Methods, The System Development Life Cycle Model Implementation on Information System of Performance Reporting IT Asset Case Study: PT Kereta Api Indonesia (Persero), and Comparative Analysis of Support Vector Machine and K-Nearest Neighbors with a Pyramidal Histogram of the Gradient for Sign Language Detection. Finally, we would like to thank reviewers for their efforts and hard work in conducting series of review phase thoroughly based on their expertise. It is our hope that the work of the authors in this issue will be a valuable resource for other researchers and will stimulate further research into the vibrant area of technology and information management in specific, and engineering in general.

Politeknik Negeri Bali, 15 July 2021

Editor-in-chief

Gusti Nyoman Ayu Sukerti, S.S., M.Hum.

ISSN: 2580-5630

Table of contents

Matrix: Jurnal Manajemen Teknologi dan Informatika
Volume 11 Issue 2 Year 2021 Pages 55-62
DOI: http://dx.doi.org/10.31940/matrix.v11i2.2447
URL: http://ojs.pnb.ac.id/index.php/matrix

KAMI index as an evaluation of academic information system security at XYZ university

Andi Sofyan Anas [1]*, I Gusti Ayu Sri Devi Gayatri Utami [2], Adam Bachtiar Maulachela [3], Akbar Juliansyah [4]

[1,2] *universitas Bumigora, Indonesia*
[3,4] *Universitas Pendidikan Mandalika , Indonesia*

*Corresponding Author: *andi.sofyan@universitasbumigora.ac.id*

Abstract: XYZ University is one of the universities that has used information technology to create quality service for students and the entire academic community. This Information technology service is managed by Information Technology and Communication Center (PUSTIK) which is responsible to carry out the development, management, service, and maintaining the security of information and communication technology. Good information technology governance should be able to maintain information security. Therefore, it is necessary to evaluate information system security especially the security of academic information systems. This information system security evaluation uses *Keamanan Informasi* (KAMI) Index which refers to the ISO/IEC 27001:2013 standard to be able to determine the maturity level of information security. An evaluation of five areas of the KAMI Index shows the Information Security Risk Management area gets the lowest score at 10 out of a total of 72. The result of the KAMI Index dashboard shows that the maturity level of each area of information security is at levels I and I+ with a total score of 166. This means that the level of completeness of implement ISO 27001:2013 standard is in the inadequate category.

Keywords: KAMI index, ISO/IEC 27001:2013, evaluation, security, information

History Article: Submitted 27 March 2021 | Revised 26 April 2021 | Accepted 17 May 2021

How to Cite: A. S. Anas, I. G. A. S. D. G Utami, A. B. Maulachela, A. Juliansyah, "KAMI index as an evaluation of academic information system security at XYZ university," *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 11, no. 2, pp. 55-63, 2021.

Introduction

Until now, an organization still relies on information to help effectiveness and create quality services [1]. The implementation of Information Technology governance has now become a necessity and also a note in agencies, given the role of Information Technology which is currently increasingly important as a realization of good organizational or corporate governance [2]. Information technology governance should be able to maintain information security. According to Tata Sutabri in Edo Rizky Pratama, et al., Information is the most valuable data in the decision-making process [3]. Maintaining information security means protecting all information assets owned from threats that may arise, by taking into account the security factors of all supporting devices, networks, and other facilities that are directly or indirectly related to the information processing [4].

Educational institutions in Indonesia need to implement an information security system to safeguard their data to ensure security and authenticity. Information security is an effort to protect information assets from threats that may arise. The risk of data damage, loss, and exposure to unwanted parties is directly proportional to the increasing number of information stored, managed, and shared [5]. According to Husaini et al. In Prastiyawan et al. defines Risk as the probability of an event that can harm the company due to vulnerabilities and threats [6]. XYZ University is one of the educational institutions that has implemented an academic information system to provide information to students and the academic community. This Academic Information System is managed and developed by the Center for Information and Communication

Technology (PUSTIK). PUSTIK is responsible for carrying out the development, management, and services of information and communication technology. Information system security must be managed since the information system was built, not as a complement to the information systems [7].

Information security in the XYZ University academic information system is very necessary because it involves the data of all students, lecturers, and employees. Data that is not maintained, data integrity that cannot be maintained, affect the effectiveness and efficiency in providing and offering information to the academic community and disrupting the institution in achieving its institutional goals and strategies [8].

Because information system security is so important, a good policy with procedures is needed, including Asset management, human resource management, physical and environmental safeguards, logical security, information technology operational security, and incident handling in information security. Information system security evaluation must apply information system security audit techniques to ensure information system security meets standards and is following procedures. Evaluation is the process of evaluating objects by first taking measurements [9].

To be able to revise and improve the quality of information security of an institution, the Ministry of Communication and Information makes efforts one of which is to create an Information Security Index / Indeks Keamanan Informasi (KAMI) which is a tool to measure the level of maturity and completeness in information security. The KAMI index refers to the information security standard, namely ISO 27001 [10]. The KAMI index is not used to analyze the feasibility or effectiveness of information security, but as a tool that provides an overview of the readiness of an information security framework to leaders [11].

The specifications and requirements that must be met in building an Information Security Management System (ISMS) are regulated in ISO / IEC 27001. ISO 27001 is a standard that describes the management of information security in an organization. An overview of the needs of an organization in its efforts to implement information security concepts can be provided by ISO 27001 [12]. This standard is independent of information technology products, requires the use of a risk-based management approach, and is designed to ensure that selected security controls can protect information assets from various risks and assure the level of security for interested parties [10].

Based on the description that has been described, a study on the Evaluation of Academic Information Security Systems at XYZ University was carried out using the KAMI Index tool by referring to the ISO / IEC 27001: 2013 standard to be able to determine the maturity level of academic information system security.

Methodology

The stages carried out in this study are following this chart:

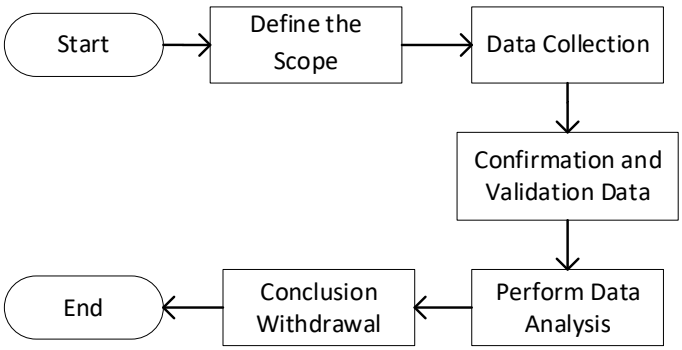


Figure 1. Research stages

Define the Scope

The research begins with defining the scope. XYZ University certainly understands the importance of information system security for campus success. Institutions or companies must pay attention to information system security because it can cause harm to the institution in the

event of leakage and system failure [13]. The confidentiality and authenticity of the data and information processed will be maintained because of the good security of the information system. Therefore, it is necessary to evaluate information security to determine the maturity level of academic information system security at XYZ University using the KAMI index by referring to the ISO / IEC 27001: 2013 standard.

Data Collection

At the stage of data collection obtained from direct observation, filling out questionnaires, and interviews with competent parties on the object to be studied. Observations were made to determine the conditions of the existing security management for the system. So that later the results of the observations are used in determining the appropriate objective control. Furthermore, to obtain primary data, interviews were conducted with several managers related to academic information systems. Interviewed managers consisted of PUSTIK, Academic Administration (BAAK) and XYZ University Vice-Chancellor II (WR II).

Confirmation and Validation Data

Data confirmation and validation were carried out to check the authenticity and validity of the data obtained from the informants. This stage is carried out by the checklist method [14]. The checklist was carried out by respondents, namely the PUSTIK, BAAK, and WR II sections. This data validation was carried out concerning the five areas of the Index.

Perform Data Analysis

The next stage is data analysis to obtain evaluation results on the level of completeness of the application of the ISO 27001: 2013 standard. The data were analyzed using the existing formulation in the Information Security Index (KAMI Index) against the previously distributed questionnaires, so that the level of maturity and completeness of information security was obtained, which was then adjusted to ISO 27001: 2013 standards. Annex A or security control in the ISO / IEC 27001: 2013 structure consists of 14 domain areas, 35 objective controls, and 114 information security controls [10].

Areas in the KAMI Index, which are used to measure the maturity level of the ISMS at an institution, summarize the 14 domain areas in the ISO 27001 structure into 5 evaluation areas [3]. The relationship between the KAMI index and ISO 27001 can be seen in Figure 2 below.

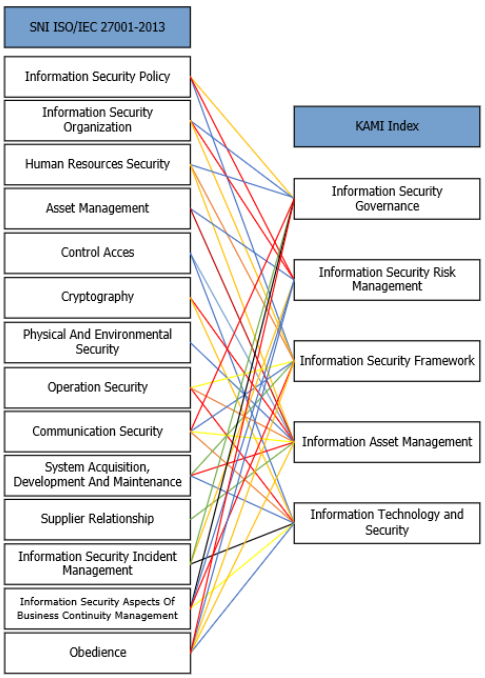


Figure 2. KAMI Index relationship with ISO 27001 [3]

Conclusion

The last stage is to conclude the results of the data analysis that has been carried out. With this conclusion, we can find out the maturity level and feasibility of academic information system security at XYZ University.

Results and Discussions

The evaluation process is carried out by answering several questions from the following areas [10]:

1. Information System Security Governance
2. Information Security Risk Management
3. Information Security Framework
4. Information Asset Management
5. Information Technology and Security.

The answer to each question is scored to generate an index score and is also used to display the evaluation results in the dashboard at the end of the process. The score for each question refers to the following Table 1.

Table 1. KAMI index score mapping

	Security Category		
Security Status	1	2	3
Is Not Done	0	0	0
In planning	1	2	3
In an application or partially applied	2	4	6
Applied thoroughly	3	6	9

Before answering each question from the five evaluation areas on the KAMI index, first, the Electronic Systems category classification is carried out. Respondents must briefly describe the Electronic System in their work unit. It aims to classify the electronic systems used into certain levels, namely low, high and strategic [10]. The correlation between the Electronic System Category and the readiness status can be seen in Figure 3 below:

Electronis System Category				
Low		**Final Score**		**Readiness Status**
10	15	0	174	Not feasible
		175	312	Needs Improvement
		313	535	Enough
		536	645	Good
High		**Final Score**		**Readiness Status**
16	34	0	272	Not feasible
		273	455	Needs Improvement
		456	583	Enough
		584	645	Good
Strategic		**Final Score**		**Readiness Status**
35	50	0	333	Not feasible
		334	535	Needs Improvement
		536	609	Enough
		610	645	Good

Figure 3. Correlation matrix of electronic system categories and readiness status

The results of the assessment for the Electronic Systems category level at XYZ University can be seen in Figure 4 below:

PART 1 : ELECTRONIC SYSTEM CATEGORY			
This section evaluates the level or category of Electronic Systems used			
[Electronic System Category] Low; High; Strategic		Status	Score
#	Characteristics of Agencies/Companies		
1.1	Investment Value of Installed Electronic Systems [A] More than Rp. 30 Billion [B] More than Rp. 3 Billion until Rp. 30 Billion [C] Less than Rp. 3 Billion	C	1
1.2	The total annual budget allocated for Electronic System management [A] More than Rp. 10 Billion [B] More than Rp. 1 Billion until Rp. 10 Billion [C] Less than Rp. 1 Billion	C	1
1.3	Have an obligation to comply with certain Regulations or Standards [A] National and International Regulations or Standards [B] National Regulations or Standards [C] There are no special Regulations	C	1
1.4	Using special cryptographic techniques for information security in Electronic Systems [A] Specific cryptographic techniques certified by the state [B] industry-standard, publicly available or self-developed cryptographic techniques [C] There is no use of cryptographic techniques	C	1
1.5	Number of Electronic System users [A] More than 5.000 users [B] 1000 until 5000 users [C] Less than 1000 users	C	1
1.6	Personal data managed by Electronic Systems [A] Personal data that has a relationship with other personal data [B] Individual personal data and / or personal data related to the ownership of a business entity [C] There is no personal data	C	1

Figure 4. Electronic system readiness status

From the results of the XYZ University Electronic System Category assessment, it was obtained a score of 10, so it was in the Low category because it was in the range of values 10-15. This low category means the importance of using Electronic Systems at XYZ University has not become a priority and there is still a lack of awareness about the importance of using Electronic Systems.

The next stage is to evaluate 5 areas on the KAMI index. Respondents did a checklist to confirm data by comparing the results of the questionnaire with the actual situation. From this checklist, the results of the level of completeness and security of information are obtained as shown in Figure 5 on the KAMI Index dashboard.

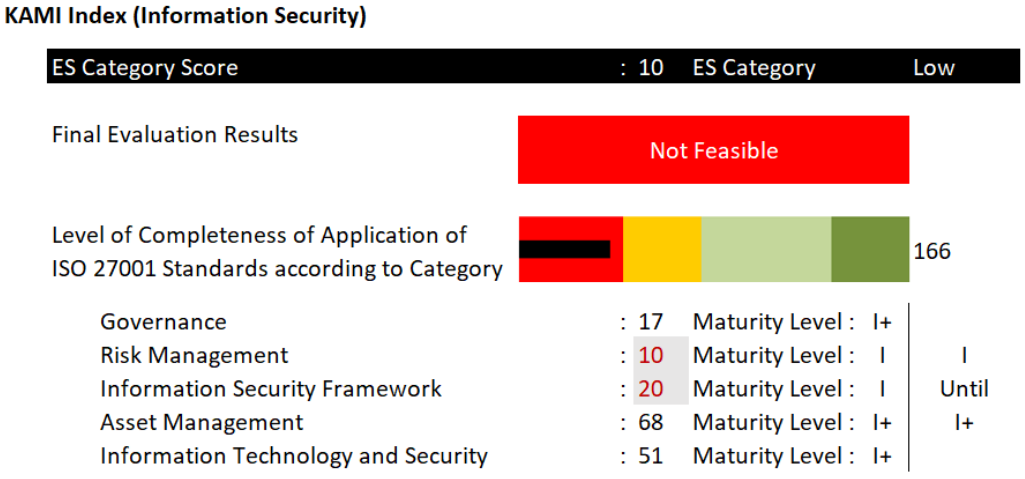


Figure 5. Information security maturity level

Based on Figure 5, it can be stated that the level of completeness of the application of the ISO 27001 Standard according to the electronic category gets a score of 166 with maturity levels at levels I to I +, and is in the red area which means the status is not feasible.

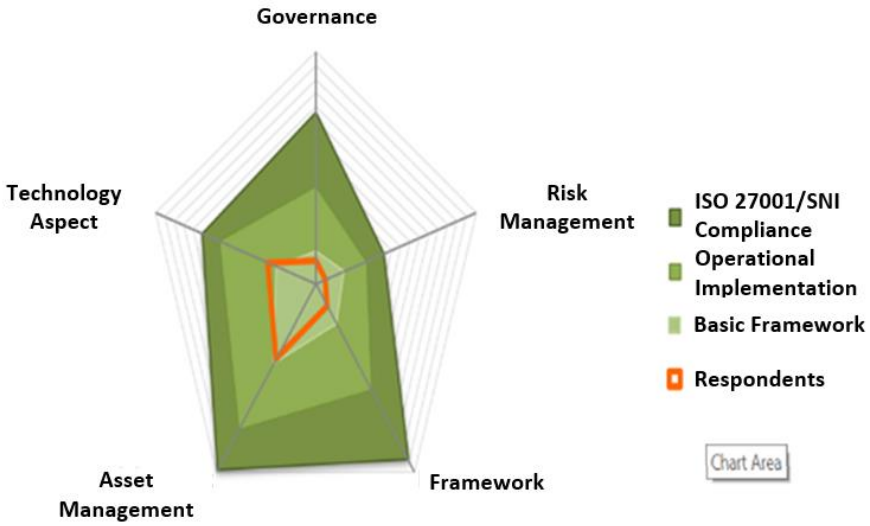


Figure 6. Radar diagram Information security maturity level

Based on the results of the radar diagram in Figure 6, the data is obtained from the calculation of the KAMI Index, showing the Red section or Red area (Respondents) on the bar chart is the security condition of the XYZ University academic information system. From the five areas of information security, it can be observed that the academic information system of XYZ University has a technological aspect that is in the operational implementation process area. Meanwhile, governance, risk management, framework, and asset management are in the basic framework area in the information system security management process.

The following is a description of the percentage maturity level of the five areas previously assessed using the KAMI Index Version 4.0:

Table 2. Percentage of information maturity level

Annotation	Governance	Risk Management	Framework	Asset Management	Security Technology
Max Score	126	72	159	168	120
Respondents	17	10	20	68	51
Percentage	13%	14%	12%	40%	42%

Based on Table 2 it can be described as follows:

1. In the Information Security Governance area, the respondent's score was 17 (13%) from a maximum score of 126. This score was obtained from 13 scores representing maturity level II, 4 scores representing maturity level III, and 0 scores representing maturity level IV. The Information Security Governance Area is classified into the status level of maturity level I + with a score of 13. Because it has exceeded the minimum maturity level, which is 12 but does not exceed the requirements to reach the minimum value of maturity level II, namely 36. The status of the maturity level of the security governance area This information relates to domain control areas A5 (Information Security), A7 (Human resource security) at ISO 27001: 2013.
2. In the Information Security Risk Management area, the respondent's score is 10 (14%) of the maximum score of 72, the score is 8 scores representing maturity level II, 2 scores representing maturity level III, 0 scores representing maturity level IV, and 0 The score represents the level of maturity V. The Information Security Risk Management Area is classified into the status level of maturity level I with a score of 8. Therefore it has not

exceeded the minimum maturity level of 14 and also does not exceed the requirements to reach a minimum value of maturity level II, namely 20. The relationship between the status of the maturity level of the information security risk management area with ISO 27001: 2013 is in the domain control areas A5 (Information Security) and A8 (Asset Management).

3. In the Information Security Management Framework area, it was found that respondents were 20 (12%) from a maximum score of 159, obtained from 10 scores representing maturity level II, 10 scores representing maturity level III, 0 scores representing maturity level IV, and 0 scores represent the level of maturity V. The Information Security Management Framework Area is classified into the status level of maturity level I with a score of 10. Therefore it has not exceeded the minimum maturity level of 15 and also does not exceed the requirements to reach a minimum value of maturity level II, namely 24. The relationship between the status of this area's maturity level with ISO 27001: 2013 is in the domain control areas A5 (Information Security), A11 (Physical and environmental security), and A12 (Operational security).
4. In the area of Information Asset Management, it was found that respondents were 68 (40%) from a maximum score of 168, where 50 scores represented maturity level II, 18 scores represented maturity level III. The Information Asset Management Area is classified into the status level of maturity level I + with a score of 50. Because it has exceeded the minimum maturity level, which is 25, it does not exceed the requirement to reach the minimum value of maturity level II, which is 62. The relationship between the status of the maturity level of the information asset management area with ISO 27001: 2013 there are domain control areas A7 (Human resource security), A8 (Asset Management), A11 (Physical and environmental security), and A12 (operational security).
5. In the area of Information Technology and Security, it was found that respondents were 51 (42%) from a maximum score of 120, obtained from 23 scores representing maturity level II, 28 scores representing maturity level III, 0 scores representing maturity level IV. The area or part of Information Asset Management is classified into the status level of maturity level I + with a score of 23. Because it has exceeded the minimum maturity level, which is 18 but does not exceed the requirements to reach the minimum value of maturity level II, namely 28. The relationship status of the maturity level of the technology area and information security with ISO 27001: 2013 is in the domain control areas A9 (access control) and A12 (operation security).

Conclusion

The conclusions that can be generated from this research on the evaluation of academic information system security using the KAMI Index and ISO 27001: 2013 are:

1. The results of the Electronic Systems category assessment at XYZ University get a score of 10 which means it is in a low category. This indicates that there is still a lack of interest in using electronic systems and low awareness of electronic systems.
2. The level of security and completeness of the application of the ISO 27001: 2013 standard according to the electronic category gets a score of 166, which means it is in the inappropriate category and at the maturity level I to I +. The Risk Management Area, Information Security Framework Area is at maturity level I, while the Governance, Asset Management, and Information Technology and Security areas are at maturity level I +. The cause of the low level of completeness and maturity of information security at the XYZ University Academic Information System has not implemented all security requirements or is still in planning.
3. Evaluation of the five areas of the KAMI Index shows that the Information Security Risk Management area gets the lowest score, 10 out of a total of 72, and is at maturity level I.
4. The results of the evaluation of the five KAMI index areas show a relationship with ISO 27001: 2013 in the domain areas A5, A7, A8, A9, A11, and A12.

Advice is given to researchers who will conduct information system security research using the KAMI index to provide recommendations based on ISO 27001: 2013. Recommendations are given by looking at what deficiencies exist in each area of the KAMI Index and comparing them with ISO 27001: 2013 controls.

Acknowledgments

With humility, the author would like to express his gratitude and appreciation to the Leaders and Academic Community of XYZ University, especially for the Vice-Chancellor II, the Head of BAAK, and the Head of PUSTIK, because for their help this research can be completed. And thank you for your cooperation during the author's research and data collection.

References

[1] D. Riani, *Audit Keamanan Sistem Informasi Akademik (SIAKAD) Universitas Lampung Menggunakan Standar ISO/IEC 27001*. Lampung: Universitas Lampung, 2018.

[2] M. Ikhsan, E. Darwiyanto, and D. D. J. Suwawi, "Audit keamanan sistem informasi akademik Sekolah Tinggi Farmasi Bandung berbasis risiko dengan menggunakan standar ISO 27001," *e-Proceeding Eng.*, vol. 3, no. 3, p. 5222, 2016.

[3] E. R. Pratama, Suprapto, and A. R. Perdanakusuma, "Evaluasi tata kelola sistem keamanan teknologi informasi menggunakan indeks KAMI dan ISO 27001: Studi kasus KOMINFO Provinsi Jawa Timur," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5911–5920, 2018.

[4] F. A. Basyarahil, "Evaluasi manajemen keamanan informasi menggunakan indeks Keamanan Informasi (KAMI) berdasarkan ISO/IEC 2700:2013 pada Direktorat Pengembangan Teknologi Dan Sistem Informasi (DPTSI) ITS Surabaya," Jurnal Tekniks, vol. 6, no. 1, 2017.

[5] R. Sarno and I. Iffano, *Sistem Manajemen Keamanan Informasi*. Surabaya: ITS Press, 2009.

[6] D. A. Prastiyawan, A. Ambarwati, and E. Setiawan, "Analisis manajemen risiko layanan sistem manajemen dealer menggunakan COBIT 5," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 10, no. 2, pp. 43–49, 2020.

[7] C. Chazar, "Standar manajemen keamanan informasi berbasis ISO/IEC 27001: 2005," *J. Inf.*, vol. VII, no. 2, pp. 48–57, 2015.

[8] D. Fitrianah and Y. G. Sucahyo, "Audit sistem informasi/ teknologi informasi dengan kerangka kerja cobit untuk evaluasi manajemen teknologi informasi di Universitas XYZ," *J. Sist. Inf.*, vol. 4, no. 1, p. 37, 2012.

[9] S. Arikunto, *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara, 2015.

[10] KOMINFO, Panduan Penerapan SMKI Berbasis Indeks KAMI, no. September, 2017. Jakarta: KOMINFO.

[11] Standar Nasional Indonesia and Badan Standardisasi Nasional, "Teknologi informasi – Teknik keamanan – Sistem manajemen keamanan informasi – Persyaratan," 2009.

[12] BSN, *Information technology - Security techniques - Information security management systems - Requirements (ISO/IEC 27001:2005, IDT)*. Jakarta: BSN, 2009.

[13] Mufadhol, "Kerahasiaan dan keutuhan keamanan data dalam menjaga integritas dan keberadaan informasi data," *J. Transform.*, vol. 6, 2009.

[14] N. Arman, W. Hayuhardhika, and A. Rachmadi, "Evaluasi keamanan informasi pada dinas komunikasi dan informatika Kabupaten Sidoarjo menggunakan indeks Keamanan Informasi (KAMI)," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, pp. 5750–5755, 2019.

Matrix: Jurnal Manajemen Teknologi dan Informatika
Volume 11 Issue 2 Year 2021 Pages 63-70
DOI: http://dx.doi.org/10.31940/matrix.v11i2.2419
URL: http://ojs.pnb.ac.id/index.php/matrix

Designing a word recommendation application using the Levenshtein Distance algorithm

Nadhia Nurin Syarafina [1]*, Jozua Ferjanus Palandi [2], Nira Radita [3]

[1,2,3] *Teknik Informatika STIKI Malang, Indonesia*

*Corresponding Author: *nurinnadhia@gmail.com*

Abstract: Good scriptwriting or reporting requires a high level of accuracy. The basic problem is that the level of accuracy of the authors is not the same. The low level of accuracy allows for mistyping of words in a sentence. Typing errors caused the word to become non-standard. Even worse, the word became meaningless. In this case, the recommendation application serves to provide word-writing recommendations in case of a typing error. This application can reduce the error rate of the writer when typing. One method to improve word spelling is Approximate String Matching. This method applies an approach to the string search process. The Levenshtein Distance algorithm is a part of the Approximate String-Matching method. This method, firstly, is necessary to go through the preprocessing stage to correct an incorrectly written word using the Levenshtein Distance algorithm. The application testing phase uses ten texts composed of 100 words, ten texts composed of 100 to 250 words, and ten texts composed of 250 to 500 words. The average accuracy rate of these test results was 95%, 94%, and 90%.

Keywords: levenshtein distance, approximate string matching, preprocessing, recommendation, "PUEBI"

History Article: Submitted 9 March 2021 | Revised 1 May 2021 | Accepted 21 May 2021

How to Cite: N.N.Syarafina, J.F. Palandi, and N. Radita, "Designing a word recommendation application using the Levenshtein Distance algorithm," *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 11, no. 1, pp. 64–71, 2021.

Introduction

Typing a report manuscript or an essay requires high accuracy. The level of accuracy of a writer has different levels. A low level of accuracy can result in errors in typing a term or word, which makes a term or a phrase non-standard, and may even have no meaning. Also, the factor that causes typing errors is a habit of a writer in abbreviating a word. This typing error problem often occurs when a writer writes scientific papers, proposals, or reports for school, college, or work needs. This kind of error is fatal if not corrected immediately. The wrong word can have a different meaning, or it can even be a word with a different connotation in a sentence. Spelling errors and writing non-standard terms can change the true meaning of information and lead to readers' misinterpretation [1]. We can avoid these typing errors by using an Approximate String Matching method, which is a method for matching a string based on similarity in terms of writing, both the number of characters and the composition of characters in a document [2]. This method has several types of algorithms, one of which is the Levenshtein Distance algorithm. The Levenshtein Distance algorithm has three types of string operations, namely deleting, adding, and changing. These operations are useful for calculating the distance between 2 strings. The smaller the distance between the two strings indicates that the two strings are match [3]. Fixing this problem can use the Levenshtein Distance algorithm with a high level of accuracy in terms of similarity search words. An application to recommend the right words in a text using the Levenshtein Distance algorithm is the right solution to solve this problem. This application helps minimize word writing errors, both words that are not standard or words that have no meaning.

The Indonesian language has developed very rapidly as a result of advances in science, technology, and art. The use of the Indonesian language is increasingly widespread, both verbally and in writing. Therefore, writing a manuscript requires a reference or a guideline in writing an Indonesian

language script. In this case, Language Development and Development Board has made improvements to the Indonesian spelling. The refinement resulted in a document called "Pedoman Umum Ejaan Bahasa Indonesia (PUEBI)" [4].

A spelling checker is a feature that checks words or spelling errors based on a specific language. The spelling checker looks for all kinds of mistakes in the document, then warns the user regarding the existing error and provides some suggestions to fix it [5]. Furthermore, spell checker features include a) proofreading and parsing of all words in the manuscript, b) comparing parsing results with a dictionary containing a list of correct word spelling, c) handling errors by utilizing appropriate matching techniques [6].

In the field of image processing, preprocessing means the initial process for removing noise [7]. However, the preprocessing function in this program is to get keywords that serve to match a string or to compare a document [8]. In general, the preprocessing stage consists of four steps: case folding, tokenizing, filtering, and stemming [9]. In this study, the filtering stage was eliminated so that the preprocessing process only consisted of tokenizing, case folding, and stemming. Tokenizing is generally replacing sensitive data with a unique identification symbol and stores all important information about the data [10]. Based on NLP (Natural Language Processing) research, tokenization is a technique for dividing documents into separate tokens, either with spaces or punctuation, which are useful in identifying meaningful keywords and using them for further processing [11]. The omitted input strings are numbers, symbols, and punctuation. Case folding is a stage that converts text from uppercase to lowercase [12]. Stemming is an advanced level used to remove prefixes or suffixes [10] into root words [12].

Approximate String Matching is a method in computer science applied in text search, pattern recognition, and signal processing applications [13] that allow the search results to match the given keywords [14]. Even though the results are not the same, Approximate String Matching successfully finds mismatches in typing or recommendations if the exact word was not found [14]. The approximate string matching approach is a generalization of the exact string matching approach [15]. The reasons for introducing approximate string matching are: low quality of the text, heterogeneousness of databases, spelling errors in the pattern or text, searching for foreign names, and searching with uncertainty [15].

Levenshtein distance algorithm was invented by Vladimir Levenshtein, a scientist from Russia, in 1965 [16]. The edit distance calculation is obtained from a matrix that calculates the number of differences between two strings. This algorithm runs from the top left corner of a two-dimensional array containing some characters from the initial string and the target string and has a cost value. The cost value in the lower right corner becomes the edit distance value which represents the difference between the two strings. Three types of string operations can be performed by the Levenshtein Distance algorithm, namely: Character Change Operation, Character Add Operation, and Character Delete Operation. A character-changing operation is an operation to swap a character with another character. An example for this case is writing the word "tukae" should become "tukar". In this case, the character "e" is replaced by the letter "r". The character addition operation is an operation adding characters to a string. Adding characters is not only at the end of a word but can be added at the beginning or inserted in the middle of a word. An example for this case is writing the word "belajr" should become "belajar", the character "a" is added in the middle of the word. Character deletion operations are performed to remove characters from a string. For example, in the word "rusaka", the last character is omitted to become the word "rusak". In this operation, the character "a" is deleted.

```
Int LevenshteinDistance (char s[1...m], char t[1...n])
//d is a table with m+1 rows and n+1 columns
Declare int d[0...m, 0...n]
for i from 0 to m
        d[i,0] := i
for j from 0 to n
        d[0,j] := j
for i from 1 to m
        for j from 1 to n
                if s[i] = t[j] then d[i,j] := d[i-1,j-1]
                else
                        d[i,j] := minimum(
                                d[i-1,j] + 1, //deletion
                                d[I,j-1] + 1, //insertion
                                d[i-1,j-1] + 1 ) //substitution
Return d[m,n]
```

Figure 1. Pseudocode of Levenshtein Distance

Methodology

Problems that often occur in typing a document or text are usually caused by a lack of accuracy. Besides, issues arise because the author does not re-check the document or text that has just been completed. After printing a document, a writer needs to re-examine the document. If there are ambiguous or typographical errors, the author has to reprint the document, which causes the use of time to be ineffective. We need to overcome typing errors, and it is necessary to have a way that can make us easier to check whether there are errors in typing text. This spelling correction makes recommendations for the right words according to the word list contained in the dictionary. Therefore, we need an application that automatically provides word recommendations. In this study, the application uses the Levenshtein Distance algorithm to find the distance between 2 strings in a text compared to a list of words in the dictionary.

Design is the initial stage of a series of making an application program. Before creating an application program, the first stage is to design a diagram or chart to describe the process and the final result. Generally, flowchart making is useful to describe the process of compiling an application program. Based on that flowchart, developers can build application programs according to their needs. We can see the flowchart of the word improvement recommendation application program in Figure 2.

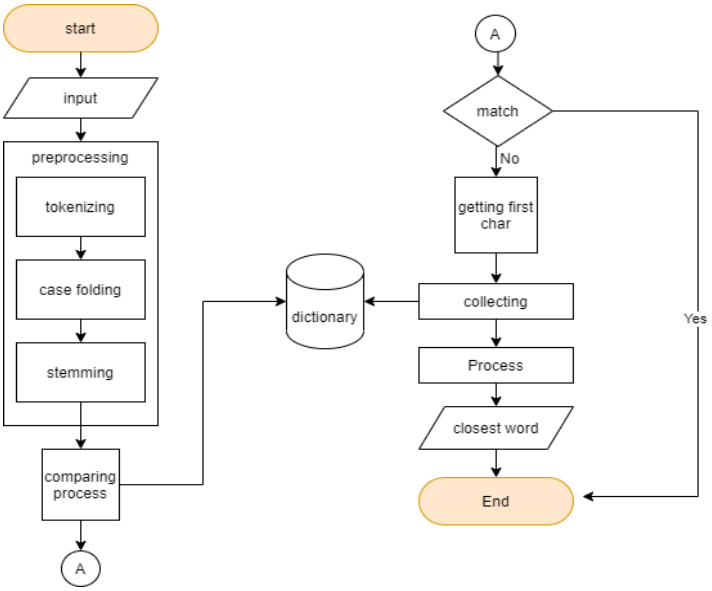


Figure 2. Application flowchart

A context diagram is a Data Flow Diagram used to define the context and system boundaries in modeling. The context diagram includes relationships with entities outside the system. It is often referred to as DFD Level 0 and is the primary determinant of a system modeled in the Data Flow Diagram. The following is a picture of DFD level 0.

Figure 3. DFD level 0

After completing the context diagram, the next step is to detail the primary process in the context diagram. The circle shows an event that needs to be detailed again so that a DFD is formed, which is called a level 1 DFD. Level 1 DFD aims to provide a more in-depth view of the entire system. The existing main process will be broken down into sub-processes. The following is a picture of DFD level 1.

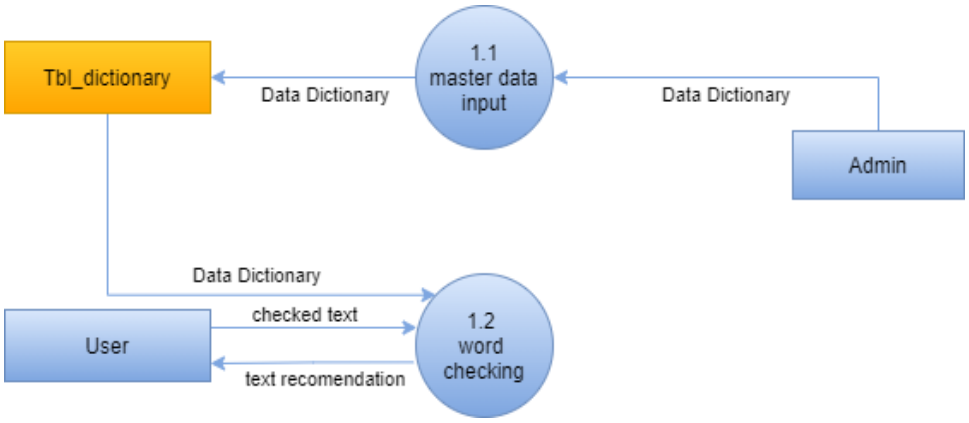


Figure 4. DFD level 1

Designing a good user interface for a system can help users learn, understand, and use it. Creating this user interface is an essential part of implementing a system. This user interface design aims to make it easier for users to use the system to be built [17]. Figure 5 and Figure 6 are the design of a user interface for a recommendation application for correcting words in a text using the Levenshtein distance algorithm.

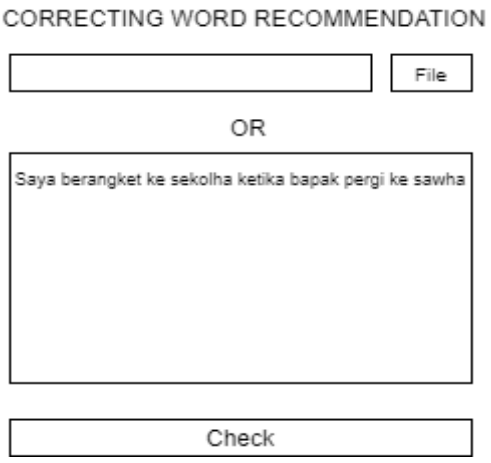


Figure 5. User interface design for the main form

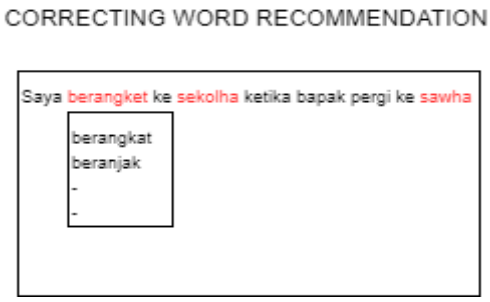


Figure 6. User interface design for recommendation form

When the user presses the Check button, the system will perform the preprocessing process. This preprocessing consists of several parts. These parts are tokenizing, case folding, and stemming. The Tokenizing process is useful for separating or removing input strings based on each constituent word or separating every word composed in a document.

```
//Tokenizing

$tokenizedText = explode(" ",$text);

$totalArrayText = count($tokenizedText);
```

Figure 7. Tokenizing

After running the tokenizing process, the system will run the case folding process. The case folding process is changing the writing of capital letters to lowercase letters in a text.

```
//Case Folding
for($a=0; $a<$totalArrayText; $a++)
{
        $tokenizedText[$a] = strtolower($tokenizedText[$a]);
}
```

Figure 8. Case folding

The last process in the preprocessing stage is stemming. Stemming is a process step that functions to remove prefixes or suffixes to become the root word. After the preprocessing operation is complete, the system will continue the counting process with Levenshtein Distance. The system will display the results in the form of word recommendations based on the Levenshtein Distance calculation results.

Results and Discussions

Testing of this recommendation application is carried out based on the user's input, which is a combination of letters, numbers, and punctuation marks. Users freely enter words, numbers, and punctuation marks. But the app checks the spelling of the word only and shows the closest recommended term. The following testing can be seen in Figure 9:

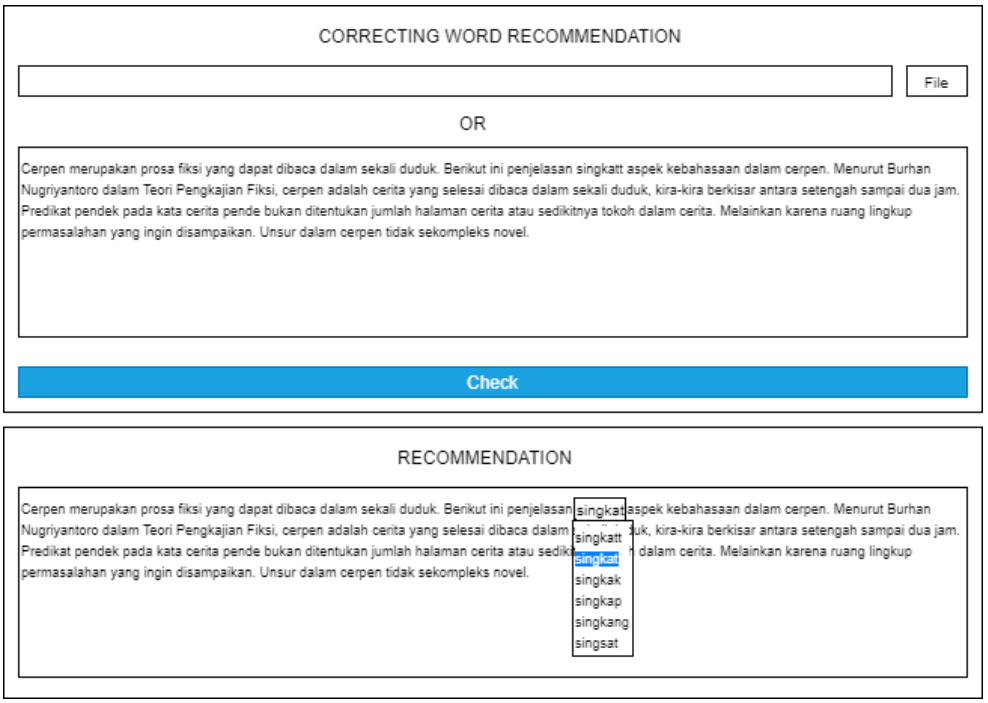


Figure 9. Testing

The following is a formula for calculating the level of accuracy of the word improvement recommendations. The formula is shown in Equation (1) below.

$$accuracy = \frac{number\ of\ words\ correct}{total\ word\ count} \times 100\% \quad (1)$$

This word improvement recommendation application was tested on ten different texts. The first was a text consisting of <100 words, the second was a text consisting of 100 to 250 words, and the third was a text consisting of 250 to 500 words. The following is a recapitulation of the average calculation results.

Results

Results should be clear and concise. The results should summarize (scientific) findings rather than providing data in great detail. Results of data analyses can be presented in tables, graphs, figures, or any combination of the three. The authors are advised to use proper variation in presenting tables, graphs, or verbal descriptions. All displayed tables and graphs should be referred to in the text. What answer was found to the research question; what did the study find? Was the tested hypothesis true?. The level of accuracy can be seen in Table 1 below.

Table 1. The level of accuracy

Number of Words	Accuracy
<100	95%
100 – 250	94%
250 - 500	90%

Tests were carried out 30 times with 30 different data. The first ten texts are less than 100 words long, the following ten texts are 100 to 250 words long, and the last ten texts are 250 to 500 words long. Table 1 shows the average level of accuracy of the above tests. The test results on these three conditions showed that the accuracy value is equally high, more than or equal to 90%. This fact means that the Approximate String-Matching method, which in this study uses the Levenshtein Distance algorithm, has a good level of accuracy. These results are in line with Braddley's statement, which states that the Approximate String Matching method has a

reasonably good level of accuracy, performance, approximate approach, and metric MAP (Mean Average Precision) value [3]. In other applications, the correction system is based solely on a dictionary or a collection of vocabulary words stored in a file. The solution to the problem offered in this study is the use of a table that contains a collection of standard words and non-standard words. Here, when a word is not found in the dictionary, it will be checked based on the standard word table. This case is also one of the reasons why the accuracy percentage is not high.

Conclusion

Based on the discussion and analysis results of the development of the Word Improvement Recommendation Application using the Levenshtein Distance Algorithm, we can conclude that:

a) The word improvement recommendation application using the Levenshtein Distance algorithm has an average accuracy rate of 95% for text <100 words, whereas for text with 100-250 words it has an accuracy rate of 94%, and for text with 250-500 words it has an accuracy rate of 90%.
b) The downside of this application is that this application still cannot detect errors for incorrect affixes and cannot compare two words with a different number of characters.

Acknowledgments

Thank you to the editors and reviewers of the Journal of Technology and Information Management for their willingness to process this article.

References

[1] Y. P. Putri and R. Lawson, "Aplikasi pengkoreksi kesalahan ejaan dan padanan kata pada tugas akhir mahasiswa," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 14, no. 2, pp. 72–75, 2019.

[2] E. S. Nasution, N. A. Hasibuan, and S. Suginam, "Implementasi algoritma approximate string matching pada aplikasi filosofi berbasis android," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 466–470, 2018.

[3] M. O. Braddley, M. Fachrurrozi, and Y. Novi, "Pengoreksian ejaan kata berbahasa Indonesia menggunakan Algoritma Levensthein Distance," in *Prosiding Annual Research Seminar*, 2017, vol. 3, no. 1, pp. 167–171.

[4] M. R. Qhadafi, "Analisis kesalahan penulisan ejaan yang disempurnakan dalam teks negosiasi siswa SMA Negeri 3 Palu," *J. Bhs. dan Sastra*, vol. 3, no. 4, pp. 1–21, 2018.

[5] A. I. Fahma, "Identifikasi kesalahan penulisan kata (typographical error) pada dokumen berbahasa Indonesia menggunakan metode N-gram dan Levenshtein Distance," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 53–62, 2018.

[6] B. V. Indriyono, "Kombinasi Damerau Levenshtein dan Jaro-Winkler Distance untuk koreksi kata bahasa Inggris," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 162–173, 2020.

[7] W. S. Ismail, P. W. Purnawan, I. Riyanto, and N. Nazori, "Sistem perekaman pelat nomor mobil pada palang pintu parkir menggunakan web kamera dan mikrokontroler," *Matrix Jurnal Manajemen Teknologi dan Informatika*, vol. 10, no. 3, pp. 103–112, 2020.

[8] N. H. Ariyani, Sutardi, and R. Ramadhan, "Aplikasi pendeteksi kemiripan isi teks dokumen menggunakan metode Levenshtein Distance," *semanTIK*, vol. 2, no. 1, pp. 279–286, 2016.

[9] L. Hermawan and M. Bellaniar Ismiati, "Pembelajaran text preprocessing berbasis simulator untuk mata kuliah information retrieval," *J. Transform.*, vol. 17, no. 2, pp. 188–199, 2020.

[10] A. I. Kadhim, "An evaluation of preprocessing techniques for text classification," *Int. J. Comput. Sci. Inf. Secur.*, vol. 16, no. 6, pp. 22–32, 2018.

[11] J. Joseph and J. R. Jeba, "Information extraction using tokenization and clustering methods," *Int. J. Recent Technol. Eng.*, vol. 8, no. 4, pp. 3690–3692, 2019.

[12] S. Suprianto, S. Sunardi, and A. Fadlil, "Aplikasi sistem temu kembali angket mahasiswa menggunakan application of information retrieval for opinion student," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 1, pp. 33–40, 2018.

[13] A. Rasool, A. Tiwari, G. Singla, and N. Khare, "String matching methodologies: A comparative analysis," *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 2, pp. 3394–3397, 2012.

[14] F. Friendly, "Jaro-Winkler Distance improvement for approximate string search using indexing data for multiuser application," in *Journal of Physics: Conference Series*, 2019, vol. 1361, no. 1, pp. 1–7.
[15] K. Al-Khamaiseh and S. Alshagarin, "A survey of string matching algorithms," *J. Eng. Res. Appl.*, vol. 4, no. 2, pp. 144–156, 2014.
[16] M. M. Yulianto, R. Arifudin, and A. Alamsyah, "Autocomplete and spell checking levenshtein distance algorithm to getting text suggest error data searching in library," *Sci. J. Informatics*, vol. 5, no. 1, pp. 67–75, 2018.
[17] I. Astrina, M. Z. Arifin, and U. Pujianto, "Penerapan algoritma FP-Growth dalam penentuan pola pembelian konsumen pada kain tenun medali mas," *Matrix Jurnal Manajemen Teknologi dan Informatika*, vol. 9, no. 1, pp. 32–40, 2019.

Matrix: Jurnal Manajemen Teknologi dan Informatika
Volume 11 Issue 2 Year 2021 Pages 71-80
DOI: http://dx.doi.org/10.31940/matrix.v11i2.2449
URL: http://ojs.pnb.ac.id/index.php/matrix

Exploratory data analysis of crime report

Irwan Setiawan [1]*, Suprihanto [2]

[1,2] *Department of Computer Engineering and Informatics, Politeknik Negeri Bandung, Indonesia*

***Corresponding Author: *irwan@jtk.polban.ac.id*

Abstract: Visualization of data is the appearance of data in a pictographic or graphical form. This form facilitates top management to understand the data visually and get the messages of difficult concepts or identify new patterns. The approach of the personal understanding to handle data; applying diagrams or graphs to reflect vast volumes of complex data is more comfortable than presenting over tables or statements. In this study, we conduct data processing and data visualization for crime report data that occurred in the city of Los Angeles in the range of 2010 to 2017 using R language. The research methodology follows five steps, namely: variables identification, data pre-processing, univariate analysis, bivariate analysis, and multivariate analysis. This paper analyses data related to crime variables, time of occurrence, victims, type of crime, weapons used, distribution, and trends of crime, and the relationship between these variables. As the result shows, by using those methods, we can gain insights, understandings, new patterns, and do visual analytics from the existing data. The variations of crime variables presented in this paper are only a few of the many variations that can be made. Other variations can be performed to get more insights, understandings, and new patterns from the existing data. The methods can be performed on other types of data as well.

Keywords: data visualization, exploratory analysis, visual analytics, data analysis, crime report

History Article: Submitted 28 March 2021 | Revised 29 April 2021 | Accepted 8 June 2021

How to Cite: I.Setiawan and Suprihanto, "Exploratory data analysis of crime report," *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 11, no. 2, pp. 72-81, 2021.

Introduction

Data visualization is the display of data in the form of images or graphics that can help decision-makers to be able to understand data visually and get new patterns hidden in the data. Visualization of complex and large amounts of data is more manageable for humans to understand when using pictures or graphics compared to being displayed in tabular or written form.

In a modern digital era, visions used in critical organization decision making gathered from Exploratory Data Analysis (EDA). EDA is the technique of studying one or more datasets to recognize the underlying structure of the data carried there [1]. EDA can be used to identify hidden patterns and correlations among variables in the data and assist people in confirming predictions from the data. Over the last few decades, academics have introduced various tools and techniques to visualize hidden correlations among data variables using simplistic diagrams and charts [2]–[8]. Visual data analysis aid domain-specific data interpretation such as analysis of CRISPR/Cas9 screens [2], analysis of container shipping slot bookings [9], analysis of executive functions during childhood [10], analysis of kindergarten students log data [11], sodium and potassium coronate stability [12], fault injection campaigns [13], employee demographics and earnings [14], airport waiting times [15], analysis of medical data [16] to perform analytics tasks, and analysis of Airbnb's super host profile [17]. Crime is a risk that must be faced and managed. The results from the EDA can be used as input for performing identification, analysis, and plans for handling potential risks that exist in the city [18].

Unemployment, poverty, urbanization, and rapid population growth are the primary causes of social di-lemmas. One of these problems implicit in every city is a crime. For example, as reported in [19], Indonesian police reported that the crime rate per 100,000 population in 2017 is 129 people. Although it experienced a decline from 2016, which numbered 140 people, the decline occurred less than 10 percent to lessen criminality rates, police have collected a large

amount of data to analyze. The study of criminal activity and the forecast of the number of crimes remains one of the most exciting problems for researchers. Research related to crime has been widely carried out [20], [21], [22], [23].

In this study, data processing and visualization were carried out for crime report data that occurred in the city of Los Angeles in the range of 2010 to 2017. Visualization of data related to crime variables, time of occurrence, victims, types of crime, weapons used, distribution, and crime trends, and the relationship between these variables is elaborated to be further used by decision-makers to conduct further analysis.

Methodology

The research methodology, as shown in Figure.1, follows five steps, namely: variables identification, da-ta pre-processing, univariate analysis, bivariate analysis, and multivariate analysis.

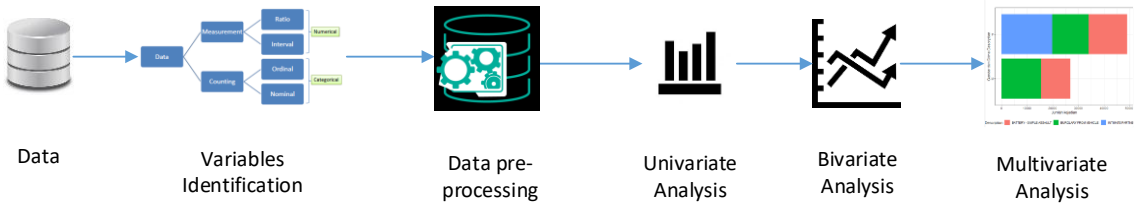


Figure 1. EDA steps

Variables identification: this is an essential step to clearly distinguish and understand the meaning of each variable in a dataset before analyzing the data. Datasets commonly have numerical, ordinal, or nominal variables [1]. An essential characteristic of numerical data is that we can apply many mathematical operations to it. A nominal, categorical, or factor variable cannot apply to mathematical operations. Ordinal variables, also referred to as ordered categorical variables or ordered factors, is a non-numeric value but possess an inherent order.

Data pre-processing: this is the second step of the EDA process. This process performs data integration (such as finding redundant attributes and tuple duplication and inconsistency), data cleaning, imputation of missing values [24], dealing with noisy data, and data reduction [25].

Univariate data analysis: the objective of the univariate analysis is to get a better understanding of each attribute. In this step, we analyze each attribute to understand how each attribute looks like. We use the ggplot2 package to visualize the data. Bivariate data analysis: the objective of the bivariate analysis is to analyze relationships between two attributes. In this step, we compare two attributes to analyze the correlation between them. We use the ggplot2 package to visualize the data. Multivariate data analysis: the objective of multivariate analysis is to get a more in-depth investigation from more than two attributes. In this step, we compare three or more attributes to analyze the correlation between them. We use the ggplot2 package to visualize the data.

Results and Discussions

The raw data are collected from the Los Angeles Police Department. The dataset reflects incidents of crime in the City of Los Angeles from 2010. The dataset represents a transcribed report from the original crime report, which is typed on paper. The original data includes over one point nine million data points for the period of 1st January 2010 to 25th November 2019. The crime report attribute includes division of records number made up of a two-digit year and five digits area ID, date reported, date occurred, time occurred, an area which referred to as geographic areas within the department, area name which represents a name designation that references a landmark of the surrounding community that is responsible for, reporting district number made up of a four-digit code that represents a sub-area within a geographic area, crime code which indicates the crime committed, modus operandi, victim age, and sex, victim descent, premise code which represents the type of structure, vehicle, or location where the crime took place, the weapon used, the status of the case, criminal code, the location which represents the

street address of crime incident rounded to the nearest hundred blocks to maintain anonymity, cross street, latitude, and longitude.

The data pre-processing step consists of removing the missing data, changing the data type of some at-tributes, rename the name of attributes, and finding redundant attributes and tuple duplication and inconsistency. In this study, because there are many NULL values in the data range 2018 to 2019, the range of data to be explored is from 1st January 2010 to 31st December 2017. From 1,900,312 crime report data will only be used 1,895,619 data.

Using R programming language and charts, we can analyze the crime data according to its variables, time of occurrence, victims, type of crime, weapons used, distribution of incidents, and trends of crime. Figures 2, 3, 4, and 5 show the distribution of crime incidents per year, per month, per day, and date respectively, from 2010 through 2017. Figures 2, 3, 4, 5, 6, 7, and Table 1 are an example of the results of univariate analysis. Figures 9, 12, and 14 are an example of the results of bivariate analysis. Figures 10, 11, and 13 are an example of the results of multi-variate analysis.

Figure. 2 shows the number of crime incidents distributed between 2010 and 2017. At the end of 2010, the Los Angeles Police Department recorded 208,883 crime reports. The number of crime incidents decreases significantly and reaches the minimum at the end of 2013. Since then, the number of crime incidents increases and reaches the maximum at the end of 2017. Signifi-cantly from 2014 to 2015, the number of crime incidents increased by almost ten percent.

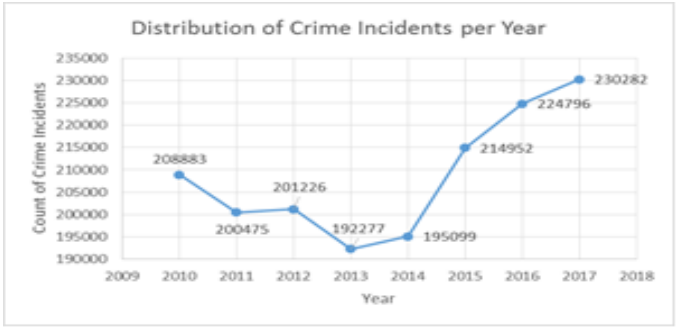


Figure 2. Distribution of crime incidents per year between 2010 and 2017

Figure 3 illustrates the distribution of crime per month from 2010 through 2017. The chart shows that February is the lowermost month of crime incidents (141,088) in Los Angeles. For the other months, it fluctuates between 150,000 and 165,000.

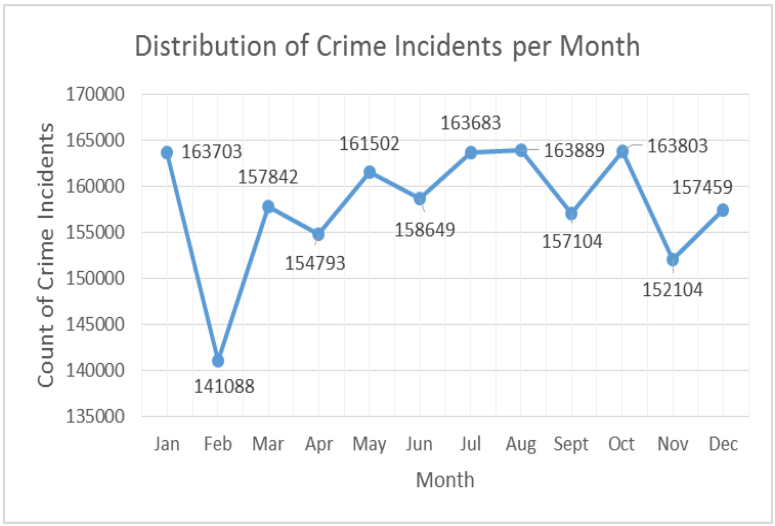


Figure 3. Distribution of crime incidents per month from 2010 through 2017

Figure 4 illustrates the distribution of crime incidents per day from 2010 through 2017. From Monday to Thursday, crime incidents fluctuate between 26,500 to 27,000, and it increases

surprisingly to 291,092 incidents on Friday. Then, it goes down to the lowest on Sunday (260,735).

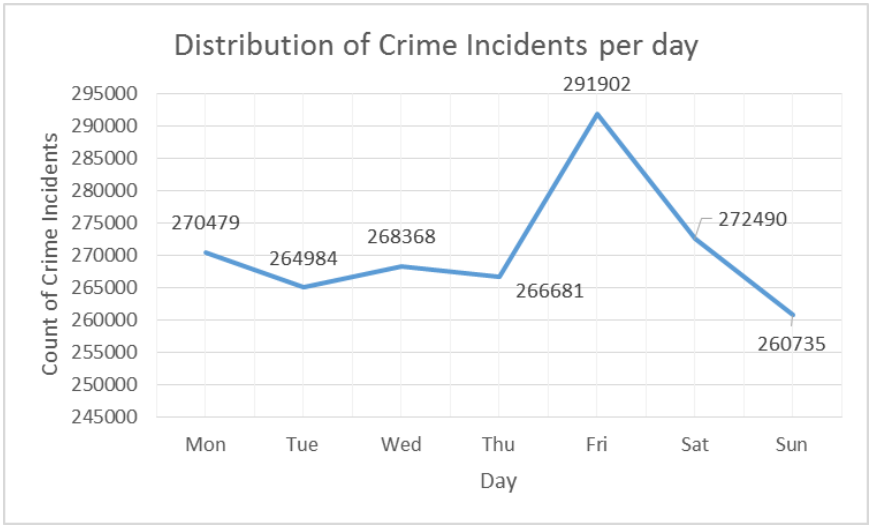


Figure 4. Distribution of crime incidents per day from 2010 through 2017

Figure 5 shows the highest ten of count of incidents according to the time when the incidents occur for the years 2010 through 2017. Interestingly enough, we can see from Figure 5. that crimes mostly happen at 12 o'clock.

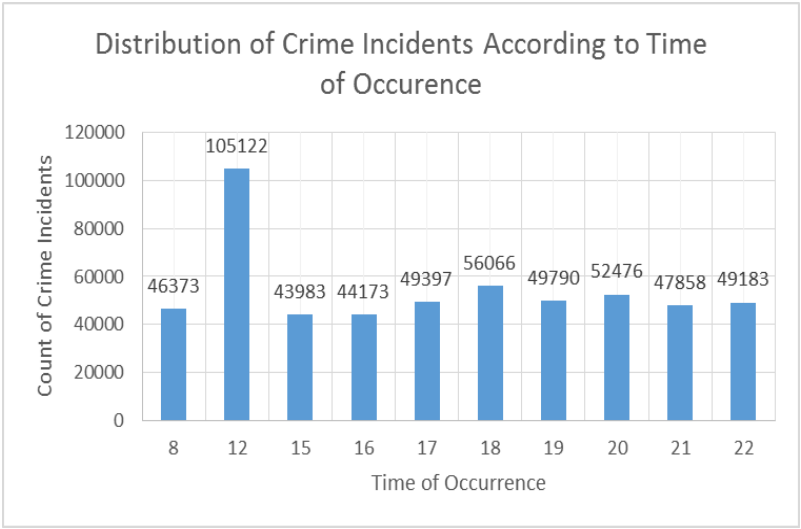


Figure 5. Distribution of crime incidents according to the time when the incidents occur during 2010 through 2017

Figure 6 shows that crime incidents from the 2nd through the 30th day of every month fluctuates between 56,000 and 66,000 incidents. Surprisingly, we can see that most crimes occur on the 1st of every month (96,879 incidents), and the lowest is on the 31st (36,851 incidents).

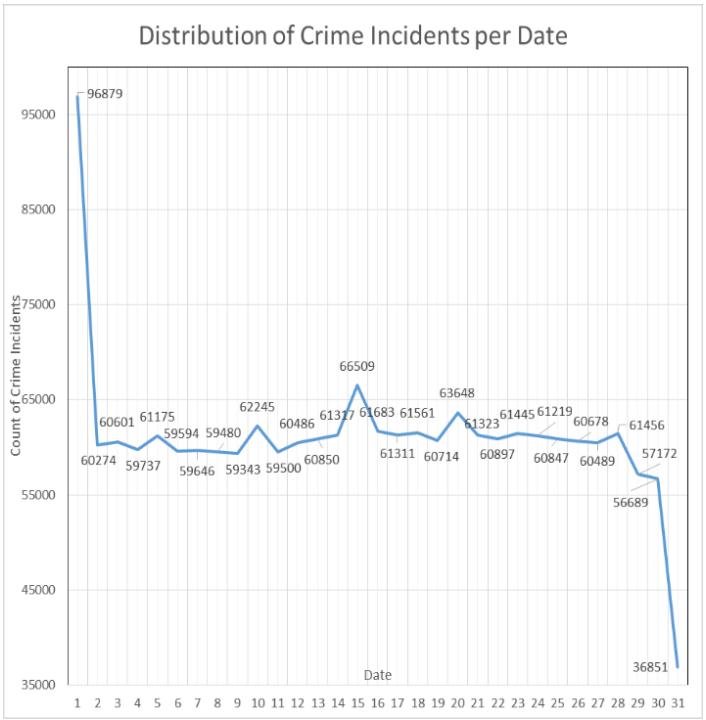


Figure 6. Distribution of crime incidents per date during 2010 through 2017

Figure 7 takes a detailed look at what type of crimes happen at 12 o'clock. It is found that most of the crimes that happen at that time are theft of identities, such as theft of name, theft of identifying number, theft of credit card number, and theft of social security number.

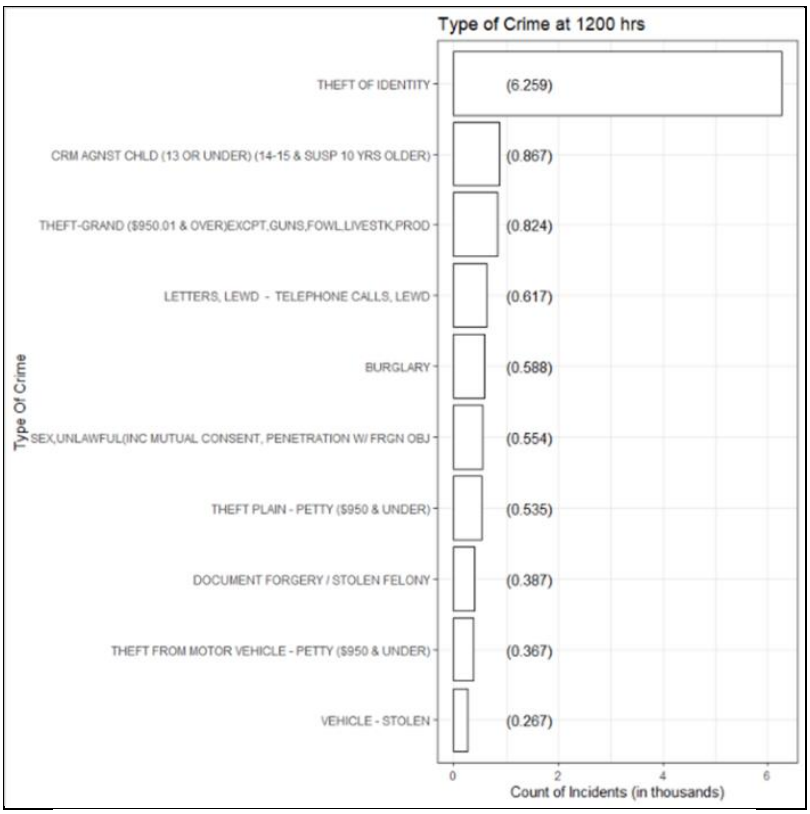


Figure 7. The top ten types of crimes happen at 12 o'clock

Figure 8 describes the distribution of crime incidents according to the gender of the victim from 2010 until 2017. It shows that through those years, the count of crime incidents is nearly the same for the male victim and female victim, with males suffers more crimes than females.

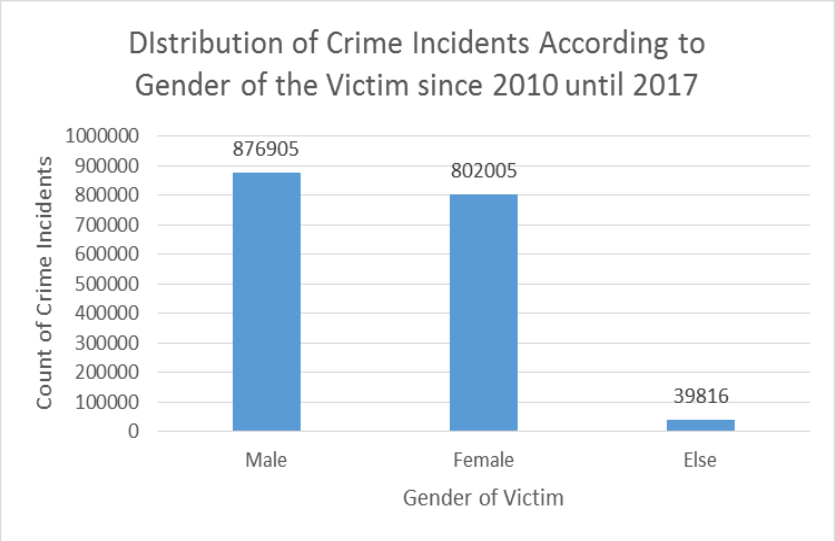


Figure 8. Distribution of Crime Incidents According to Gender of the Victim from 2010 until 2017

Figure 9 takes a look at the distribution of crime incidents according to the gender of the victim and their ages from 2010 until 2017. It shows that males and females between ages 20 and 55 suffer more crimes. Fe-males age 20 and 35 suffer most crimes. Taking a more in-depth look at what type of crimes happens to females age between 20 and 35, as shown in Figure 10, we find that they suffer the most from intimate partner-simple assault. As for males of that range of age, we find that most crime that happens to them is burglary from a vehicle.

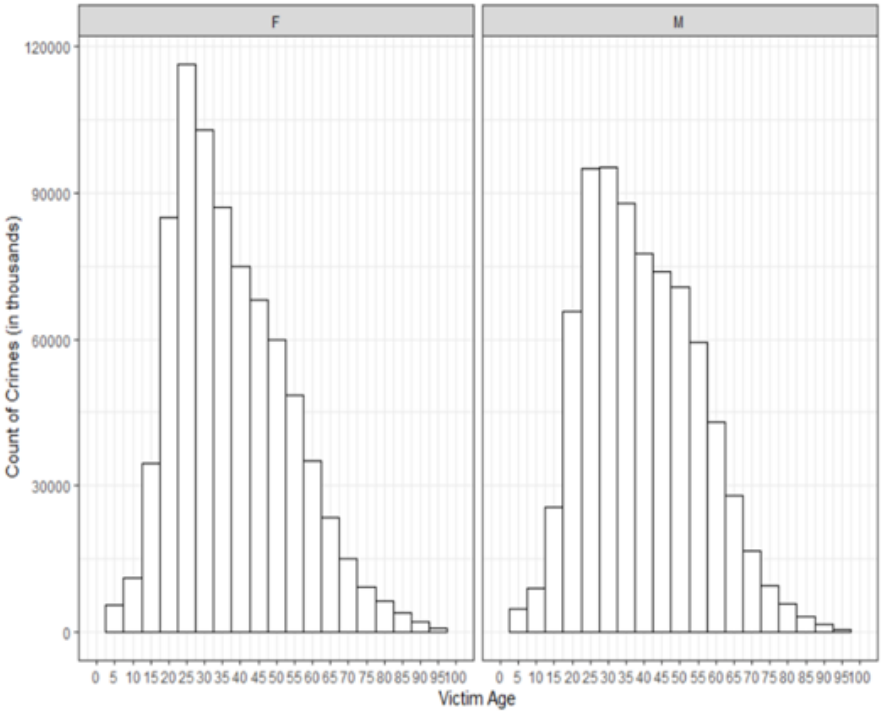


Figure 9. Distribution of Crime Incidents According to Gender of the Victim and Their Age from 2010 to 2017

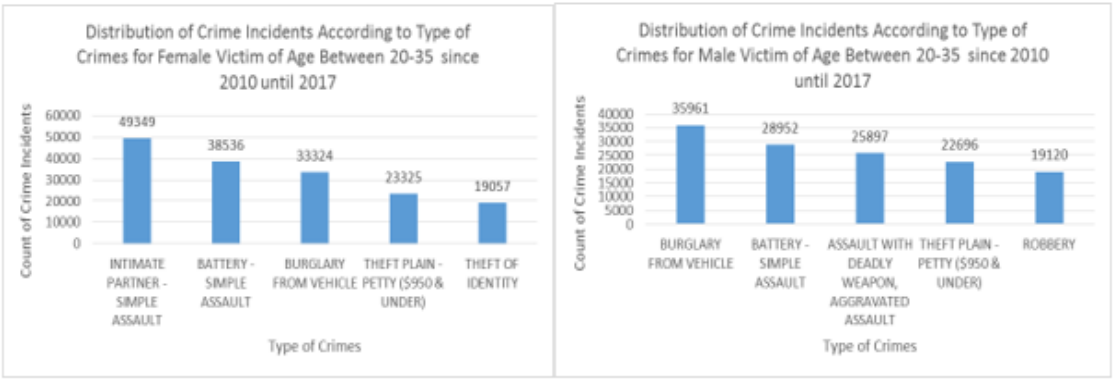


Figure 10. Top 5 of crime incidents according to the type of crimes for female victims and male victims of age between 20 and 35 from 2010 until 2017

Figure 11 shows the top 5 crime incidents according to the type of crimes for female victims and male victims from 2010 until 2017. It is interesting to note that both victims suffer the most from battery-simple as-sault. It is also interesting to know that, within the top 5 types of crime, female victims suffer intimate partner-simple assault, while male victims do not. However, in these top 5 types of crime, male victims suffer assault with a deadly weapon, while female victims do not.

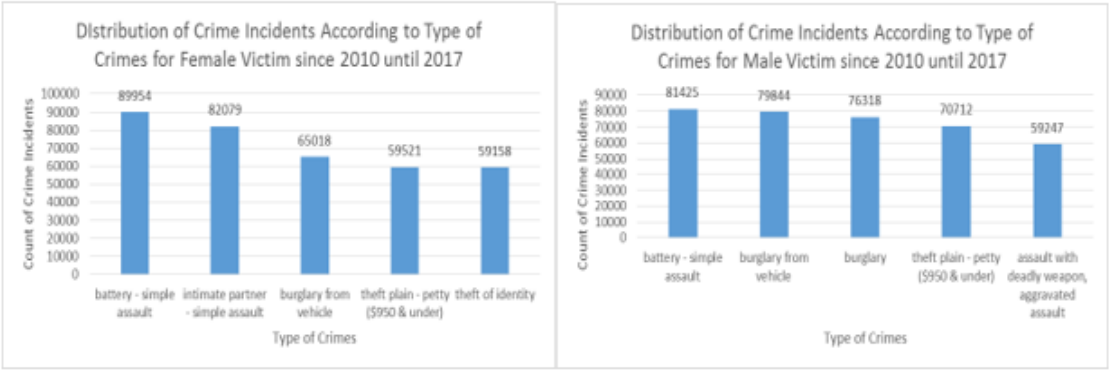


Figure 11. Top 5 of Crime Incidents According to Type of Crimes and Gender since 2010 until 2017

Figure 12 shows the distribution of crime incidents according to the premises of happening from 2010 until 2017. From the figure.12, we find that most crimes happen on the street. The second place and third place of happening are, respectively, a single-family dwelling and multi-unit dwelling such as apartments, duplex, etc.

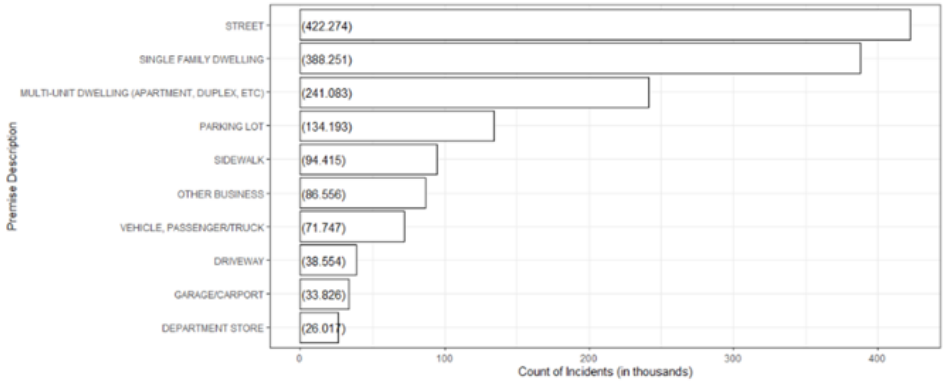


Figure 12. Distribution of Crime Incidents According to Premises since 2010 until 2017

Figure 13 shows the type of crimes that happens on those premises. Table 1 shows the count of a type of crime that happens on those premises.

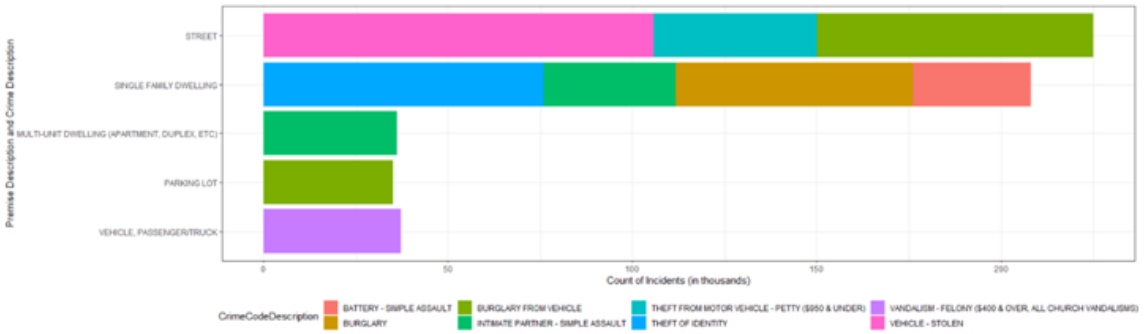


Figure 13. Type of crimes happens on the premises

Table 1. Count of type of crimes happens on the premises

Premises	Crime Type	Count
Street	Vehicle - Stolen	105,757
Street	Burglary From Vehicle	74,978
Street	Theft From Motor Vehicle - Petty ($950 & Under)	44,282
Single Family Dwelling	Theft Of Identity	75,914
Single Family Dwelling	Burglary	64,177
Single Family Dwelling	Intimate Partner - Simple Assault	35,915
Single Family Dwelling	Battery - Simple Assault	32,059
Multi-Unit Dwelling (Apartment, Duplex, etc.)	Intimate Partner - Simple Assault	36,125
Parking Lot	Burglary From Vehicle	35,015
Vehicle, Passenger/Truck	Vandalism - Felony ($400 & Over, All Church Vandalisms)	37,214

Figure 14 shows the map of the top ten locations of crimes from 2010 until 2017. The Southwest area with reporting district number 0363 is the most unsafe area with 9,609 incidents.

Figure 14. Map of Top Ten Location of Crimes (circled) since 2010 until 2017

Conclusion

This paper presents the result of Exploratory Data Analysis (EDA) using univariate analysis, bivariate analysis, and multivariate analysis. R programming language applied to 1,895,619 rows and 28 columns of Los Angeles Crime Report Data from 2010 until 2017. As the result shows, by using those methods, we can gain insights, understandings, and new patterns from the existing data. By performing EDA we can analyze the data using tables and various types of charts such as line charts, bar charts, stacked charts, and geo charts.

The variations of crime variables presented in this paper are only a few of the many variations that can be made. Other variations can be performed to get more insights, understandings, and new patterns from the existing data. The methods can be performed on other types of data as well.

References

[1] R. K. Pearson, *Exploratory Data Analysis Using R*, 1st Editio. CRC Press/Taylor & Francis Group, 2018.

[2] J. Winter, M. Breinig, F. Heigwer, D. Brügemann, S. Leible, O. Pelz, T. Zhan, and M. Boutros."CaRpools: An R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens," Bioinformatics, vol. 32, pp. 632–634, 2016.

[3] R. L. Nuzzo, "Histograms: A Useful Data Analysis Visualization," *PM R*, 2019, doi: 10.1002/pmrj.12145.

[4] P. Vermeesch, "Exploratory analysis of provenance data using R and the provenance package," *Minerals*, vol. 9, no. 3, 2019, doi: 10.3390/min9030005.

[5] A. Perer and B. Shneiderman, "Integrating statistics and visualization," p. 265, 2008, doi: 10.1145/1357054.1357101.

[6] N. Verbeeck, R. M. Caprioli, and R. Van de Plas, "Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry," *Mass Spectrom. Rev.*, p. mas.21602, Oct. 2019, doi: 10.1002/mas.21602.

[7] J. Camacho, R. A. Rodríguez-Gómez, and E. Saccenti, "Group-Wise Principal Component Analysis for Exploratory Data Analysis," *J. Comput. Graph. Stat.*, vol. 26, no. 3, pp. 501–512, 2017, doi: 10.1080/10618600.2016.1265527.

[8] A. Watson, S. Bateman, and S. Ray, "PySnippet: Accelerating exploratory data analysis in Jupyter Notebook through facilitated access to example code," *CEUR Workshop Proc.*, vol. 2322, pp. 6–9, 2019.

[9] H. Zhao, Q. Meng, and Y. Wang, "Exploratory data analysis for the cancellation of slot booking in intercontinental container liner shipping: A case study of Asia to US West Coast Service," *Transp. Res. Part C Emerg. Technol.*, vol. 106, pp. 243–263, Sep. 2019, doi: 10.1016/j.trc.2019.07.009.

[10] M. M. Richard's, S. Vernucci, F. Stelzer, I. Introzzi, and J. Guàrdia-Olmos, "Exploratory data analysis of executive functions in children: a new assessment battery," *Curr. Psychol.*, pp. 1–8, May 2018, doi: 10.1007/s12144-018-9860-4.

[11] M. Feng, D. Brenner, and A. Coulson, "Using exploratory data analysis to support implementation and improvement of education technology product," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11626 LNAI, pp. 79–83, doi: 10.1007/978-3-030-23207-8_15.

[12] N. V. Bondarev, "Classification and Prediction of Sodium and Potassium Coronates Stability in Aqueous-Organic Media by Exploratory Data Analysis Methods," *Russ. J. Gen. Chem.*, vol. 89, no. 2, pp. 281–291, Feb. 2019, doi: 10.1134/S1070363219020191.

[13] F. Cerveira *et al.,* "Exploratory data analysis of fault injection campaigns," in *Proceedings - 2018 IEEE 18th International Conference on Software Quality, Reliability, and Security, QRS 2018*, Aug. 2018, pp. 191–202, doi: 10.1109/QRS.2018.00033.

[14] K. Hara, A. Adams, K. Milland, S. Savage, B. V. Hanrahan, J. P. Bigham, and C. Callison-Burch. "Worker demographics and earnings on amazon mechanical turk: An exploratory analysis," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1–6, 2019.

[15] H. B. Sankaranarayanan, G. Agarwal, and V. Rathod, "An exploratory data analysis of airport wait times using big data visualisation techniques," *2016 Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2016*, pp. 324–329, 2016, doi: 10.1109/CSITSS.2016.7779379.

[16] S. Yamada, Y. Yamamoto, K. Umezawa, S. Asai, H. Miyachi, M. Hashimoto, and S. Inokuchi. "Exploratory data analysis for medical data using interactive data visualization," *2016 14th International Conference on ICT and Knowledge Engineering (ICT&KE),* 2016, pp. 7-11, doi: 10.1109/ICTKE.2016.7804091.

[17] I. Setiawan, "Analisis Eksplorasi Dan Visualisasi Profil Superhost Airbnb Kota Madrid Dan Amsterdam," *JTT (Jurnal Teknol. Ter.*, vol. 6, no. 2, p. 156, 2020, doi: 10.31884/jtt.v6i2.274.

[18] I. Setiawan, "Pengembangan Prototipe Aplikasi Manajemen Risiko Berbasis ISO 31000," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 10, no. 1, pp. 26–33, 2020, doi: 10.31940/matrix.v10i1.1817.

[19] S. S. P. & Keamanan, "Statistik Kriminal 2018," 2018.

[20] S. Pradhan, Divyansh, M. Pandey, and S. S. Rautaray, "Analysis of Suicides in India—A Study Using the Techniques of Big Data," in *Lecture Notes in Networks and Systems*, vol. 41, Springer, 2019, pp. 327–338.

[21] C. Catlett, E. Cesario, D. Talia, and A. Vinci, "A data-driven approach for spatio-Temporal crime predictions in smart cities," in *Proceedings - 2018 IEEE International Conference on Smart Computing, SMARTCOMP 2018*, Jul. 2018, pp. 17–24, doi: 10.1109/SMARTCOMP.2018.00069.

[22] M. Feng, J. Zheng, Y. Han, J. Ren, and Q. Liu, "Big Data Analytics and Mining for Crime Data Analysis, Visualization and Prediction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10989 LNAI, pp. 605–614, doi: 10.1007/978-3-030-00563-4_59.

[23] J. Zhao and C. L. Rogalin, "Heinous Crime or Unfortunate Incident: Does Gender Matter?," *Soc. Psychol. Q.*, vol. 80, no. 4, pp. 330–341, 2017, doi: 10.1177/0190272517728923.

[24] I. Setiawan, "Time series air quality forecasting with R Language and R Studio," *J. Phys. Conf. Ser.*, vol. 1450, p. 12064, Feb. 2020, doi: 10.1088/1742-6596/1450/1/012064.

[25] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, 1st Editio., vol. 72. Springer International Publishing, 2015.

Matrix: Jurnal Manajemen Teknologi dan Informatika
Volume 11 Issue 2 Year 2021 Pages 81-94
DOI: http://dx.doi.org/10.31940/matrix.v11i2.2306
URL: http://ojs.pnb.ac.id/index.php/matrix

DSS for best e-commerce selection using AHP-WASPAS and AHP-MOORA methods

Gede Surya Mahendra [1]*, I Wayan Widi Karsana [2], Anak Agung Istri Ita Paramitha [3]

[1]*Informatics Engineering Study Program, STMIK STIKOM Indonesia, Indonesia*
[2]*Recording and Health Information Study Program, Universitas Dhyana Pura, Indonesia*
[3]*Information Systems Study Program, STMIK Primakara, Indonesia*

*****Corresponding Author: *gede.mahendra@stiki-indonesia.ac.id*

Abstract: eIQ Consumer Pulse 2019 determines the 6 largest e-commerce in Indonesia, namely Blibli, Bukalapak, JD.id, Lazada Indonesia, Shopee and Tokopedia. This is due to an increase in e-commerce transactions. The COVID-19 pandemic, which emphasizes social distancing and physical distancing, has also played a role in increasing these transactions. Indonesia is also predicted to become the market leader in Southeast Asia due to this. On the other hand, each e-commerce customer has their preferences in choosing e-commerce to use in transactions. Various criteria make customers confused due to intense competition between e-commerce companies. DSS is a solution in choosing the right e-commerce for each customer's preferences. The AHP-WASPAS and AHP-MOORA methods can be used in calculations for determining favorite e-commerce in Indonesia. The CRISP-DM framework also helps in preparing the research flow well. 3 decision-makers are used to provide weighting criteria using AHP. The results of this study indicate that the Tokopedia alternative is the best e-commerce, with a preference value of 0.8964 for AHP-WASPAS and 0.4245 for AHP-MOORA. The second and third places are Bukalapak and Lazada, respectively. The weighting of the criteria by the decision-maker, the alternative normalization process and the calculation technique for the preference value have a significant impact on the ranking results.

Keywords: AHP, DSS, e-commerce, MOORA, WASPAS

History Article: Submitted 28 April 2021 | Revised 18 May 2021 | Accepted 15 June 2021

How to Cite: G.S. Mahendra, I.W.W Karsana, A.A.I.I Paramitha, "DSS for best e-commerce selection using AHP-WASPAS and AHP-MOORA methods," *Matrix : Jurnal Manajemen Teknologi dan Informatika*, vol. 11, no. 2, pp. 82-95, 2021.

Introduction

Given that 96% of internet users in Indonesia have used e-commerce, it is hoped that e-commerce in Indonesia can develop rapidly and become a leader in the Southeast Asian market [1]. According to 2019 data, the value of e-commerce transactions in Indonesia is the US $ 21 billion and is estimated to reach the US $ 82 billion in transaction value by 2025 [2]. The current situation regarding the impact of COVID-19 on the global e-commerce industry shows that daily web traffic has increased significantly by more than 50%, which may be due to the social and physical restrictions experienced by consumers [3]. The government also protects consumers when transacting through e-commerce through preventive and repressive methods [4]. This situation allows e-commerce to penetrate deeper and gain growth in the Indonesian market. E-commerce is also required to always maintain its excellence in providing quality products and services. Indonesia has a lot of e-commerce, and Indonesia has the 6 largest e-commerce companies consisting of Blibli, Bukalapak, JD.id, Lazada, Shopee, and Tokopedia [5]. Behind the convenience provided, there are still several negative factors, such as product mismatches, delivery problems, security of payment methods, and customer service. Due to intense competition, many consumer considerations, it is difficult to choose between the same products but the prices offered are different, so that consumers are still confused about choosing the right and trusted e-commerce transaction. To solve this problem, a decision support system (DSS) can be used to provide advice in choosing the right e-commerce. DSS itself is an effective system that can assist

users in making complex decisions. This system uses decision rules, analysis models, comprehensive databases, and decision-maker knowledge [6], [7], [8].

In this study, the method used was a combination of AHP-WASPAS and AHP-MOORA. This combination of methods was chosen because AHP is a functional hierarchy with the main input being human perception [9], [10], [11], [12], [13]. The WASPAS method can reduce errors or optimize evaluation to select the highest and lowest scores [14], [15], [16], [17]. The MOORA method has a good level of selectivity in determining an alternative [18], [19], [20], [21]. In previous studies, the combination of these methods in the DSS has been applied well, decision-makers can weigh the criteria and greatly influence the results of recommendations but have not been implemented in software [22], [23], [24]. Regarding the choice of e-commerce, several studies have compared different alternative criteria and methods, and achieved good results [25], [26], [27], [28]. However, there has been no comprehensive analysis on how to compare how the combination of these methods is implemented in the software, taking into account the weighting of criteria for different Decision Maker.

Therefore, this study aims to be able to perform calculations manually and implement a combination of the AHP-WASPAS and AHP-MOORA methods in a software product. The urgency of this research, if not realized, could result in obstruction of the development of the DSS method which can only reach the calculation and design stages manually, thus hindering innovation in the DSS field. Based on the background previously described, it is necessary to realize a combination of the AHP-WASPAS and AHP-MOORA methods to determine the best e-commerce using DSS.

Methodology

The research method used in this study follows the various stages of the CRISP-DM model [29]. Data-related problems such as data mining and DSS can use the CRISP-DM method, which is expected to analyze business problems and current conditions, provide appropriate data conversion to provide a model that can evaluate the effectiveness, and record the results obtained. CRISP-DM solves this problem by defining a process model related to data mining and DSS, regardless of the problem department or technology used.

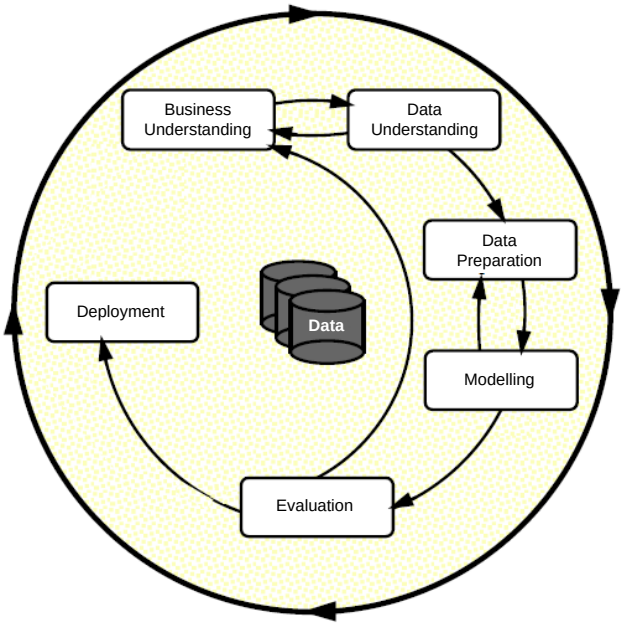


Figure 1. Current process model phase related to CRISP-DM [29]

Business understanding is the stage used to determine business goals, analyze business conditions, and determine the objectives of the DSS. At this stage, a thorough understanding is

carried out based on the results of the analysis of observations, interviews, and supporting documents for the objectives and results of the research. Several options can be found when determining the best e-commerce in Indonesia. Based on the alternatives obtained, calculations are made to determine the ranking. The best e-commerce results can be the best recommendations for consumers to make digital transactions. On the other hand, an e-commerce that has not achieved the best results can still improve its performance to gain a better market share. When determining the number and alternative criteria for the best e-commerce candidates, refer to the assessment in the 2019 Consumer Pulse eIQ survey and get 6 alternatives namely Blibli, Bukalapak, JD.id, Laza-da Indonesia, Shopee, and Tokopedia. The decision-makers used are 3 netizens who are actively using e-commerce. The weights of the criteria were obtained from the Decision Maker and were calculated using AHP, while the evaluation of the alternative ranking used the WASPAS and MOORA methods.

At data understanding stage, it starts with the process of data collection, data analysis, and evaluation of the quality of the data used in the study. To be able to use the AHP-WASPAS and AHP-MOORA methods correctly, appropriate criteria and alternative data are needed. The criteria used in this study include (C1) reputation, (C2) price, (C3) product, (C4) customer service, (C5) delivery, (C6) application & UX, (C7) payment, and (C8) security. & Policy. Reputation is a good name for e-commerce in society. Price is the nominal amount that consumers have to pay, and how cheap it is compared to competing e-commerce. Products cover product range, product authenticity, product selection, and product quality. Customer service is good customer service. De-livery includes the speed of delivery, the conditions of free shipping, and the time it takes to process the shipment. App & UX includes easy-to-use apps, mobile apps, and UX that satisfy users. Payment methods cover a variety of payment methods, payment processing, and the quality of each payment method. Security and policies relating to returns, membership programs, promotions, and safety.

At data preparation stage includes selecting the data used and published to be included in the DSS calculation. At this stage, data cleaning is also carried out to repair, remove or ignore the noise in the data. At the business understanding stage, the tools, techniques, or methods used in this study have been selected. In this Modeling stage, AHP-WASPAS and AHP-MOORA methods were chosen to determine the best e-commerce in Indonesia. Before continuing the research, you can do a test design with the data to prove the method can be used. A flowchart of the method used can be seen in Figure 2.

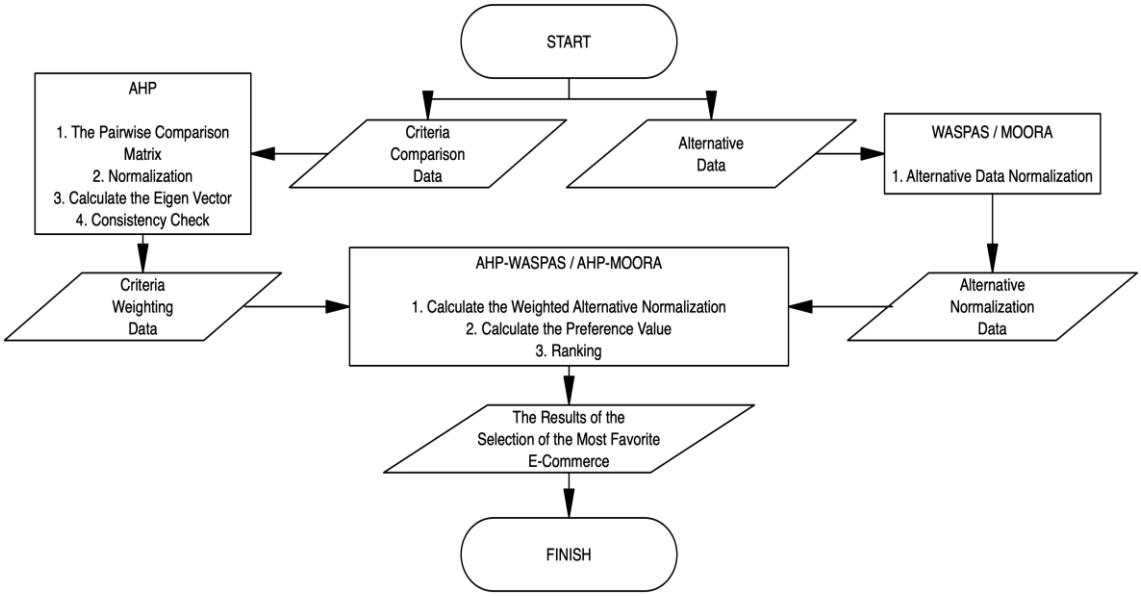


Figure 2. Flowchart of using the AHP-WASPAS and AHP-MOORA methods

The first step is to prepare comparison data between the criteria provided by the decision-maker as a resource and alternative data is Indonesian e-commerce data based on the 2019 eIQ Consumer Pulse survey. Starting from the determination of the pairwise comparison matrix, the

AHP method is used to determine the criteria weighting data, then normalize it, compute eigenvectors and check hierarchy consistency. To calculate the consistency index using AHP can be seen in equation (1), and to calculate the consistency ratio can be seen in equation (2) [30].

$$CI = \frac{(\lambda_{max} - n)}{(n-1)} \quad (1)$$

$$CR = \frac{CI}{IR} \quad (2)$$

Furthermore, the WASPAS method and the MOORA method are used to normalize alternative data to produce normalized alternative data. Criteria weight data results from the calculation of the AHP method, and alternative data normalized using the WASPAS and MOORA methods are used for weighted normalization calculations and calculating preference values, as well as producing ratings based on preference values that can determine the best e-commerce ranking. This can be a reference for customers or input as a refinement of e-commerce, which still lacks in some aspects. To calculate the preference value using WASPAS using equation (3), which consists of calculations using WSM in equation (4) and WPM in equation (5) [14]. Normalization using MOORA using equation (6) and preference value using MOORA using equation (7) [19].

$$Q_i = \frac{1}{2}\left(Q_i^{(1)} + Q_i^{(2)}\right) \quad (3)$$

$$Q_1^{(1)} = \sum_{j=1}^{n} x_{ij} w_j \quad (4)$$

$$Q_1^{(2)} = \prod_{j=1}^{n} x_{ij} w_j \quad (5)$$

$$X^*_{ij} = \frac{x_{ij}}{\sqrt{\left[\sum_{j=1}^{m} (x_{ij})^2\right]}} \quad (6)$$

$$y^*_j = \sum_{j=1}^{g} w_j x^*_{ij} - \sum_{i=g+1}^{n} w_j x^*_{ij} \quad (7)$$

At the evaluation stage, testing is carried out based on the results of the DSS recommendations and the performance of the methods used. Calculations must be tested manually, and the results obtained when implemented in software have the same value to have compatibility between the two. Sensitivity testing is used to compare the performance between the WASPAS and MOORA methods to measure which method is more sensitive to changes in weighting criteria, therefore the more sensitive the better. At the Deployment stage, a deployment plan is carried out based on previous assessments. If the test results show good results, further implementation can be planned. Apart from deployment planning, monitoring and maintenance plan can also be planned to produce a final report on the research results.

Results and Discussions

Results

This research is based on questionnaire data from users who are very familiar with e-commerce, the questionnaire is transformed using the AHP method into weighting criteria and e-commerce data as an alternative. The number of Decision Makers used to produce weighting criteria is 3 people, and the amount of e-commerce data used is 6 companies. The calculation starts using the AHP method. There are 8 criteria, namely (C1) reputation, (C2) price, (C3) product, (C4) customer service, (C5) delivery, (C6) application & UX, (C7) payment methods, and (C8) security & policies. Pairwise comparison matrices from Decision Maker 1. 2 and 3 are shown in Table 1. Table 2 and Table 3. Equally important weights are not shown in the table. EQ means equally important, MD means moderate important, ST means strongly important, DM means demonstrated important and EX means extremely important.

Table 1. Pairwise comparison matrix from decision maker 1

Criteria		C1	C2	C3	C4	C5	C6	C7	C8
Reputation	C1	-	MD		MD	MD	MD	MD	
Price	C2		-						
Product	C3		MD	-					
Customer Service	C4				-				
Delivery	C5					-			
Application & UX	C6						-		
Payment Methods	C7							-	
Security & Policies	C8		MD	MD	MD	MD	MD		-

Table 2. Pairwise comparison matrix from decision maker 2

Criteria		C1	C2	C3	C4	C5	C6	C7	C8
Reputation	C1	-	MD		MD	MD	MD	MD	
Price	C2		-						
Product	C3	MD	MD	-					
Customer Service	C4				-				
Delivery	C5					-			
Application & UX	C6						-		
Payment Methods	C7							-	
Security & Policies	C8	MD	ST	ST	ST	ST	ST	ST	-

Table 3. Pairwise comparison matrix from decision maker 3

Criteria		C1	C2	C3	C4	C5	C6	C7	C8
Reputation	C1	-	MD		MD	MD	MD	MD	
Price	C2		-						
Product	C3	MD		-	MD	MD	MD		
Customer Service	C4				-				
Delivery	C5					-			
Application & UX	C6						-		
Payment Methods	C7							-	
Security & Policies	C8	MD				MD	MD		-

Furthermore, the calculation is focused on Decision Maker 1. The calculation steps for other decision-makers are the same as the calculation for Decision Maker 1. The criteria comparison matrix for decision-maker 1 is translated, based on the Saaty scale presented in Table 4. The results of the pairwise comparison matrix transformation using Saaty scale are presented in Table 5.

Table 4. Saaty scale

Intensity	Description / Linguistics
1	Equaly Important (EQ)
3	Moderate Important (MD)
5	Strongly Important (ST)
7	Demonstrated Important (DM)
9	Extremely Important (EX)
2,4,6,8	Intermediate Value

Table 5. Pairwise comparison matrix translated using the saaty scale from decision maker 1

Criteria		C1	C2	C3	C4	C5	C6	C7	C8
Reputation	C1	1	3	1	3	3	3	3	1
Price	C2	1/3	1	1/3	1	1	1	1	1/3
Product	C3	1	3	1	1	1	1	1	1/3
Customer Service	C4	1/3	1	1	1	1	1	1	1/3
Delivery	C5	1/3	1	1	1	1	1	1	1/3
Application & UX	C6	1/3	1	1	1	1	1	1	1/3
Payment Methods	C7	1/3	1	1	1	1	1	1	1
Security & Policies	C8	1	3	3	3	3	3	1	1
SUM		4 2/3	14	9 1/3	12	12	12	10	4 2/3

Normalization in the AHP method is done by dividing the element values by the number of column values. The eigenvector value is generated based on the number of criteria for each row, as follows.

$$C_{11} = \frac{1}{4^2/_3} = 0.214 \qquad C_{12} = \frac{3}{14} = 0.214 \qquad C_{13} = \frac{1}{9^1/_3} = 0.107 \qquad C_{14} = \frac{3}{12} = 0.250$$

$$C_{15} = \frac{3}{12} = 0.250 \qquad C_{16} = \frac{3}{12} = 0.250 \qquad C_{17} = \frac{3}{10} = 0.300 \qquad C_{18} = \frac{1}{4^2/_3} = 0.214$$

$$EV\ C_1 = \frac{0.214 + 0.214 + 0.107 + 0.250 + 0.250 + 0.250 + 0.300 + 0.214}{8} = \frac{1.800}{8} = 0.2250$$

For the following criteria, use the same formula to produce the eigenvector values shown in table 6. as follows.

Table 6. Eigen vector for decision-maker 1

Criteria		Eigen Vector
Reputation	C1	0.225
Price	C2	0.075
Product	C3	0.120
Customer Service	C4	0.084
Delivery	C5	0.084
Application & UX	C6	0.084
Payment Methods	C7	0.101
Security & Policies	C8	0.227

After obtaining the eigenvector for each criterion, λ_{max} can be calculated from the pairwise comparison matrix multiplied by the eigenvector[30]. Each product yield is divided by the feature vector, and the average value is λ_{max}. Use the following steps to determine the λ_{max} of Decision Maker 1.

$$\lambda = \begin{bmatrix} 1.0 & 3.0 & 1.0 & 3.0 & 3.0 & 3.0 & 3.0 & 1.0 \\ 0.3 & 1.0 & 0.3 & 1.0 & 1.0 & 1.0 & 1.0 & 0.3 \\ 1.0 & 3.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0.3 \\ 0.3 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0.3 \\ 0.3 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0.3 \\ 0.3 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0.3 \\ 0.3 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 1.0 & 3.0 & 3.0 & 3.0 & 3.0 & 3.0 & 1.0 & 1.0 \end{bmatrix} \begin{bmatrix} 0.225 \\ 0.075 \\ 0.120 \\ 0.084 \\ 0.084 \\ 0.084 \\ 0.102 \\ 0.227 \end{bmatrix} = \begin{bmatrix} 1.857 \\ 0.619 \\ 0.999 \\ 0.699 \\ 0.699 \\ 0.699 \\ 0.850 \\ 1.893 \end{bmatrix}$$

$$\lambda_{max} = \frac{\left(\frac{0.1857}{0.225} + \frac{0.619}{0.075} + \frac{0.999}{0.120} + \frac{0.699}{0.084} + \frac{0.699}{0.084} + \frac{0.699}{0.084} + \frac{0.850}{0.102} + \frac{1.893}{0.227}\right)}{8}$$

$$\lambda_{max} = \frac{8.254 + 8.254 + 8.348 + 8.326 + 8.326 + 8.326 + 8.351 + 8.346}{8} = 8.317$$

After getting λ_{max}, the following steps can be used to calculate the decision-maker 1 consistency index, using equation (1).

$$CI = \frac{(8.317-8)}{(8-1)} = \frac{0.317}{7} = 0.0453$$

After obtaining the consistency index (CI), then the consistency ratio can be calculated for Decision Maker 1 [30]. Based on the Alonso-lamata RI values, which are presented in Table 7. considering the number of criteria is 8. the IR used is 1.4056. CR can be calculated using equation (2).

Table 7. Alonso-Lamata RI values

Number of Elements	3	4	5	6	7	8
Alonso-Lamata RI Values	0.5245	0.8815	1.1086	1.2479	1.3417	1.4056

$$CR = \frac{0.0453}{1.4056} = 0.0322$$

Because CR is less than 0.1. the hierarchy is considered consistent, so the calculation is declared true and can be used as a criterion weight. The same steps as Decision Maker 1 are also used to calculate the comparison matrix between Decision Maker 2 and 3 to obtain the eigenvector value of all Decision Maker. To find the weighted average of all Decision Maker, the geometric mean (GEOMEAN) calculation is carried out based on the weighting of the criteria for all Decision Maker, the results of which are presented in Table 8. If the number of weighted criteria from GEOMEAN does not equal 1. the weighting of the criteria is normalized. The steps for obtaining the weighted average criteria are as follows.

Table 8. Eigen value by the Three Decision Maker and the Geometric Mean (GEOMEAN)

Criteria		**EV N1**	**EV N2**	**EV N3**	**Geomean EV**	**Normalized Geomean EV**
Reputation	C1	0.225	0.165	0.189	0.192	0.196
Price	C2	0.075	0.062	0.102	0.078	0.080
Product	C3	0.120	0.124	0.197	0.143	0.146
Customer Service	C4	0.084	0.070	0.087	0.080	0.082
Delivery	C5	0.084	0.070	0.073	0.075	0.077
Application & UX	C6	0.084	0.070	0.073	0.075	0.077
Payment Methods	C7	0.102	0.070	0.102	0.090	0.092
Security & Policies	C8	0.227	0.369	0.177	0.245	0.251
SUM		1	1	1	0.979	1

After getting the weighted criteria results, continue using the WASPAS method and the MOORA method to calculate the preference value. When using the WASPAS method to calculate, starting from the alternative normalization calculation, calculating the preference value and ranking. The e-commerce data used includes Blibli (EC1), Bukalapak (EC2), JD.ID (EC3), Lazada (EC4), Shopee (EC5) and Tokopedia (EC6). Based on predetermined e-commerce data, the results are shown in Table 9 below.

Table 9. E-commerce alternative value

Alternative		**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**
Blibli	EC1	14.8	29.9	138.6	18.4	48.1	19.8	13.6	15.0
Bukalapak	EC2	13.0	21.9	141.8	23.3	32.7	32.4	13.7	21.4
JD.ID	EC3	12.9	26.4	143.7	8.3	54.2	22.8	17.2	13.7
Lazada	EC4	13.7	23.5	143.0	15.8	46.2	19.5	16.1	22.2
Shopee	EC5	10.9	23.6	144.6	18.7	48.4	23.0	9.9	22.4
Tokopedia	EC6	14.3	21.7	145.5	21.4	30.7	31.6	12.2	22.7

Based on these alternative data, the WASPAS method can be used to calculate the alternative normalization. By dividing the criteria value by the maximum value of the criteria in the column, welfare conditions can be standardized using the WASPAS method [14]. The following

calculation shows an example of the criteria in the EC1 alternative. For the next alternative, use the same formula to generate the priority and rating values shown in Table 10, as follows.

$r_{11} = \frac{14.8}{14.8} = 1;$ $r_{12} = \frac{29.9}{29.9} = 1;$ $r_{13} = \frac{138.6}{145.5} = 0.953;$ $r_{14} = \frac{18.4}{23.3} = 0.790;$

$r_{15} = \frac{48.1}{54.2} = 0.887;$ $r_{16} = \frac{19.8}{32.4} = 0.611;$ $r_{17} = \frac{13.6}{17.2} = 0.791;$ $r_{18} = \frac{15.0}{22.7} = 0.661;$

Table 10. E-commerce alternative normalization data using WASPAS

Alternative		C1	C2	C3	C4	C5	C6	C7	C8
Blibli	EC1	1.00	1.00	0.95	0.79	0.89	0.61	0.79	0.66
Bukalapak	EC2	0.88	0.73	0.97	1.00	0.60	1.00	0.80	0.94
JD.ID	EC3	0.87	0.88	0.99	0.36	1.00	0.70	1.00	0.60
Lazada	EC4	0.93	0.79	0.98	0.68	0.85	0.60	0.94	0.98
Shopee	EC5	0.74	0.79	0.99	0.80	0.89	0.71	0.58	0.99
Tokopedia	EC6	0.97	0.73	1.00	0.92	0.57	0.98	0.71	1.00

After obtaining the alternative normalization value, the calculation of the preference value in WASPAS is obtained from the combination of additive importance and multiplicative importance values [14]. The following calculation shows the calculation of the value of the importance of the AHP-WASPAS additive in alternative 1, using formula (4), and the calculation of the AHP-WASPAS multiplicative importance value in alternative 1, using formula (5).

$$Q_1{}^{(1)} = \sum \begin{bmatrix} (0.196 \times 1.00); (0.080 \times 1.00); (0.146 \times 0.95); (0.082 \times 0.79) \\ (0.077 \times 0.89); (0.077 \times 0.61); (0.092 \times 0.79); (0.251 \times 0.66) \end{bmatrix}$$
$$Q_1{}^{(1)} = 0.196 + 0.080 + 0.139 + 0.064 + 0.068 + 0.047 + 0.073 + 0.166 = 0.8330$$

$$Q_1{}^{(2)} = \prod \begin{bmatrix} (1.00^{0.196}); (1.00^{0.080}); (0.95^{0.146}); (0.79^{0.082}) \\ (0.89^{0.077}); (0.61^{0.077}); (0.79^{0.092}); (0.66^{0.251}) \end{bmatrix}$$
$$Q_1{}^{(2)} = 1.000 \times 1.000 \times 0.993 \times 0.981 \times 0.991 \times 0.963 \times 0.979 \times 0.901 = 0.8196$$

For the next alternative, use the same formula to produce additive and multiplicative importance values. Furthermore, the calculation of preference values can be done using AHP-WASPAS using formula (3), with the results of the preference values shown in Table 11.

$$Q_1 = \frac{1}{2}(0.8330 + 0.8196) = 0.8263$$

Table 11. Preference value and ranking using AHP-WASPAS

No	Alternative		Preference Value	Ranking
1	Tokopedia	EC6	0.8964	Ranked 1st
2	Bukalapak	EC2	0.8834	Ranked 2nd
3	Lazada	EC4	0.8814	Ranked 3rd
4	Shopee	EC5	0.8356	Ranked 4th
5	Blibli	EC1	0.8263	Ranked 5th
6	JD.ID	EC3	0.7737	Ranked 6th

After getting the preference value on the AHP-WASPAS, to find the preference value using the MOORA method, it can be done by calculating alternative normalization, calculating the optimization value, and ranking it. Based on predetermined alternative data, the following calculation shows an example of calculating the normalized value for the EC1 alternative using formula (6).

$$X^*{}_{11} = \frac{14.8}{\sqrt{[(14.8)^2+(13)^2+(12.9)^2+(13.7)^2+(10.9)^2+(14.3)^2]}} = 0.453$$
$$X^*{}_{12} = \frac{29.9}{\sqrt{[(29.9)^2+(21.9)^2+(26.4)^2+(23.5)^2+(23.6)^2+(21.7)^2]}} = 0.495$$

$$X^*_{13} = \frac{138.6}{\sqrt{[(138.6)^2+(141.8)^2+(143.7)^2+(143)^2+(144.6)^2+(145.5)^2]}} = 0.396$$
$$X^*_{14} = \frac{18.4}{\sqrt{[(18.4)^2+(23.3)^2+(8.3)^2+(15.8)^2+(18.7)^2+(21.4)^2]}} = 0.411$$
$$X^*_{15} = \frac{48.1}{\sqrt{[(48.1)^2+(32.7)^2+(54.2)^2+(46.2)^2+(48.4)^2+(30.7)^2]}} = 0.444$$
$$X^*_{16} = \frac{19.8}{\sqrt{[(19.8)^2+(32.4)^2+(22.8)^2+(19.5)^2+(23)^2+(31.6)^2]}} = 0.318$$
$$X^*_{17} = \frac{13.6}{\sqrt{[(13.6)^2+(13.7)^2+(17.2)^2+(16.1)^2+(9.9)^2+(12.2)^2]}} = 0.397$$
$$X^*_{18} = \frac{15}{\sqrt{[(15)^2+(21.4)^2+(13.7)^2+(22.2)^2+(22.4)^2+(22.7)^2]}} = 0.307$$

For the next alternative, use the same formula to produce the alternative normalized values shown in Table 12 as follows.

Table 12. E-commerce alternative normalization data using MOORA

Alternative		C1	C2	C3	C4	C5	C6	C7	C8
Blibli	EC1	0.45	0.49	0.40	0.41	0.44	0.32	0.40	0.31
Bukalapak	EC2	0.40	0.36	0.41	0.52	0.30	0.52	0.40	0.44
JD.ID	EC3	0.40	0.44	0.41	0.19	0.50	0.37	0.50	0.28
Lazada	EC4	0.42	0.39	0.41	0.35	0.43	0.31	0.47	0.46
Shopee	EC5	0.33	0.39	0.41	0.42	0.45	0.37	0.29	0.46
Tokopedia	EC6	0.44	0.36	0.42	0.48	0.28	0.51	0.36	0.47

After obtaining the alternative normalization value, the calculation of the preference value in MOORA is continued with the calculation of the optimization value. The following calculation shows the preference value calculated using AHP-MOORA [19]. Table 13 shows the results of preference values and ratings using AHP-MOORA, using formula (7).

$$y^*_1 = \sum \begin{bmatrix} (0.196 \times 0.45); (0.080 \times 0.49); (0.146 \times 0.40); (0.082 \times 0.41); \\ (0.077 \times 0.44); (0.077 \times 0.32); (0.092 \times 0.40); (0.251 \times 0.31) \end{bmatrix} = 0.3919$$

Table 13. Preference value and ranking using AHP-MOORA

No	Alternative		Preference Value	Ranking
1	Tokopedia	EC6	0.4245	Ranked 1st
2	Bukalapak	EC2	0.4186	Ranked 2nd
3	Lazada	EC4	0.4159	Ranked 3rd
4	Shopee	EC5	0.3955	Ranked 4th
5	Blibli	EC1	0.3919	Ranked 5th
6	JD.ID	EC3	0.3707	Ranked 6th

Discussions

The results of manual calculations using AHP-WASPAS and AHP-MOORA have also been applied to web-based software, and the results are in accordance with manual calculations that have been tested previously using Microsoft Excel. Figure 3 shows the pairwise comparisons normalization between criteria using AHP. Figure 4 shows the eigenvector calculations results using AHP.

Figure 5 shows the Web-Based Implementation of the Alternative Data. Figure 6 shows the preference value calculation results using AHP-WASPAS. Figure 7 shows the preference value calculation results using AHP-MOORA. Figure 8 shows the preference value comparison using AHP-WASPAS and AHP-MOORA. Figure 9 shows the preference value graph comparison using AHP-WASPAS and AHP-MOORA.

Matriks AHP

Nama Kriteria		C1	C2	C3	C4	C5	C6	C7	C8
		BEN	BEN	BEN	BEN	BEN	BEN	BEN	BEN
		CF	SF	CF	CF	SF	CF	CF	CF
Reputasi	C1	1.0000	3.0000	1.0000	3.0000	3.0000	3.0000	3.0000	1.0000
Harga	C2	0.3333	1.0000	0.3333	1.0000	1.0000	1.0000	1.0000	0.3333
Produk	C3	1.0000	3.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3333
Customer Service	C4	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3333
Pengiriman	C5	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3333
Aplikasi & UX	C6	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.2000
Metode Pembayaran	C7	0.3333	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3333
Keamanan & Kebijakan	C8	1.0000	3.0000	3.0000	3.0000	3.0000	5.0000	3.0000	1.0000
SUM		4.6667	14.0000	9.3333	12.0000	12.0000	14.0000	12.0000	3.8667

Figure 3. Web-based implementation of the pairwise comparisons normalization between criteria using AHP

Eigen Vector

Nama Kriteria	Nilai Kriteria
Reputasi	0.2198
Harga	0.0733
Produk	0.1179
Customer Service	0.0822
Pengiriman	0.0822
Aplikasi & UX	0.0779
Metode Pembayaran	0.0822
Keamanan & Kebijakan	0.2645
SUM	1.0000

Figure 4. Web-based implementation of the eigenvector calculations results using AHP

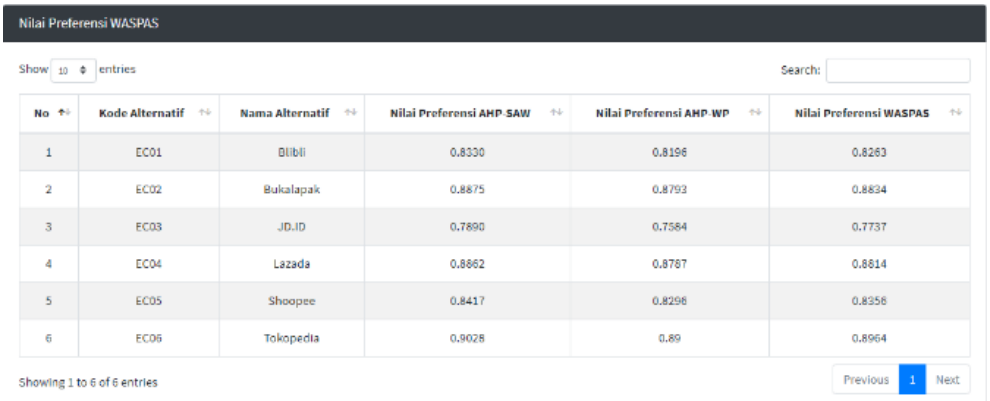
Data Alternatif

Show 10 entries Search:

Kode Alternatif	(C1)	(C2)	(C3)	(C4)	(C5)	(C6)	(C7)	(C8)
EC01	14.8	29.9	138.6	18.4	48.1	19.8	13.6	15
EC02	13	21.9	141.8	23.3	32.7	32.4	13.7	21.4
EC03	12.9	26.4	143.7	8.3	54.2	22.8	17.2	13.7
EC04	13.7	23.5	143	15.8	46.2	19.5	16.1	22.2
EC05	10.9	23.6	144.6	18.7	48.4	23	9.9	22.4
EC06	14.3	21.7	145.5	21.4	30.7	31.6	12.2	22.7
MIN	10.9	21.7	138.6	8.3	30.7	19.5	9.9	13.7
MAX	14.8	29.9	145.5	23.3	54.2	32.4	17.2	22.7

Showing 1 to 6 of 6 entries Previous 1 Next

Figure 5. Web-based implementation of the alternative data



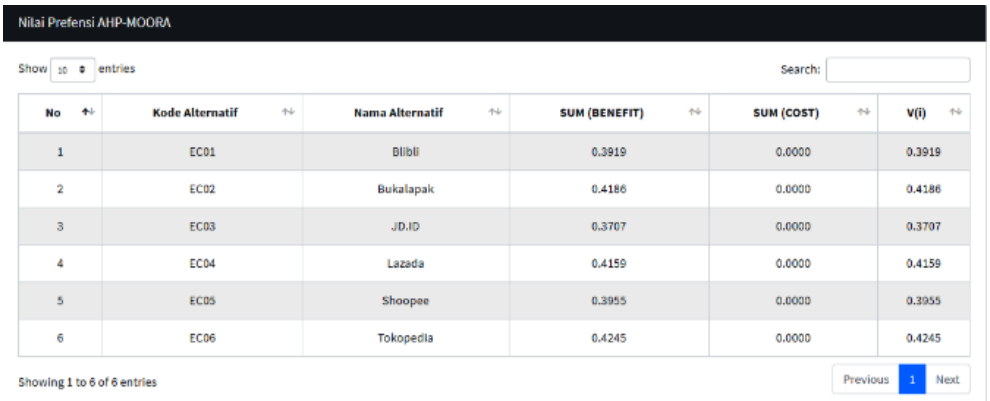
Nilai Preferensi WASPAS

Show 10 entries Search:

No	Kode Alternatif	Nama Alternatif	Nilai Preferensi AHP SAW	Nilai Preferensi AHP WP	Nilai Preferensi WASPAS
1	EC01	Blibli	0.8330	0.8196	0.8263
2	EC02	Bukalapak	0.8875	0.8793	0.8834
3	EC03	JD.ID	0.7890	0.7584	0.7737
4	EC04	Lazada	0.8862	0.8787	0.8814
5	EC05	Shoopee	0.8417	0.8296	0.8356
6	EC06	Tokopedia	0.9028	0.89	0.8964

Showing 1 to 6 of 6 entries Previous 1 Next

Figure 6. Web-based implementation of the preference value calculations results using AHP-WASPAS



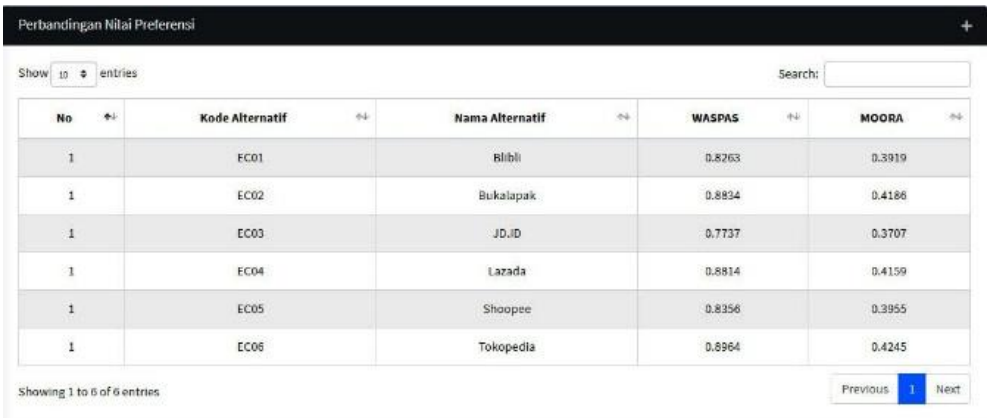
Nilai Prefensi AHP-MOORA

Show 10 entries Search:

No	Kode Alternatif	Nama Alternatif	SUM (BENEFIT)	SUM (COST)	V(i)
1	EC01	Blibli	0.3919	0.0000	0.3919
2	EC02	Bukalapak	0.4186	0.0000	0.4186
3	EC03	JD.ID	0.3707	0.0000	0.3707
4	EC04	Lazada	0.4159	0.0000	0.4159
5	EC05	Shoopee	0.3955	0.0000	0.3955
6	EC06	Tokopedia	0.4245	0.0000	0.4245

Showing 1 to 6 of 6 entries Previous 1 Next

Figure 7. Web-based implementation of the preference value calculations results using AHP-MOORA



Perbandingan Nilai Preferensi

Show 10 entries Search:

No	Kode Alternatif	Nama Alternatif	WASPAS	MOORA
1	EC01	Blibli	0.8263	0.3919
1	EC02	Bukalapak	0.8834	0.4186
1	EC03	JD.ID	0.7737	0.3707
1	EC04	Lazada	0.8814	0.4159
1	EC05	Shoopee	0.8356	0.3955
1	EC06	Tokopedia	0.8964	0.4245

Showing 1 to 6 of 6 entries Previous 1 Next

Figure 8. Web-based implementation of the preference value comparison using AHP-WASPAS and AHP-MOORA

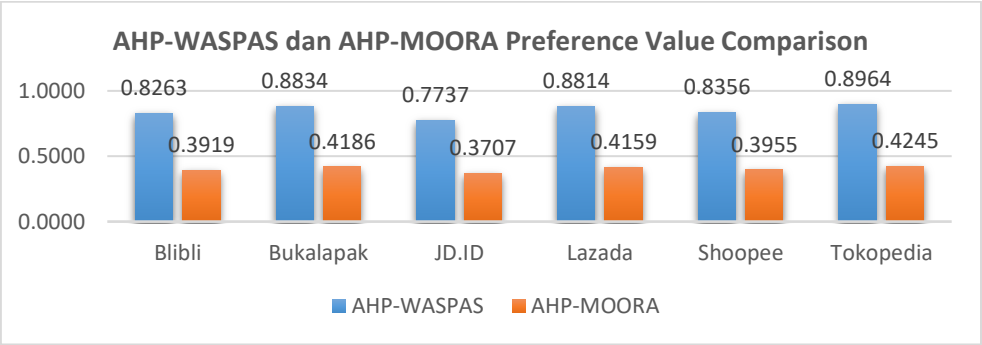


Figure 9. Preference value comparison graph using AHP-WASPAS and AHP-MOORA

Conclusion

Based on the comparative research results of the AHP-WASPAS and AHP-MOORA methods on DSS to determine the best e-commerce in Indonesia, it can be concluded that the AHP-WASPAS and AHP-MOORA methods can be used to determine the best e-commerce in Indonesia. Can be calculated manually and implemented into web-based software. The best e-commerce in Indonesia is determined based on weighting the criteria of 3 users who actively use e-commerce and based on the results of alternative evaluations obtained from the 2019 eIQ Consumer Pulse survey, which states that the e-commerce alternative "Tokopedia" as the best e-commerce has value. AHP-WASPAS preference is 0.8964 and AHP-MOORA is 0.4245. The second and third ranks were "Bukalapak" and "Lazada", respectively. The weighting of the criteria by the Decision Maker, the alternative normalization process, and the calculation technique for the preference value have an impact on the ranking.

Acknowledgments

Thanks the authors would like to say to STMIK STIKOM Indonesia, for funding in the STIKI Lecturer development research grant program for the implementation year of 2020. with contract number 07/04/LPPM/PPDS/VI/2020. Thank you to the Editorial Team of the Information and Technology Management Journal (MATRIX) for providing the opportunity for authors to publish this article.

References

[1] Y. Pusparisa and S. Fitra, "96% pengguna internet di indonesia pernah menggunakan e-commerce," *katadata.co.id*, pp. 1–1, 2019.

[2] Y. Pusparisa and S. Fitra, "Transaksi e-commerce Indonesia terbesar di Asia Tenggara," *katadata.co.id*, pp. 1–1, 2019.

[3] D. A. Widiastuti, "Covid-19 berdampak signifikan bagi e-commerce," *tek.id*, pp. 1–1, 2020.

[4] S. Rongiyati, "Pelindungan konsumen dalam transaksi dagang melalui sistem elektronik," *Negara Hukum: Membangun Hukum untuk Keadilan dan Kesejahteraan*, vol. 10, no. 1, pp. 1–25, 2019.

[5] EIQ, "EIQ consumer pulse: Uncovering the value of Indonesia's top online platform," *https://*ecommerceiq.*asia*, pp. 1–1, 2018.

[6] G. S. Mahendra and P. G. S. C. Nugraha, "Komparasi metode AHP-SAW dan AHP-WP pada SPK penentuan e-commerce terbaik di Indonesia, cComparison of AHP-SAW and AHP-WP methods on DSS to determine the best e-commerce in Indonesia," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 08, no. 4, pp. 346–356, 2020.

[7] N. K. A. P. Sari, "Implementation of the AHP-SAW method in the decision support system for selecting the best tourism village," *Jurnal Teknik Informatika C.I.T Medicom*, vol. 13, no. 1, pp. 22–31, 2021.

[8] H. T. Hidayat and N. G. A. P. H. Saptarini, "sistem pendukung keputusan pemilihan mas-kapai terbaik pada rute domestik dengan menggunakan metode Multi Factor Evaluation Process (MFEP)," *Jurnal Manajemen Teknologi dan Informatika*, vol. 4, no. 3, 2017.

[9] G. S. Mahendra and E. Hartono, "Komparasi analisis konsistensi metode AHP-MAUT dan

AHP-PM pada SPK penempatan siswa OJT," *Jurnal Teknologi Informasi Informasi dan Komputer*, vol. 7, no. 2, 2021.

[10] S. R. Arianto, S. Siswanti, and W. L. Y. Saptomo, "Sistem pendukung keputusan penerima bantuan pangan non tunai dengan metode hybrid AHP- SAW," *Transformtika,* vol. 17, no. 2, pp. 200–208, 2020.

[11] G. M. A. Azis, I. Cholissodin, and M. T. Furqon, "Sistem pendukung keputusan untuk rekomendasi wirausaha menggunakan metode AHP-TOPSIS (Studi kasus Kab. Probolinggo)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 11, pp. 1204–1214, 2017.

[12] I. Fahmi, F. Kurnia, and G. E. S. Mige, "Perancangan sistem promosi jabatan menggunakan kombinasi Analytical Hierarchy Process (AHP) dan Profile Matching (PM)," *Jurnal Spektro*, vol. 2, no. 1, pp. 26–34, 2019.

[13] P. I. Ciptayani and N. G. A. P. H. Saptarini, "Sistem pendukung keputusan pemilihan manajer proyek menggunakan analytical hierarchy process," *Matrix : Jurnal Manajemen Teknologi dan Informatika*, vol. 4, no. 2, 2017.

[14] G. S. Mahendra and I. G. B. Subawa, "Perancangan metode AHP-WASPAS pada sistem pendukung keputusan penempatan ATM," *Prosiding Seminar Nasional Pendidikan Teknik Informatika (SENAPATI) Ke-10*, vol. 1, no. 1, pp. 122–128, 2019.

[15] V. Amalia, D. Syamsuar, and L. Atika, "Komparasi metode WP SAW dan WASPAS dalam penentuan penerima beasiswa penelusuran minat dan kemampuan," *Jurnal Informatika*, vol. 6, no. 1, pp. 114–121, 2019.

[16] P. Fitriani and T. S. Alasi, "Sistem pendukung keputusan dalam menentukan judul skripsi mahasiswa dengan metode WASPAS , COPRAS dan EDAS berdasarkan penilaian dosen," *Jurnal Media Informatika Budidarma*, vol. 4, no. 4, pp. 1051–1061, 2020.

[17] M. Ickhsan, D. Anggraini, R. Haryono, S. H. Sahir, and Rohminatin, "Sistem pendukung keputusan pemberian kredit usaha rakyat menggunakan metode Weighted Aggregated Sum Product Assesment (WASPAS)," *Jurnal Riset Komputer (JURIKOM)*, vol. 5, no. 2, pp. 97–102, 2018.

[18] C. Lukita, C. Nas, and W. Ilham, "Analisis pengambilan keputusan penentuan prioritas utama dalam peningkatan kualitas mata pelajaran dengan menggunakan metode perbandingan WASPAS dan MOORA," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 3, pp. 130–137, 2020.

[19] I. G. Hendrayana and G. S. Mahendra, "Perancangan metode AHP-MOORA pada sistem pendukung keputusan pemilihan paket wisata," *Prosiding Seminar Nasional Pendidikan Teknik Informatika (SENAPATI) Ke-10*, vol. 1, no. 1, pp. 143–149, 2019.

[20] W. Brauers, "Location theory and multi-criteria decision making : An application of the MOORA method," *IDEAS*, vol. 12, no. 3, pp. 241–252, 2018.

[21] D. Assrani, N. Huda, R. Sidabutar, I. Saputra, and O. K. Sulaiman, "Penentuan penerima bantuan siswa miskin menerapkan metode Multi Objective Optimization on The Basis of Ratio Analysis (MOORA)," *Jurnal Riset Komputer (JURIKOM)*, vol. 5, no. 1, pp. 1–5, 2018.

[22] A. D. S. Sitepu and R. Sianipar, "Sistem pendukung keputusan penerima bantuan JAMKESMAS dengan menerapkan metode AHP dan WASPAS," *Seminar Nasional Sains & Teknologi Informasi (SENSASI),* pp. 265–271, 2019.

[23] Mesran, D. I. Zalukhu, S. Handayani, and D. U. Sutiksno, "Penerapan WASPAS dan MOORA dalam menentukan lokasi tempat pembuangan sampah akhir," *Seminar Nasional Sains & Teknologi Informasi (SENSASI),* pp. 347–353, 2018.

[24] N. W. A. Ulandari, "Implementasi metode MOORA pada proses seleksi beasiswa bidikmisi di Institut Teknologi dan Bisnis STIKOM Bali," *Jurnal Eksplora Informatika*, vol. 10, no. 1, pp. 53–58, 2020.

[25] I. Pratiwi and W. Andalia, "Perancangan model keputusan multikriteria pemilihan layanan e-commerce untuk kepuasan pelanggan," *Prosiding Seminar Nasional Sains dan Teknologi,* pp. 1–10, 2018.

[26] I. Pratistha, I. P. A. Mahadewa, and P. Sugiartawan, "Sistem pendukung keputusan kelompok pemilihan e-commerce/marketplace menggunakan metode profile matching dan BORDA," *Jurnal Sistem* Informasi *dan Komputer Terapan Indonesia (JSIKTI)*, vol. 1, no. 1, pp. 13–24, 2018.

[27] A. Setiyadi and R. D. Agustia, "Penerapan metode AHP dalam memilih marketplace e-

commerce berdasarkan software quality and evaluation ISO/IEC 9126-4 untuk UMKM," *Jurnal IKRA-ITH Informatika*, vol. 2, no. 3, pp. 61–70, 2018.

[28] H. S. Setiawan and A. Pamuji, "Membangun sistem pendukung keputusan untuk rekomendasi pada e-commerce melalui penerapan logika fuzzy tahani," *Faktor Exacta*, vol. 9, no. 4, pp. 341–352, 2016.

[29] R. Wirth and J. Hipp, "CRISP-DM : Towards a standard process model for data mining," *Semantic Scholar*, vol. 1, no. 24959, pp. 1–11, 2000.

[30] G. S. Mahendra and I. P. Y. Indrawan, "Metode AHP-TOPSIS pada sistem pendukung keputusan penentuan penempatan ATM," *JST (Jurnal Sains dan Teknologi)*, vol. 9, no. 2, pp. 130–142, 2020.

Journal of Manajemen Teknologi dan Informatika
Volume 11 Issue 2 Year 2021 Pages 95-106
DOI: http://dx.doi.org/10.31940/matrix.v11i2.2274
URL: http://ojs.pnb.ac.id/index.php/matrix

The system development life cycle model implementation on information system of performance reporting IT asset case study: PT Kereta Api Indonesia (Persero)

Muhamad Agung Santoso [1]*, Rina Fiati [2], Ratih Nindyasari [3]

[1,2,3] *Universitas Muria Kudus, Indonesia*

*Corresponding Author: *smuhamadagung@gmail.com*

Abstract: The asset performance monitoring should be done to find out and review the reported data whether it has been in line with the reality or not. It is important to identify the occurring problems to be immediately handled. The information system development aimed to build the existing system. Thus, it will be more effective and efficient to produce information. This research applied the System Development Life Cycle (SDLC) with Waterfall Model to design, create, maintain the information and process the system. During the creation of this information system, the authors applied the PHP language program, Framework Codeigniter, MySQL, database, and Visual Studio Code as the editorial codes. The implementation of the SIPERITA application, could be used excellently were proven effective and efficient because the application facilitated the monitoring activities by providing single application access in promoting a checking process of the condition asset then the user administrator can see a diagram of the monitoring results on the home page so that if there is a problem with the asset, direct action can be taken and with this application, users can print the reports based on daily, monthly, and annually.

Keywords: *Information, SDLC, SIPERITA Application, System*

History Article: Submitted 16 April 2021 | Revised 30 April 2021 | Accepted 17 June 2021

How to Cite: M. A. Santoso, R. Fiati, and R. Nindyasari, "The System Development Life Cycle Model Implementation on Information System of Performance Reporting IT Asset (Case Study: PT Kereta Api Indonesia (Persero))," *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 11, no. 2, pp.96-108, 2021.

Introduction

The asset condition monitoring process in the IT Operation Unit of the PT Kereta Api Indonesia has been using the *Google Form* feature as the data input. Then, the data is processed into a monthly report. The report is usually done manually with Microsoft Excel. It takes time and is not efficient to work on them [1].

This research [2] explains that a monitoring process will produce a monthly recapitulated data report. The system has the purpose to make the data accurate and the information clear. Thus, the data recapitulation and report processes will run smoothly and prevent further data duplication.The applied method [3] was Waterfall in which the original concept of the method is to see problems systematically and structurally from the beginning until the end. By creating new data supported with an asset management-data-based system, the information of the corporate asset report would be accessible quickly, easily, and efficiently.

A study by [4] states that the use of information technology could maximize the web-based inventory system application that could be accessed by the central official party and branch offices. This inventorying process consists of noting, storing, and reporting from the branch offices. It could also notice each branch office inventory data quickly and accurately. In this research, the created application was done with Macromedia Dreamweaver, PHP language program, and MySQL database.

On the other hand, a study by [5] found that asset management could be optimal by using an information system. The applied method in this research was a subject-oriented approach

with UML (Unified Modeling Language). It was by using a case diagram, class diagram, and activity diagram to facilitate understanding a designed system. This research produced an information system prototype of asset management. It was implemented at the BMKG office (Meteorology, Climatology, and Geophysics Agency) in Jambi.

The management asset is an art and knowledge to combine the resource management. It consists of planning the asset need, inventorying, legal audit of operating, maintaining, removing, and effective and efficient asset transfer [6][7]. A study by [8] found that the use of the CICT Asset Inventory and Management system facilitated the monitoring of the asset. Monitoring had a purpose to conduct research, review, identify problems, and measure progress [9]. The other website applications had been developed to support the main website application [10] [11][12].

Based on the explanation, this research aimed to apply a website-based application, the application name is Information System of Reporting IT Asset (SIPERITA), at the IT Operation Unit of the State-Owned Kereta Api Indonesia.

Methodology

The Research Stages

In this research, the authors applied the System Development Life Cycle (SDLC) with Waterfall Model. According to [13], the life cycle of this system has functions to describe the main stages and steps of each phase. The stages of the Waterfall Model could be seen in Figure 1.

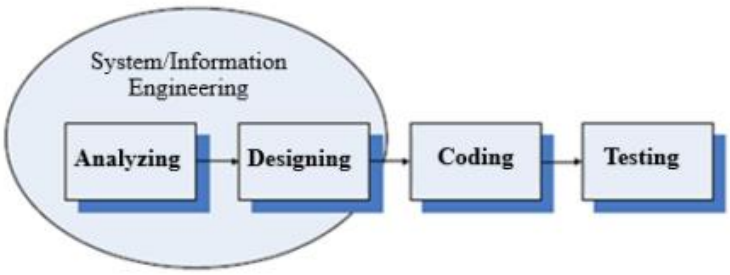


Figure 1. Waterfall model

1. The Need Analysis
   This stage was a process to collect the needed data. It was to specify the functional and non-functional needs during the asset performance monitoring processes. It included data input monitoring, report print, and other management features. This system allowed the user to carry out the monitoring process easier than the previous system.
2. The System Design
   This stage was done to design the website-based application, SIPERITA, with a Unified Modeling Language planning model. This stage facilitated the logical understanding realization of the designed system by determining the detailed specification of each information system based on the analytical stage results.
3. The Program Code Creation
   This stage was done by applying the design to the software program. The results of this stage were in line with the arranged design during the designing stage.
4. The Testing System
   This stage was done by using Black Box and Beta tests. The test was the stage to check the conditions of a system whether the output had been in line with the functional need specification.
5. The Supports and Maintenance
   The support and maintenance stages were the last stages concerning the program utility and adjustment. The activities consisted of revision, revising, and developing the implemented application.

Data Collection Methods

The method of collecting data related to the asset performance monitoring at the IT Operation unit applied several methods. They were:

Field Research

The researchers obtained the data by going into the field and conducted interviews and observation.

1. Interview
   The researchers interviewed the employees of the IT Operation unit to obtain accurate and relevant data. Thus, they could create an appropriate information system plan based on their needs.
2. Observation
   The researchers observed the asset performance monitoring process directly at the IT Operation unit of the State-Owned Kereta Api Indonesia.

Document Analysis

It was an activity to collect information about the required documents in a system. The document includes an ISO checklist form and a 2019th performance report. The purposes of the document analysis were to find out and understand what documents were needed and used in a running system.

Results and Discussions

The Need Analysis

The need data of the website-based application production, SIPERITA, for the IT Operation unit of the State-Owned Kereta Api Indonesia consisted of functional needs dan non-functional needs.

Functional Need

The functional need of an information system could be seen in Table 1.

Table 1. Functional Need

Number	Administrator	Executor
1	Login function for administrator	Login function for the executor
2	The data user management function	The asset performance monitoring data input
3	The asset performance monitoring data input	The data user management function
4	The asset performance monitoring report print	

Non-Functional Need

The non-functional need of the information system produced the following details.

Table 2. Non-Functional Need

Number	Data	Criteria
1	The data monitoring needs	Daily generator set, weekly generator set, UPS and PAC
2	The data user needs	Administrator and Executor
3	The need for the daily asset generator monitoring criteria information	Type of asset, date and time monitoring, volt charger, engine battery, water, oil, fuel daily tank, frequency, indicator, and storage tank
4	The need for the weekly asset generator monitoring criteria information	Type of asset, weekly period, the executor name, date and time monitoring, oil press KPA, water temp °C, fuel meter, hertz engine, RPM engine, and hours engine
5	The need for UPS monitoring criteria information	Type of asset, date and time monitoring, AC temperature, AC humidity, room temperature, room humidity, and internal service indicator
6	The need for PAC monitoring criteria information	Type of asset, date and time monitoring, UPS input voltage, UPS output voltage, frequency, battery room temperature, capacity module installed, capacity KVA by module, and total capacity KVA
7	The asset performance report information	Daily report and monthly report of asset performance

The Designing

The system design of the SIPERITA application used the User Case Diagram, Activity Diagram, Sequence Diagram, and Class Diagram [14].

The Use Case Diagram

The application SIPERITA has two actors, the administrators, and the executors. The administrator is the user who operates the server and the executor is the user who becomes a client and sends data to the server. Figure 2 is the description of the SIPERITA application.

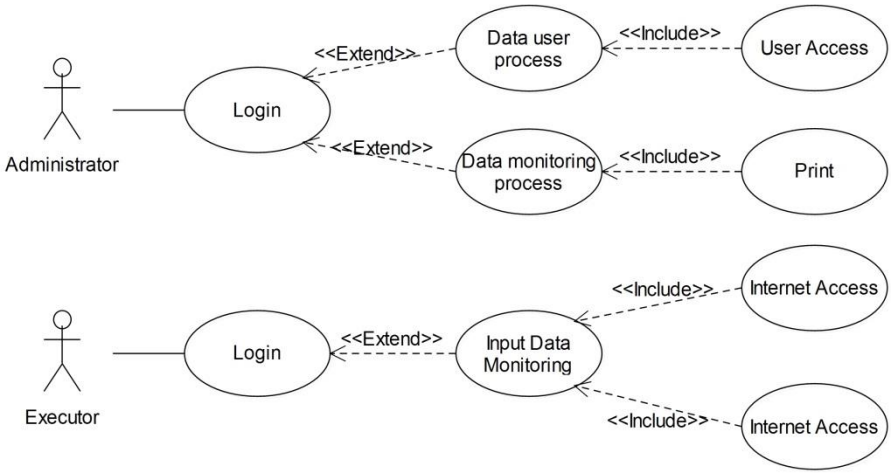


Figure 2. Use case diagram

The Administrator and the Executor users have the facility to enter into their respective systems. The Administrator user has the authority to create and change user data for system users. The Administrator user has the authority to view, modify and delete monitoring data and print the report. The Executor user can only add the asset monitoring data.

Activity Diagram

Figure 3 shows several data monitoring management that consists of adding, changing, and removing the asset performance monitoring data.

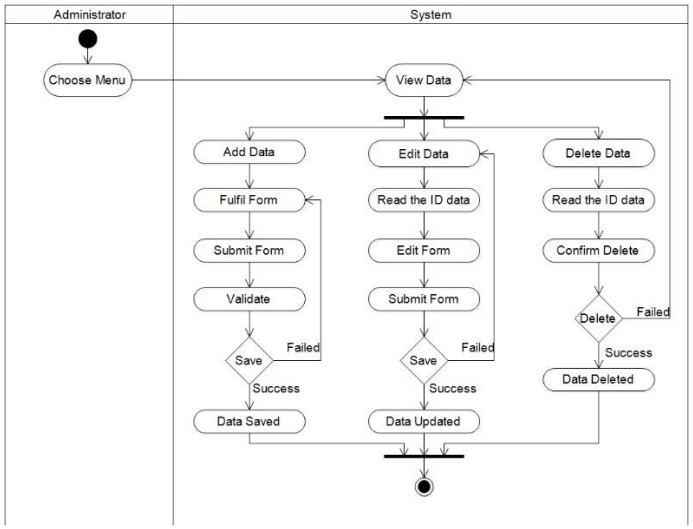


Figure 3. The activity of the data monitoring process

The Sequence Diagram

Figure 4 explains how the executor users added the data monitoring and the administrator users changed, removed, and printed the data monitoring.

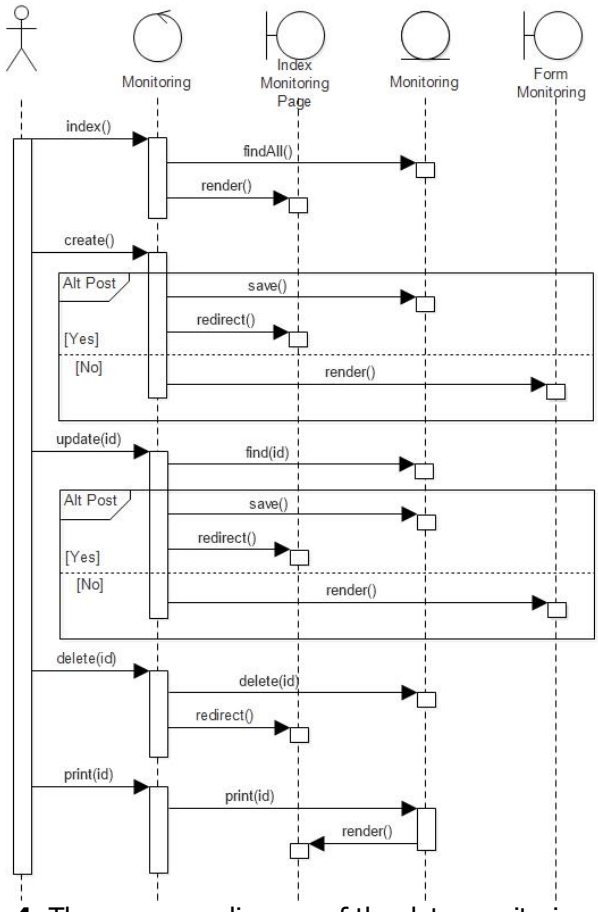


Figure 4. The sequence diagram of the data monitoring process

Class Diagram

This diagram is a connection of the classes and the explanations of each class detail within a design model of a system. It also shows the entity's regulations and responsibilities to determine the system's behavior. Figure 5 portrays the *Class Diagram* design of the SIPERITA application.

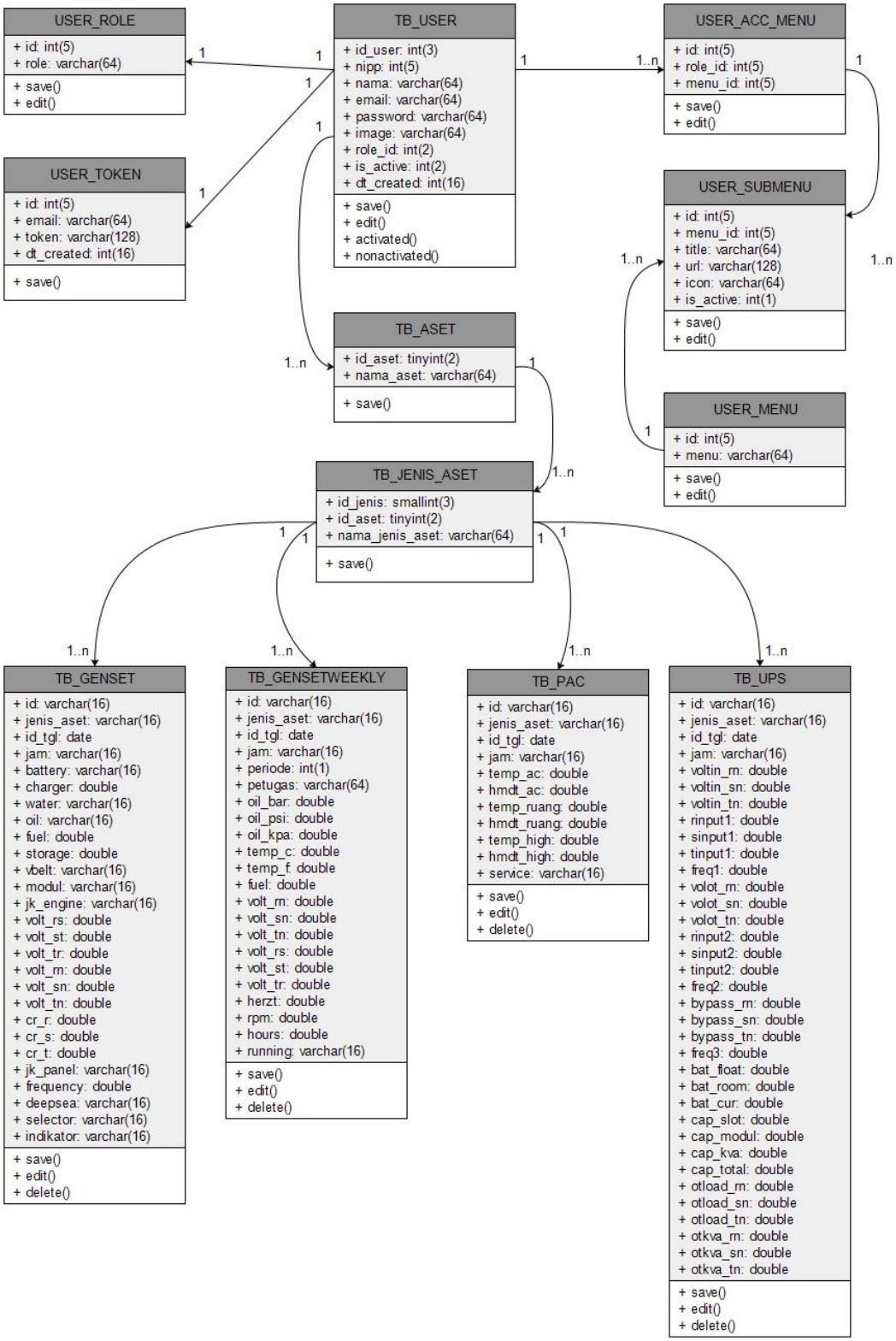


Figure 5. Class diagram

Implementation

The system implementation at this stage was a program coding creation process of the previously defined system plan results. The display of the program would be used by the users to interact with the established information system. During the program creation, the applied language programs were *PHP, framework CodeIgniter, MySQL as the database, and Microsoft Visual Code Studio as the editorial code* [15]. The applied system was used as the measurement or the test of the created program that was already established for further development [16].

Administrator Page

The initial login page display for all user levels is shown in figure 6. When the NIPP is not registered, it will display the warning. Figure 7 shows the main homepage of the administrator, on this page will appear a diagram of the results of asset monitoring. As illustrated in figure 7, four menus consist of Home, Report, Print of Report, and the User Menu.

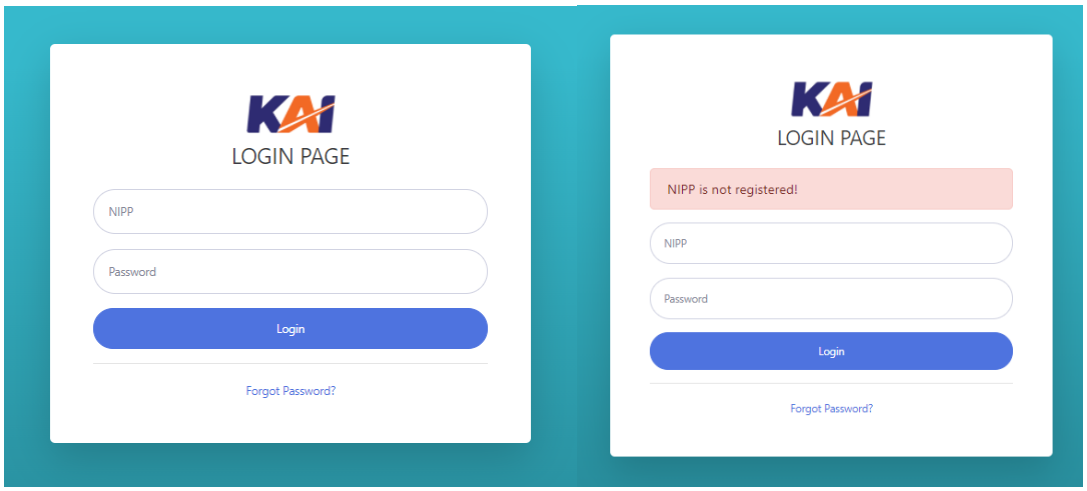


Figure 6. Login page

Figure 7. The administrator homepage

On the report page, the administrator user could see the data monitoring and asset based on the provided filters. They were inputted in the filter adjustment column as shown in figure 8. Meanwhile on this page also possible to edit or change one of the data monitoring. Figure 9 shows the page for editing data and it will display the alert when successfully edited. Figure 10

shows the filter type of the data that will be print out. This page was used to print the data monitoring stored in the *database.* The administrators had to select the type of the preferred assets to be printed. The report could be printed based on the dates, months, or years. The data user page showed the user lists in the SIPERITA application. As illustrated the figure 11, on this page the administrators could add the new users, non-activated users, and activate users.

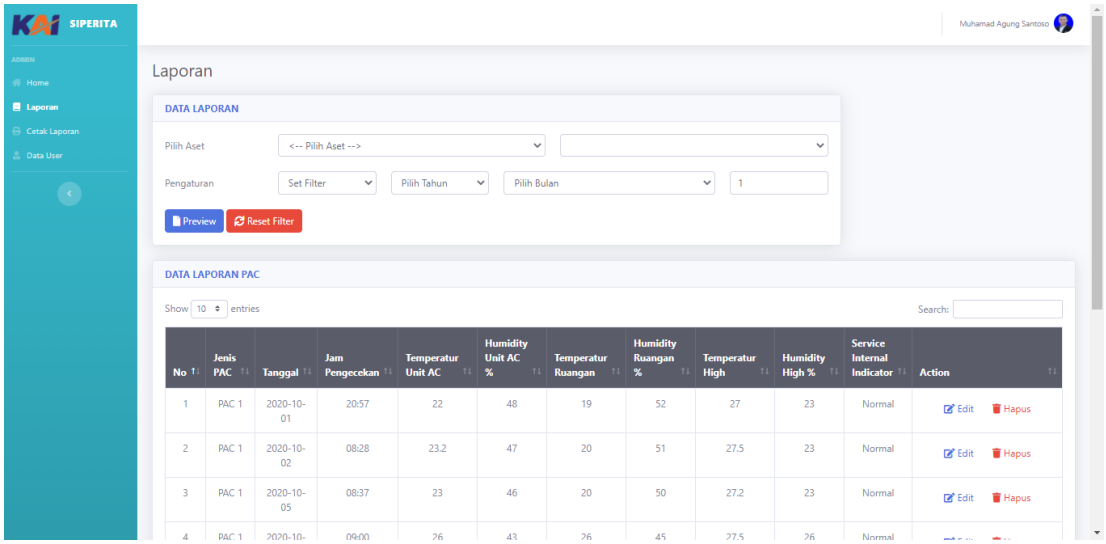


Figure 8. The report page

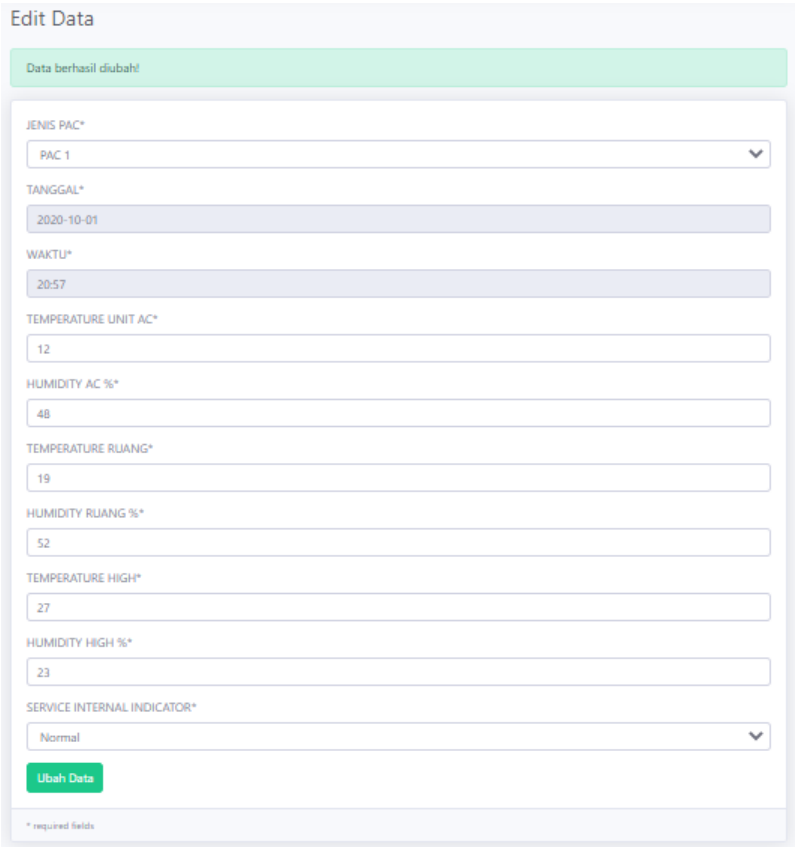


Figure 9. Edit data page

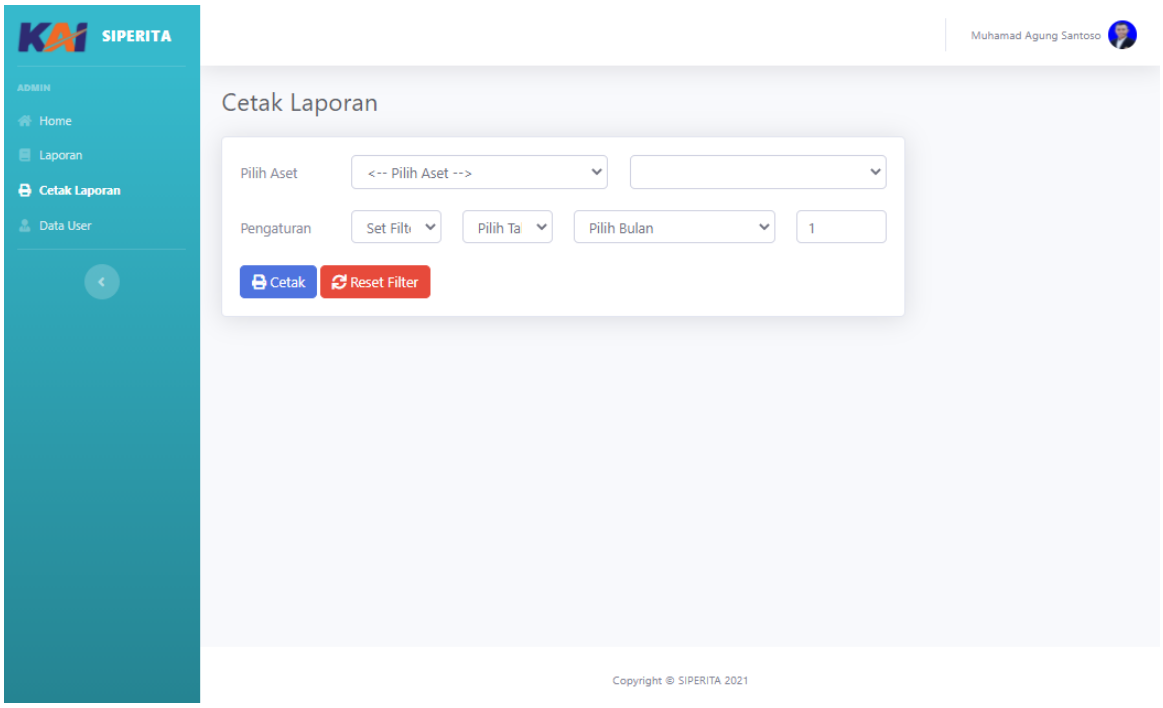


Figure 10. The data report printing page

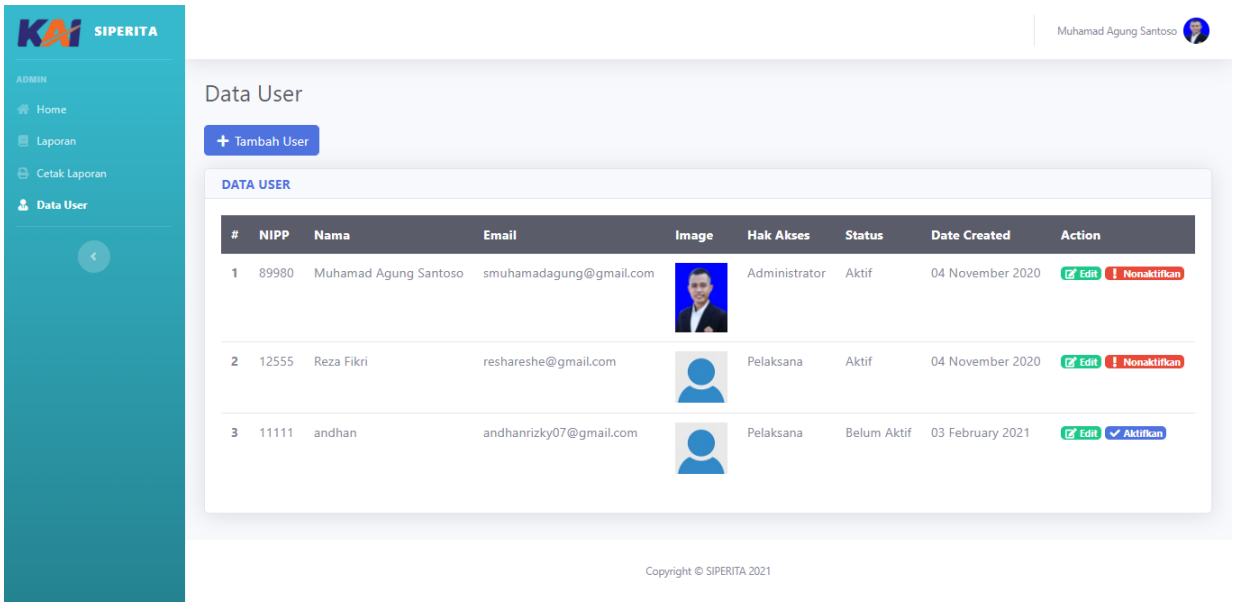


Figure 11. The data user page

Executor Page

Figure 12 is a page when someone successfully logged in as an executor and will display the homepage. On this page, they can view the adding data page has the asset lists that can be monitored by this information system. By selecting the monitored asset, the input form of data monitoring would appear based on the criteria of each asset as shown in the following figure 13. When the data is successfully added, it will display the alert.

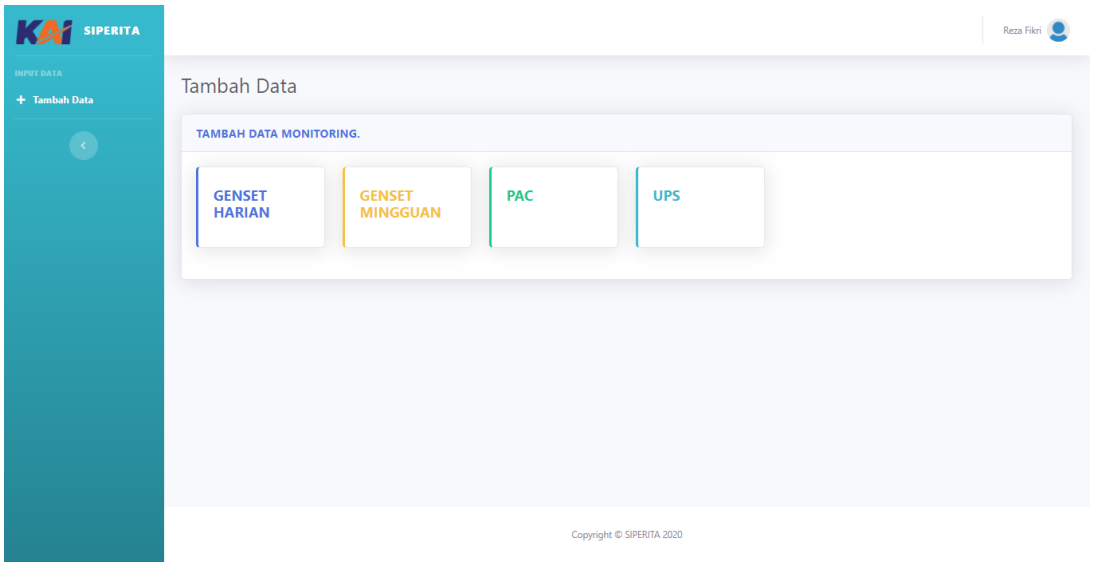


Figure 12. The executor homepage

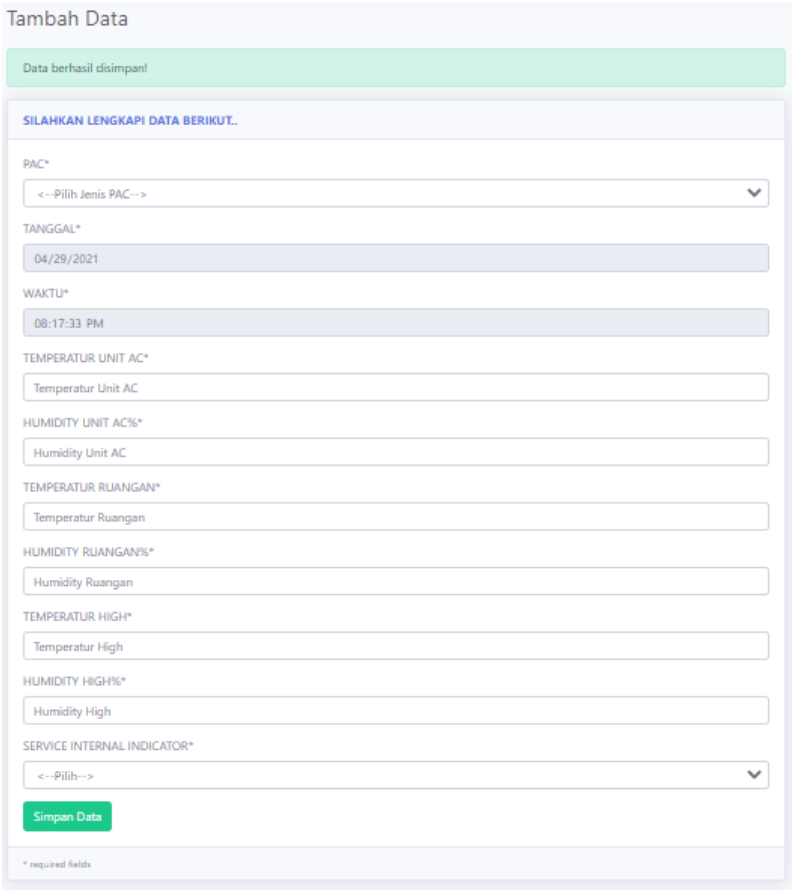


Figure 13. The adding data page

The Testing System

The test carried out for the respondents about the functionality, interface, and system utility was done with a questionnaire. It consisted of the respondents' statements and consents, then user satisfaction ratings are measured using the Likert Scale method [17]. The Siperita application users need to fill in nine questions and were filled in by ten respondents. The statements are shown in Table 3. In this study, there are five Likert scales used as shown in Table 4 and the interval value response categories shown in Table 5.

Table 3. Statements for the respondents

Number	Statements
1	The easiness to operate the application
2	The convenience to use the application in general
3	The application suitability to improve the asset monitoring process effectiveness
4	The time efficiency for asset monitoring process compared to the previous system
5	The application suitability to print the report
6	The accuracy of application utility based on the use
7	The application suitability based on the user access rights
8	The accuracy of the button and menu functions with the preferred targets
9	The accuracy of the displayed message (successful or fail)

Table 4. Likert scale

Variable	Response Categories
1	Strongly disagree
2	Disagree
3	Neither agree nor disagree
4	Agree
5	Strongly agree

Table 5. Classification based on interval value

Percentage	Response Categories
0% - 19.99%	Strongly disagree
20% - 39.99%	Disagree
40% - 59.99%	Neither agree nor disagree
60% - 79.99%	Agree
80% - 100%	Strongly agree

Based on the results of beta testing the users give 88.67%. So, it can be concluded that the respondents "totally agree" that the SIPERTA application was feasible to implement because of the ease in carrying out the monitoring process.

Conclusion

Based on the results and discussion, the implementation of the SIPERITA application, in the IT Operation unit of the PT Kereta Api Indonesia (Persero) could be used excellently. With the website-based application, it was because the application facilitated the monitoring activities by providing single-application access. Then the user administrator can see a diagram of the monitoring results on the home page so that if there is a problem with the asset, direct action can be taken. In addition, it could improve the effectiveness and efficiency in promoting the asset condition monitoring process and creating the reports.

Acknowledgment

The author would like to thank all employees of the IT Operation unit of PT Kereta Api Indonesia (Persero) that gave me chance and supporting this research. Thanks to the editorial and reviewing teams of Journal MATRIX for this scientific article publication process.

References

[1] M. A. Santoso, R. Fiati, and R. Nindyasasri, *Sistem Informasi Performance Reporting IT Asset (SIPERITA) Pada Unit IT Operation PT Kereta Api Indonesia (Persero)*. Kudus: Universitas Muria Kudus, 2020.

[2] G. T. Mardiani, "Sistem Monitoring Data Aset Dan Inventaris PT Telkom Cianjur Berbasis Web," *Komputa J. Ilm. Komput. dan Inform.*, vol. 2, no. 1, pp. 1–6, 2013.

[3] Suparni and Hadiyansyah, "Sistem Informasi Monitoring Inventory IT Aset (SIMONAS) Berbasis Web Pada PT. Metrocom Global Solusi Jakarta," *J. Penelit. Tek. Inform.*, vol. 3, no. 1, pp. 91–98, 2018.

[4] M. Tabrani and E. Pudjiarti, "Penerapan Metode Waterfall Pada Sistem Informasi Inventori PT Pangan Sehat Sejahtera," *J. Infokar*, vol. 1, no. 2, pp. 30–40, 2017.

[5] P. R. Sari and J. Devitra, "Analisis Dan Perancangan Sistem Informasi Manajemen Aset Pada Kantor Badan Meteorologi, Klimatologi Dan Geofisika (BMKG) Provinsi Jambi," *J. Manaj. Sist. Inf.*, vol. 2, no. 3, pp. 573–591, 2017.

[6] A. G. Sugima, *Manajemen Aset Pariwisata: Pelayanan Berkualitas agar Wisatawan puas dan Loyal*, 1st ed. Bandung: Guardaya Intimarta, 2013.

[7] A. D. Setiawati and I. M. D. J. Sulastra, "Rancang Bangun Sistem Informasi Inventory Berbasis Client Server pada PT Arpan Bali Utama di Denpasar," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 10, no. 3, pp. 94–102, 2020.

[8] A. I. Reyes, "Implementation of an Asset Management and Maintenance System for the College of Information and Communications Technology," *Int. J. Innov. Appl. Stud.*, vol. 12, no. 2, pp. 418–423, 2015.

[9] I. P. A. Pratama and I. G. T. Suryawan, "Rancang Bangun Sistem Monitoring Pelaksanaan dan Hasil Audit Mutu Internal." Jurnal S@CIES Volume 7, Nomor 2, pp. 74–81, 2017.

[10] S. Yaakub and J. Devitra, "Analisis Pemodelan Sistem Informasi Manajemen Aset Berbasis Web Pada Politeknik Jambi," *J. Manaj. Sist. Inf.*, vol. 2, no. 3, pp. 610–628, 2017.

[11] N. Husna and L. Fitria, "Sistem Informasi Persediaan Alat Tulis Kantor di Bpjs Kesehatan Cabang Langsa," *J. _COM (Jurnal Infromatika dan Teknol. Komput.*, vol. 01, no. 01, pp. 32–40, 2020.

[12] T. A. Cinderatama, A. F. Dianta, F. S. Efendi, and K. Eliyen, "Web and android-based application for monitoring tuberculosis (TB) patients in Kediri City," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 11, no. 1, pp. 1–10, 2021.

[13] H. Zenita and R. Fiati, "SDLC Model for Implementation of E-Blangko on Department of Population and Civil Registration," *Klik - Kumpul. J. Ilmu Komput.*, vol. 6, no. 1, p. 58, 2019.

[14] M. D. Irawan and S. A. Simargolang, "Implementasi E-Arsip Pada Program Studi Teknik Informatika," *J. Teknol. Inf.*, vol. 2, no. 1, p. 67, 2018.

[15] B. Raharjo, *Belajar Otodidak Framework Codeigniter*, vol. 5, no. 1. Bandung: Informatika Bandung, 2015.

[16] J. Dermawan and S. Hartini, "Implementasi Model Waterfall Pada Pengembangan Sistem Informasi Perhitungan Nilai Mata Pelajaran Berbasis Web Pada Sekolah Dasar Al-Azhar Syifa Budi Jatibening," *Paradigma*, vol. 19, no. 2, pp. 142–147, 2017.

[17] M. Hani'ah, Y. Kurniawan, and I. F. Rozi, "livE (onLine – java Exercise) java programming language learning system for lab and online test," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 11, no. 1, pp. 1–10, 2021.

Matrix: Jurnal Manajemen Teknologi dan Informatika
Volume 11 Issue 2 Year 2021 Pages 107-118
DOI: http://dx.doi.org/10.31940/matrix.v11i2.2433
URL: http://ojs.pnb.ac.id/index.php/matrix

Comparative analysis of support vector machine and k-nearest neighbors with a pyramidal histogram of the gradient for sign language detection

Imantoko [1]*, Arief Hermawan [2], Donny Avianto [3]

[1,2,3] *Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Indonesia*

*Corresponding Author: *imantoko@student.uty.ac.id*

Abstract: The communication method using sign language is very efficient considering that the speed of information delivery is closer to verbal communication (speaking) compared to writing or typing. Because of this, sign language is often used by people who are deaf, speech impaired, and normal people to communicate. To make sign language translation easier, a system is needed to translate symbols formed from hand movements (in the form of images) into text or sound. This study aims to compare performance such as accuracy and computation time of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) with Pyramidal Histogram of Gradient (PHOG) for feature extraction, to know which one is better at recognizing sign language. Yield, both combined methods PHOG-SVM and PHOG-KNN can recognize images from hand movements that form certain symbols. The system built using the SVM classification produces the highest accuracy of 82% at PHOG level 3, while the system built with the KNN classification produces the highest accuracy of 78% at PHOG level 2. The total computation time of the fastest training and testing by the SVM model is 236.53 seconds at PHOG level 3, while the KNN model is 78.27 seconds at PHOG level 3. In terms of accuracy, PHOG-SVM is better, but in terms of computation time, PHOG-KNN takes the place.

Keywords: Support Vector Machine, K-Nearest Neighbors, Pyramidal Histogram of Gradient, Analysis, Sign Language

History Article: Submitted 28 April 2021 | Revised 12 May 2021 | Accepted 25 June 2021

How to Cite: Imantoko, A. Hermawan, and D. Avianto, "Comparative Analysis Of Support Vector Machine And K-Nearest Neighbors with Pyramidal Histogram of Gradient For Sign Language Detection," *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 11, no. 2, pp.109-120, 2021.

Introduction

Sign language is one of several communication methods that can be used by deaf and mute persons as well as normal people to communicate. Communication is carried out by mimicking and hand movements that form symbols so that they interpret a letter or word. The method of communication using sign language is very efficient considering that the speed of information delivery is closer to verbal communication (speaking) compared to writing or typing. To make sign language translation easier, a system is needed to translate symbols formed from hand movements (in the form of images) into text or sound.

Image translation to recognize a certain pattern or shape can be done using the Pyramid Histogram of Oriented Gradient (PHOG) feature extraction method and the Support Vector Machine (SVM) classification method. Research on sign language recognition using these methods has been conducted [1] with an accuracy of up to 86%. Meanwhile, research [2] obtained an accuracy of up to 91.8% using a combination of the Haar Classifier and K-Nearest Neighbors (KNN) with different datasets. Therefore, this study aims to compare performance such as accuracy and computation time of SVM and KNN with PHOG for feature extraction and the same dataset like [1] to find out which one is better at recognizing sign language.

Research on sign language recognition has also been conducted by [3] using the Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) methods resulting in an average accuracy value of 60.58%. In addition, [4] also conducted research using the Generalized Learning Vector Quantization (GLVQ) method with an alpha value of 0.9 resulting in the

highest accuracy of 71.37%. Another study by [5] used the Hebb Rule to recognize static sign language and resulted in an accuracy of 80.37%.

Research on letter pattern recognition apart from sign language has also been carried out by [6] using Adaline and concluded that Adaline is effective to be applied in the development of letter pattern recognition applications if the reduced alpha value is 0.05, the enlarged tolerance value is 0.1 and uses bipolar activation function. Research by [7] used an artificial neural network (ANN) to recognize the character pattern of the Karo script. The ANN method studied is perceptron and can recognize up to 100%.

The SVM and KNN classifications have been studied by [8] using Partial Least Square (PLS) as dimension reduction, SVM, and KNN as classifications. This study produced the highest accuracy of 98.54% on leukemia data with PLS-KNN, 100% on lung data with KNN, 66.52% on breast data with PLS-KNN, and 85.60% on colon data with PLS-SVM.

Research on the introduction of tajwid law has been carried out by [9] using the KNN with Local Mean which has succeeded in achieving the highest number of 96.43%.

Another research about image recognition was also conducted by [10], but this study focused on license plate recognition using Object Character Recognition (OCR). The success rate for recognizable license plates was 75%.

Methodology

Pyramidal Histogram of Gradient (PHOG)

The way the PHOG method works is to do HOG according to the size of the cells and the orientation bin then combine them into 1 feature like the formula [11].

Dataset

The dataset used is an open-source dataset from [12]. Consists of a collection of near-infrared (near-infrared) images and skeletal information obtained from the Leap Motion sensor. The composition of the dataset includes 16 types of hand movements including palm, L, fist-moved, down, index, ok, palm m, C, heavy, hang, two, three, four, five, palm u, and up from 15 subjects (five male and 10 female). This dataset has a total of 13000 hand motion infrared images. However, based on [1] the types of hand movements used are palm, L, down, index, ok, and C. Out of 13,000 images, [1] states that only 3800 images are used, with the division of 3300 images for training data and 500 test data. However, in the discussion in [1], only 486 images were used, so that the total images used were 3786 images with random sampling. The distribution of the number of images for each type of hand movement can be seen in tables 1 and 2.

Table 1. Training data

Class	Subtotal
Palm	550
L	550
Down	550
Index	550
Ok	550
C	550
Total	3300

Table 1 shows the six types of training data hand movements (classes). Palm as many as 550 images, L as many as 550 images, Down as many as 550 images, Index as many as 550 images, Ok as many as 550 images, and C as many as 550 images. The total of images for training is 3300 images.

Table 2. Testing data

Class	Subtotal
Palm	80
L	80
Down	86
Index	70
Ok	90
C	80
Total	486

Table 2 shows the six types of test data hand movements (classes). Palm as many as 80 images, L as many as 80 images, Down as many as 86 images, Index as many as 70 images, Ok as many as 90 images, and C as many as 80 images. The total number of images for training is 486 images.

System Design

This study aims to compare two classification methods, namely SVM and KNN to find out which one is better at recognizing sign language in terms of accuracy and computation time. The feature extraction used for the comparison of the two methods is PHOG levels 1, 2, and 3. The three main stages in this research are Pre-processing & Feature Extraction (PPFE), Training, and Testing.

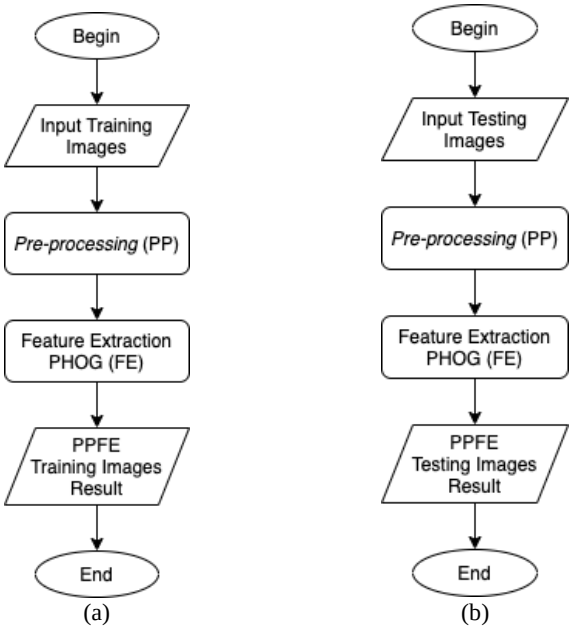


Figure 1. (a) PPFE training data stage, (b) PPFE test data stage

Figure 1 shows the PPFE stage for training data (a) and test data (b). Overall, the process for both data is the same, namely through pre-processing and extraction of PHOG features. The output of this stage is training data and test data that have been pre-processed and feature extraction. The output training data is used for the SVM and KNN Training stage, while the test data is used for the SVM and KNN Testing stage.

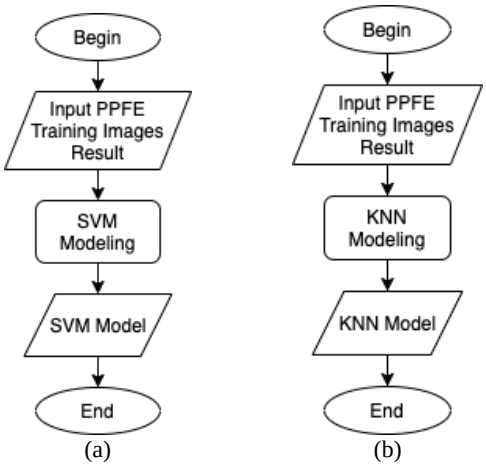


Figure 2. (a) SVM training data stage, (b) KNN training data stage

Figure 2 shows the training stage for SVM (a) and KNN (b). The input of this stage is the PPFE output training data. The output of the PPFE training data processed with SVM and KNN modeling. Each of them produces the SVM model and the KNN model that are used for the Testing stage.

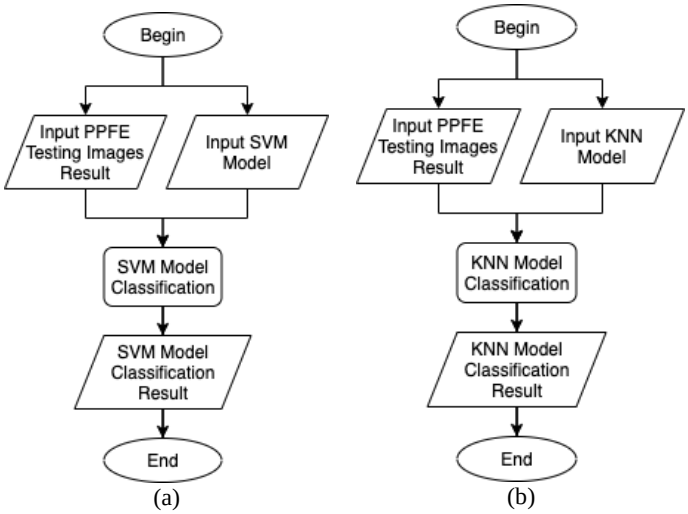


Figure 3. (a) SVM testing data stage, (b) KNN testing data stage

Figure 3 shows the testing stage of the SVM (a) and KNN (b) models. At this stage, the input is the PPFE output test data and classification model. The model classifies the PPFE test data output and produces the predictive results of the test data image. The result presented as Confusion Matrix, Precision, Recall, F1-Score, Accuracy, and Computation Time.

Pre-processing

The image from [12] varies in size from 408 x 264 to 420 x 273, so it is necessary to equalize the image size. All image sizes are converted to 408 x 264 using the crop method. When viewed from the largest size, which is 420 x 273, the reduction that occurs is only 9 to 12 pixels.

Feature Extraction

After the image pre-processing, the characteristics of the image are extracted using the PHOG method level 1, 2, and 3 for the training image and the test image. Each PHOG level applies HOG with a bin of 9, pixels per cell 8 x 8, and cells per block 2 x 2.

Training

Training or modeling using the SVM and KNN methods on 3300 training data that has passed the pre-processing and feature extraction processes. SVM uses a polynomial kernel configuration, degree 3, and tolerance of 0.00001 while KNN uses a number of neighbors of 3 (k = 3), the configuration of each method applied to all PHOG levels. The result of this training is that there are 3 SVM models and KNN models each following each PHOG level.

Testing

Testing was carried out with the SVM and KNN models on 486 test data that had passed the pre-processing and feature extraction processes. The test data of the PHOG level 1 feature extraction results were tested with the SVM model and the KNN level 1 model, and so on. The metric used for evaluating the recognition performance is the confusion matrix (CM) presented in graphical form. From CM values, four different measurements have been collected.

Precision (P) is the ratio of true positive predictions compared to the overall positive predicted outcome. The precision formula is as follows:

$$P = \frac{tp}{tp + fp} \tag{1}$$

Recall (R) is the ratio of true positive predictions compared to the total number of true positive data. The recall is calculated through:

$$R = \frac{tp}{tp + fn} \tag{2}$$

F1-score (F) is a weighted comparison of the average precision and recall. F1-score is calculated as:

$$F = 2 \times \frac{P \times R}{P + R} \tag{3}$$

Accuracy (A) is the ratio of Correct predictions (positive and negative) to the overall data. The accuracy formula is as follows:

$$A = \frac{tp + tn}{tp + tn + fp + fn} \tag{4}$$

Where *tp* represents true positive samples, *tn* the true negative samples, *fp* the false-positive samples, and *fn* the false-negative samples [13].

Results and Analysis

Correct Prediction Graph

The results of the SVM Confusion Matrix model for each PHOG level in [1] represents in graphical form can be seen in Figure 4.

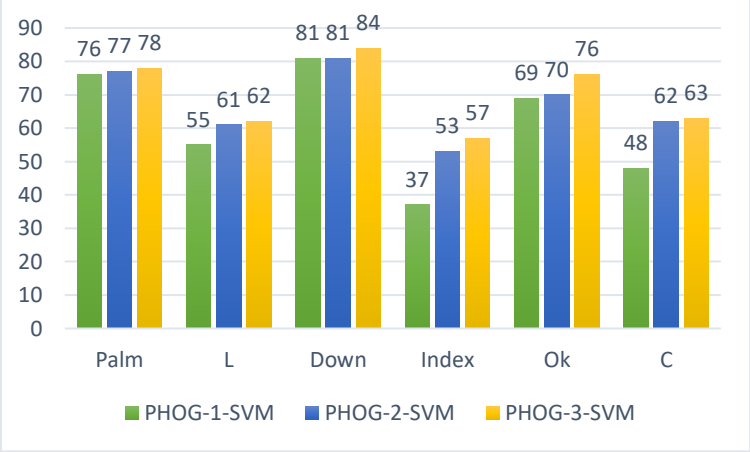


Figure 4. Correct prediction graph PHOG-SVM [1]

Figure 4 shows the number of correct predictions for each class using PHOG Level 1-SVM, PHOG Level 2-SVM, and PHOG Level 3-SVM, respectively in the study [1]. Palm got a correct prediction of 76, 77, and 78. L got a correct prediction of 55, 61, and 62. Down got a correct prediction of 81, 81, and 84. The index got a correct prediction of 37, 53, and 57. Ok got a correct prediction of 69, 70, and 76. C got a correct prediction of 48, 62, and 63.

Meanwhile, the results of the SVM Confusion Matrix model for each PHOG level in this research represents in graphical form can be seen in figure 5.

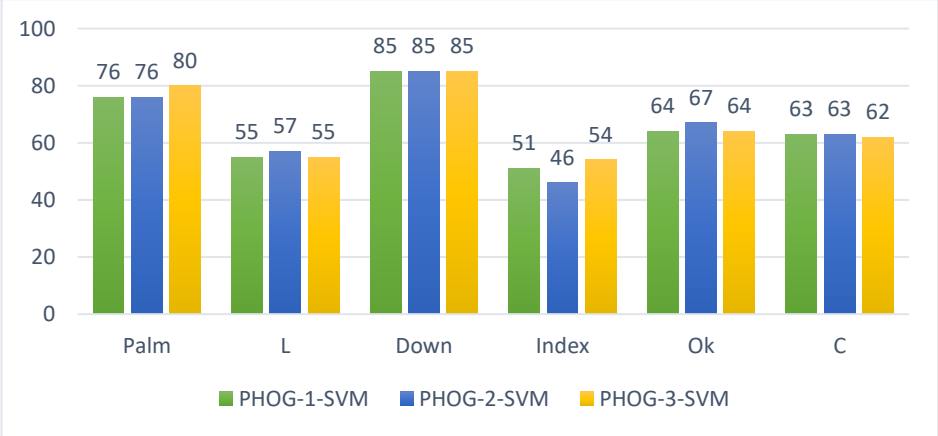


Figure 5. Correct prediction graph PHOG-SVM

Figure 5 shows the number of correct predictions for each class using PHOG Level 1-SVM, PHOG Level 2-SVM, and PHOG Level 3-SVM, respectively in this study. Palm got a correct prediction of 76, 76, and 80 out of 80. L got a correct prediction of 55, 57, and 55 out of 80. Down got a correct prediction of 85, 85, and 85 out of 86. Index got a correct prediction of 51, 46, and 54 out of 70. Ok got a correct prediction of 64, 67, and 64 out of 90. C got a correct prediction of 63, 63, and 62 out of 80.

In the other hand, the results of the KNN Confusion Matrix model for each PHOG level in this research represents in graphical form can be seen in figure 6.

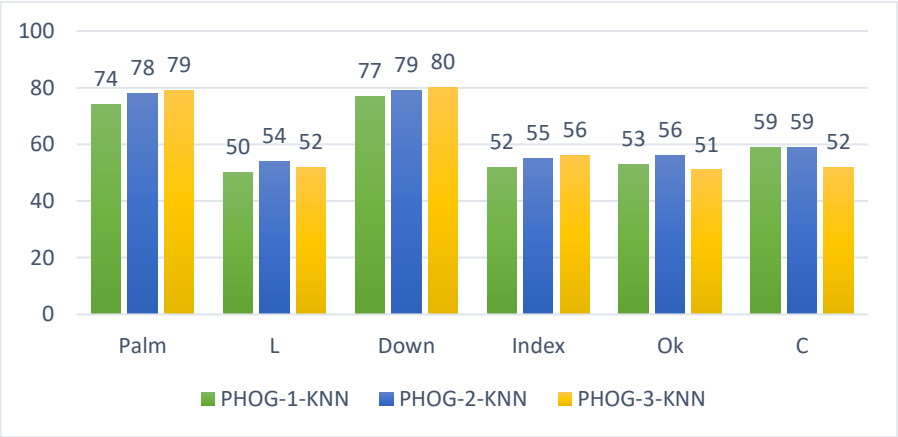


Figure 6. Correct prediction graph PHOG-KNN

Figure 6 shows the number of correct predictions for each class using PHOG Level 1-KNN, PHOG Level 2-KNN, and PHOG Level 3-KNN, respectively in this study. Palm got a correct prediction of 74, 78, and 79 out of 80. L got a correct prediction of 50, 54, and 52 out of 80. Down got a correct prediction of 77, 79, and 80 out of 86. Index got a correct prediction of 52, 55, and 56 out of 70. Ok got a correct prediction of 53, 56, and 51 out of 90. C got a correct prediction of 59, 59, and 52 out of 80.

Precision, Recall, F1-Score, and Accuracy

The results of the Precision, Recall, F1-Score, and Accuracy of the SVM model for each PHOG level in [1] can be seen in table 3.

Table 3. Precision, recall, F1-score, and accuracy PHOG-SVM [1]

Class	Model	Precision	Recall	F1-Score
Palm	PHOG-1-SVM	0,93	0,95	0,94
	PHOG-2-SVM	0,96	0,97	0,97
	PHOG-3-SVM	0,96	0,97	0,97
L	PHOG-1-SVM	0,62	0,69	0,65
	PHOG-2-SVM	0,82	0,78	0,79
	PHOG-3-SVM	0,82	0,78	0,79
Down	PHOG-1-SVM	0,92	0,94	0,93
	PHOG-2-SVM	0,98	0,98	0,98
	PHOG-3-SVM	0,98	0,98	0,98
Index	PHOG-1-SVM	0,70	0,53	0,60
	PHOG-2-SVM	0,73	0,82	0,77
	PHOG-3-SVM	0,73	0,81	0,77
Ok	PHOG-1-SVM	0,67	0,77	0,72
	PHOG-2-SVM	0,84	0,84	0,84
	PHOG-3-SVM	0,84	0,84	0,84
C	PHOG-1-SVM	0,67	0,60	0,63
	PHOG-2-SVM	0,75	0,78	0,76
	PHOG-3-SVM	0,85	0,79	0,82
Accuracy	PHOG-1-SVM	-	-	0,75
	PHOG-2-SVM	-	-	0,83
	PHOG-3-SVM	-	-	0,86

Table 3 shows the precision, recall, f1-score for each class using PHOG-1-SVM, PHOG-2-SVM, PHOG-3-SVM respectively and accuracy PHOG-1-SVM, PHOG-2-SVM, PHOG-3-SVM respectively in study [1]. Palm obtained a precision of 0.93, 0.96, and 0.96; recall of 0.95, 0.97, and

0.97; f1-score of 0.94, 0.97, and 0.97. L obtained a precision of 0.62, 0.82, and 0.82; recall of 0.69, 0.78, and 0.78; f1-score of 0.65, 0.79, and 0.79. Down obtained a precision of 0.92, 0.98, and 0.98; recall of 0.94, 0.98, and 0.98; f1-score of 0.93, 0.98, and 0.98. Index obtained a precision of 0.70, 0.73, and 0.73; recall of 0.53, 0.82, and 0.81; f1-score of 0.60, 0.77 and 0.77. Ok obtained a precision of 0.67, 0.84, and 0.84; recall of 0.77, 0.84, and 0.84; f1-score of 0.60, 0.77, and 0.77. C obtained a precision of 0.67, 0.75, and 0.85; recall of 0.60, 0.78, and 0.79; f1-score of 0.63, 0.76, and 0.82. The accuracy obtained is 0.75, 0.83, and 0.86.

The results of the Precision, Recall, F1-Score, and Accuracy of the SVM model for each PHOG level in the research conducted can be seen in table 4.

Table 4. Precision, recall, F1-score, and accuracy PHOG-SVM

Class	Model	Precision	Recall	F1-Score
Palm	PHOG-1-SVM	0,99	0,95	0,97
	PHOG-2-SVM	0,99	0,95	0,97
	PHOG-3-SVM	0,99	1,00	0,99
L	PHOG-1-SVM	0,79	0,69	0,73
	PHOG-2-SVM	0,77	0,71	0,74
	PHOG-3-SVM	0,86	0,69	0,76
Down	PHOG-1-SVM	0,96	0,99	0,97
	PHOG-2-SVM	0,96	0,99	0,97
	PHOG-3-SVM	1,00	0,99	0,99
Index	PHOG-1-SVM	0,61	0,61	0,67
	PHOG-2-SVM	0,62	0,66	0,64
	PHOG-3-SVM	0,62	0,77	0,69
Ok	PHOG-1-SVM	0,71	0,71	0,71
	PHOG-2-SVM	0,72	0,74	0,73
	PHOG-3-SVM	0,72	0,71	0,72
C	PHOG-1-SVM	0,82	0,79	0,80
	PHOG-2-SVM	0,80	0,79	0,79
	PHOG-3-SVM	0,78	0,78	0,78
Accuracy	PHOG-1-SVM	-	-	0,81
	PHOG-2-SVM	-	-	0,81
	PHOG-3-SVM	-	-	0,82

Table 4 shows the precision, recall, f1-score for each class using PHOG-1-SVM, PHOG-2-SVM, PHOG-3-SVM respectively and accuracy PHOG-1-SVM, PHOG-2-SVM, PHOG-3-SVM respectively in this study. Palm obtained a precision of 0.99, 0.99, and 0.99; recall of 0.95, 0.95, and 1.00; f1-score of 0.97, 0.97, and 0.99. L obtained a precision of 0.79, 0.77, and 0.86; recall of 0.69, 0.71, and 0.69; f1-score of 0.73, 0.74, and 0.76. Down obtained a precision of 0.96, 0.96, and 1.00; recall of 0.99, 0.99, and 0.99; f1-score of 0.97, 0.97, and 0.99. Index obtained a precision of 0.61, 0.62, and 0.62; recall of 0.61, 0.66, and 0.77; f1-score of 0.67, 0.64 and 0.69. Ok obtained a precision of 0.71, 0.72, and 0.72; recall of 0.71, 0.74, and 0.71; f1-score of 0.71, 0.73, and 0.72. C obtained a precision of 0.82, 0.80, and 0.78; recall of 0.79, 0.79, and 0.78; f1-score of 0.80, 0.79, and 0.78. The accuracy obtained is 0.81, 0.81, and 0.82.

The results of the Precision, Recall, F1-Score, and Accuracy of the KNN model for each PHOG level in the research conducted can be seen in table 5.

Table 5. Precision, recall, F1-score, and accuracy PHOG-KNN

Class	Model	Precision	Recall	F1-Score
Palm	PHOG-1-KNN	0,85	0,93	0,89
	PHOG-2-KNN	0,87	0,97	0,92
	PHOG-3-KNN	0,91	0,99	0,95
L	PHOG-1-KNN	0,67	0,62	0,65
	PHOG-2-KNN	0,70	0,68	0,69
	PHOG-3-KNN	0,68	0,65	0,66
Down	PHOG-1-KNN	0,88	0,90	0,89
	PHOG-2-KNN	0,95	0,92	0,93
	PHOG-3-KNN	0,95	0,93	0,94
Index	PHOG-1-KNN	0,65	0,74	0,69
	PHOG-2-KNN	0,70	0,79	0,74
	PHOG-3-KNN	0,60	0,80	0,68
Ok	PHOG-1-KNN	0,72	0,59	0,65
	PHOG-2-KNN	0,72	0,62	0,67
	PHOG-3-KNN	0,72	0,57	0,63
C	PHOG-1-KNN	0,72	0,74	0,73
	PHOG-2-KNN	0,75	0,74	0,74
	PHOG-3-KNN	0,71	0,65	0,68
Accuracy	PHOG-1-KNN	-	-	0,75
	PHOG-2-KNN	-	-	0,78
	PHOG-3-KNN	-	-	0,76

Table 5 shows the precision, recall, f1-score for each class using PHOG-1-KNN, PHOG-2-KNN, PHOG-3-KNN respectively and accuracy PHOG-1-KNN, PHOG-2-KNN, PHOG-3-KNN respectively in this study. Palm obtained a precision of 0.85, 0.87, and 0.91; recall of 0.93, 0.97, and 0.99; f1-score of 0.89, 0.92, and 0.95. L obtained a precision of 0.67, 0.70, and 0.68; recall of 0.62, 0.68, and 0.65; f1-score of 0.65, 0.69, and 0.66. Down obtained a precision of 0.88, 0.95, and 0.95; recall of 0.90, 0.92, and 0.93; f1-score of 0.89, 0.93, and 0.94. Index obtained a precision of 0.65, 0.70, and 0.60; recall of 0.74, 0.79, and 0.80; f1-score of 0.69, 0.74 and 0.68. Ok obtained a precision of 0.72, 0.72, and 0.72; recall of 0.59, 0.62, and 0.57; f1-score of 0.65, 0.67, and 0.63. C obtained a precision of 0.72, 0.75, and 0.71; recall of 0.74, 0.74, and 0.65; f1-score of 0.73, 0.74, and 0.68. The accuracy obtained is 0.75, 0.78, and 0.76.

Computation Time

This study also compared the computation time required by the SVM and KNN methods for training 3300 images and testing 486 images. Computation time is calculated in seconds. The hardware used is the Macbook Pro 2019. The time required for each method to PHOG level n can be seen in Table 6. However, research [1] did not take into account the computation time, so table 6 only shows the computation time carried out in this study only.

Table 6. Computation time in seconds

Model	Training	Testing	Total
PHOG-1-SVM	353,26	35,85	389,11
PHOG-2-SVM	280,26	29,86	310,12
PHOG-3-SVM	212,64	23,89	236,53
PHOG-1-KNN	3,41	113,84	117,25
PHOG-2-KNN	1,10	98,53	99,63
PHOG-3-KNN	0,59	77,68	78,27

Table 6 shows the computation time of each model at the training stage (3300 images), testing stage (486 images), and the total time for the two stages in seconds. The PHOG-1-SVM model requires a computational training time of 353.26 seconds, testing 35.85 seconds, and a

total of 389.11 seconds. The PHOG-2-SVM model requires a computational training time of 280.26 seconds, testing 29.86 seconds, and a total of 310.12 seconds. The PHOG-3-SVM model requires a computational training time of 212.64 seconds, testing 23.89 seconds, and a total of 236.53 seconds. The PHOG-1-KNN model requires a computational training time of 3.41 seconds, testing 113.84 seconds, and a total of 117.25 seconds. The PHOG-2-KNN model requires a training computation time of 1.10 seconds, testing 98.53 seconds, and a total of 99.63 seconds. The PHOG-3-KNN model requires a computational training time of 0.59 seconds, testing 77.68 seconds, and a total of 78.27 seconds.

Analysis

Figure 4 and Figure 5 are expected to have similar results, considering that these figures use the same dataset source, pre-processing method, feature extraction, and classification. However, the results are quite different, in Figure 4 the total correct prediction is 366, 404, and 420, respectively, while Figure 5 are 394, 394, and 400 respectively. Quite a big increase occurs in Figure 4, but only a little in Figure 5.

On the other hand, Figure 4 and Figure 5 have similarities, namely prediction or testing in class L and Index has poor results, followed by classes Ok and C, while the best prediction results are in the Palm and Down classes.

Switch to Precision, Recall, F1-Score, and accuracy results. A clear difference occurs in table 3 as a whole experience an increase in the precision, recall, f1-score, and accuracy results. A clear difference can be seen in the accuracy of Table 3, respectively, namely 75%, 83%, and 86%, while Table 4 hardly experience a significant increase, namely 81%. 81%, and 82%.

The difference in the results of confusion matrix, precision, recall, f1-score, and accuracy is very possible due to two factors. The first factor is the selection of the image dataset from 13000, only 3786 images were selected, it is very possible that the image used in [1] is not the same as the research conducted. The second factor is [1] not explicitly mentioning the SVM configuration used, so in this study, the authors tried several configurations and the best results were obtained on the SVM configuration with a polynomial kernel, degree 3, and tolerance 0.00001.

This difference can still be tolerated considering the results of the two SVMs have not been missed too far and the research conducted has used the source dataset, pre-processing methods, feature extraction, and the same classification as previous studies. For the SVM and KNN comparisons to be valid, comparisons are only made to the SVM and KNN models built in this study.

Returning to the results of the confusion matrix, the SVM model in Figure 5 has better results than the KNN model in Figure 6 where the total correct predictions on PHOG-1-SVM, PHOG-2-SVM, and PHOG-3-SVM respectively were 394, 394, and 400, while the PHOG-1-KNN, PHOG-2-KNN, and PHOG-3-KNN were 365, 381, and 370 respectively. The SVM model experienced an increase in the number of correct predictions although not significant, in the KNN model there was an increase from PHOG-1-KNN to PHOG-2-KNN then get off at PHOG-3-KNN.

There are similarities to the results of PHOG-1-SVM, PHOG-2-SVM, PHOG-3-SVM, PHOG-1-KNN, PHOG-2-KNN, and PHOG-3-KNN, namely predictions in class L and Index have poor results, followed by classes Ok and C, while the results best predictions for the Palm and Down classes. If we review the L and Index images, in plain view there are indeed similarities in several images as in Figures 7 and 8.

(a) (b)

Figure 7. (a) Sample image L, (b) Sample image index

Figure 7 shows the visible similarities between the L (a) image sample and the index (b) image sample.

(a) (b)

Figure 8. (a) Sample image Ok, (b) Sample image C

Figure 8 shows the visible similarity between the Ok image sample (a) and the image sample C (b).

Then the results of precision, recall, f1-score, and accuracy from table 4 show that PHOG-1-SVM, PHOG-2-SVM, and PHOG-3-SVM have increased insignificantly, even though there are classes that have low precision, recall, and f1-scores such as L, Index, Ok, and C. Low f1-score results indicate that the model is difficult to distinguish from other classes. Low F1-score due to low precision and recall. The accuracy obtained by the PHOG-1-SVM and PHOG-2-SVM models is 81% and PHOG-3-SVM is 82%. Meanwhile, based on table 5 the PHOG-1-KNN, PHOG-2-KNN, and PHOG-3-KNN models have lower precision, recall, f1-score, and accuracy results than the SVM model. The accuracy obtained by the PHOG-1-KNN model was 75%, PHOG-2-KNN increased to 78%, and there was a decrease in PHOG-3-KNN to 76%. These results indicate that at each PHOG level the SVM model is better than the KNN model.

When viewed from the aspect of computation time based on table 6, the KNN model is superior to the SVM model in the training process. For example, the PHOG-1-KNN model training process takes 3.41 seconds, while the PHOG-1-SVM takes 353.26 seconds. The SVM model is superior to the KNN model during the testing process. It can be seen that the PHOG-1-SVM model takes 35.85 seconds, while the PHOG-1-KNN takes 113.84 seconds. Overall, the KNN model is far superior to the SVM model when viewed from the total computation time required. The PHOG-1-KNN model has a total of 117.25 seconds, while the PHOG-1-SVM Model takes 389.11 seconds. Based on table 6, the total computation time of the KNN model is 78.27 seconds, while the SVM model is 236.53 seconds at PHOG level 3. Looking from another point of view, the SVM and KNN models have the same computation time pattern where the higher the PHOG level, the faster the computation time, so it can be concluded that PHOG level affects the computation time of the two models.

Conclusion

Experiments in this study with previous research there are differences in the results of confusion matrix, precision, recall, f1-score, and accuracy. The first factor that causes the difference in the results of the two studies is the selection of the image dataset from 13000, only 3786 images were selected, it is very possible that the image used in the previous study was not the same as the research carried out. The second factor is that previous research did not explicitly mention the SVM configuration used, so in this study, the authors tried several configurations and the best results were obtained on the SVM configuration with a polynomial kernel, degree 3, and tolerance 0.00001.

Based on its accuracy, the SVM method is superior to the KNN method at all PHOG levels. PHOG1-SVM produces an accuracy of 81%, while PHOG-1-KNN is only 75%. The PHOG-2-SVM produces 81% accuracy, while the PHOG-2-KNN has an accuracy of 78%. PHOG-3-SVM experienced an increase in accuracy to 82%, while PHOG-3-KNN experienced a decrease in accuracy to 76%. Based on the total computation time required for training and testing, the KNN method is superior to the SVM method at all PHOG levels. The KNN method requires the fastest

total computation time of 78.27 seconds at PHOG level 3, while the SVM method is 236.53 seconds at PHOG level 3.

References

[1] R. Novanda, K. N. Ramadhan, and P. E. Yunanto, "Pengenalan bentuk tangan dengan ekstraksi ciri pyramid histogram of oriented gradient (PHOG) dan klasifikasi Support Vector Machine (SVM)," in *e-Proceeding of Engineering*, 2020, vol. 7, no. 2, pp. 8223–8233.

[2] R. I. Borman, B. Priyopradono, and A. R. Syah, "Klasifikasi objek kode tangan pada pengenalan isyarat alphabet bahasa isyarat indonesia (BISINDO)," in *Prosiding Seminar Nasional Informatika dan Aplikasinya (SNIA)*, 2017.

[3] D. Yolanda, K. Gunadi, and E. Setyati, "Pengenalan alfabet bahasa isyarat tangan secara real-time dengan menggunakan metode Convolutional Neural Network dan Recurrent Neural Network," *J. Infra*, vol. 8, no. 1, 2020.

[4] D. Gustiar , S. H. Sitorus , and D. M. Midyanti, "Penerjemahan bahasa isyarat menggunakan metode Generalized Learning Vector Quantization (GLVQ)," *Coding J. Komput. dan Apl.*, vol. 8, no. 3, pp. 1-8, 2020.

[5] D. Setiawan and A. Z. Falani, "Pemanfaatan Artificial Neural Network dengan metode Hebb Rule untuk pengenalan bahasa isyarat indonesia statis," *j. spirit*, vol. 12, no. 1, pp. 9–15, 2020.

[6] B. Etikasari and T. D. Puspitasari, "Pengenalan pola huruf dengan menggunakan Algoritma Adaline Bety," *J. Mnemon.*, vol. 2, no. 1, pp. 12–16, 2019.

[7] D. R. Yani, "Penerapan jaringan syaraf tiruan dalam pengenalan huruf aksara suku karo dengan metode Perceptron," *J. Inf. Sist. Res.*, vol. 1, no. 3, pp. 109–114, 2020.

[8] S. A. Naufal, A. Adiwijaya, and W. Astuti, "Analisis perbandingan klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk deteksi kanker dengan data Microarray," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, pp. 162-168, 2014.

[9] I. Fathurrahman and A. Kurniawardhani, "Pengenalan hukum tajwid pada citra Al-Quran : kajian pustaka," *J. Karya untuk Masy.*, vol. 2, no. 1, 2020.

[10] W. S. Ismail, P. W. Purnawan, I. Riyanto, and N. Nazori, "Sistem perekaman pelat nomor mobil pada palang pintu parkir menggunakan web kamera dan mikrokontroler," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 10, no. 3, pp. 103–112, 2020.

[11] F. Sthevanie, P. I. Aristya, and K. N. Ramadhani, "Pengenalan aksara bali menggunakan metode pyramid histogram of oriented gradients," *Indo-Journal Comput.*, vol. 5, no. 1, pp. 73–84, 2020.

[12] T. Mantecón, "Multi-modal dataset for hand gesture recognition," 2018. https://www.kaggle.com/gti-upm/multimodhandgestrec (accessed Jan. 18, 2021).

[13] B. Raza, Y. J. Kumar, A. K. Malik, A. Anjum, and M. Faheem, "Performance prediction and adaptation for database management system workload using Case-Based Reasoning approach," *Inf. Syst.*, vol. 76, pp. 46–58, 2018.

POLITEKNIK NEGERI BALI

Redaksi Jurnal Matrix
Gedung P3M, Politeknik Negeri Bali
Bukit Jimbaran, PO BOX 1064 Tuban, Badung, Bali.
Phone: +62 361 701981, Fax: +62 361 701128
e-mail: p3mpoltekbali@pnb.ac.id
http://ojs.pnb.ac.id/index.php/matrix