e-ISSN : 2829-4149
p-ISSN : 2829-4157

COELITE

JOURNAL OF COMPUTER ENGINEERING, ELECTRONICS AND INFORMATION TECHNOLOGY

OJS :

Publisher :

The Education University

Vol. 02| Issue. 02 | Oct 2023

Volume 2, Issue 2, October 2023

e-ISSN : 2829-4149
p-ISSN : 2829-4157

Journal of
Computer Engineering, Electronics and Information Technology
Universitas Pendidikan Indonesia

Jl. Pendidikan No.15, Cibiru Wetan, Kec. Cileunyi, Kab. Bandung, Jawa Barat 40625
Phone: (022) 7801840, Fax: (022) 7830426
E-Mail: coelite@upi.edu, Website: https://ejournal.upi.edu/index.php/COELITE

Editor in Chief
Dr. Eng. Munawir, S.Kom., M.T., (Scopus ID 57208758022)
Expertise: Data Mining, Text Mining.
Universitas Pendidikan Indonesia, Indonesia.

Executive Editor
Anugrah Adiwilaga, S.ST., M.T., (Scopus ID 57194714935)
Expertise: Robotic, Control and Intelligent Systems, AI, IoT.
Universitas Pendidikan Indonesia, Indonesia.

Editorial Board
Dr. Eng. Dinda Pramanta, M.Eng.
Expertise: Artificial Intelligence, FPGA, Robot and Neural Network.
Associate Professor at Kyushu Institute of Information sciences Japan.

Wirmanto Suteddy, S.T., M.T.,
Expertise: Embedded System, IoT, AI, VLSI, Computer Vision.
Universitas Pendidikan Indonesia, Indonesia.

Dr. Eng. Mia Rizkinia, S.T.,M.T.
Expertise: Optoelectrotechnique and Remote Sensing, Vision, Image and Graphics Processing.

Editorial Assistant
Deden Pradeka, S.T., M.Kom., (Scopus ID 57200731472)
Expertise: Software Engineering Development, Mobile Development, Cryptography, Steganography.
Universitas Pendidikan Indonesia, Indonesia.

Reviewers

Dr. Eng. Md. Zaman Molla
Expertise: Solar Cell.
Associate Professor of Physics at Ahsanullah University of Science and Technology Bangladesh.

Dr. Eng. Alfian Akbar Gozali, S.T., M.T
Expertise: Data Analytics, Machine Learning, Optimization Algorithm, Mobile Development.
Telkom University, Indonesia.

Dr. Eng. Bobby Kurniawan, S.T., M.T
Expertise: Production system and control, mathematical modeling, scheduling metaheuristics, AI.
Universitas Sultan Ageng Tirtayasa, Indonesia.

Syamsul Rizal S.T, M.T., PhD.
Expertise: Image Processing, IoT,Faulty Node, Fault Detection.
Telkom University, Indonesia.

Devi Aprianti Rimadhani Agustini, S.Si, M.Si
Expertise: Modeling and Simulation.
Universitas Pendidikan Indonesia, Indonesia.

Muhammad Taufik Dwi Putra, S.Tr.Kom., M.T.I
Expertise: Computer vision, deep learning, Machine Learning.
Universitas Pendidikan Indonesia, Indonesia.

Ana Rahma Yuniarti, S.T., M.Eng., (SCOPUS ID 57192948351)
Expertise: Biomedical Engineering, Machine Learning, Artificial Inteligence.
Universitas Pendidikan Indonesia, Indonesia.

Volume 2, Issue 2, October 2023

e-ISSN : 2829-4149
p-ISSN : 2829-4157

published by **Universitas Pendidikan Indonesia.**

Journal of Computer Engineering, Electronics and Information Technology

COELITE

Volume 2, Issue 2, October 2023
Table of Contents

Application of the K-Nearest Neighbor Method in Determining Web-Based Early Childhood Development 67-74
Rahmat Hidayat

Sentiment Analysis of Twitter Users' Opinion Data Regarding the Use of ChatGPT in Education 75-88
Jezzy Putra Munggaran, Ahmad Ali Alhafidz, Maulana Taqy, Devi Aprianti Rimadhani Agustini, Munawir Munawir

Comparison of Machine Learning and Deep Learning Algorithms for Classification of Breast Cancer 89-98
Puji Ayuningtyas, Rahmawati Rahmawati, Akhmad Miftahusalam

Decision Support System For Selection Of Productive At Sahabat Sampulo Foundation Using The Profile Matching Method 99-110
Nia Nandara, Rahman Rahman, Adhy Rizaldy

Teacher and Student Attendance System at Noor Faqih Usman Foundation Based on RFID Integrated with Raspberry Pi 111-126
Dhimaz Purnama Adjhi, Mohamad Rizal Hanafi, Rastra Wardana Nanditama, Rifqi Alamsya, Hafidz Rizki Fahriza, Anugrah Adiwilaga

Indexed by.

Google Scholar GARUDA GARBA RUJUKAN DIGITAL sinta S4 Science and Technology Index

Volume 2, Issue 2, October 2023

e-ISSN : 2829-4149
p-ISSN : 2829-4157



Published by:
Universitas Pendidikan Indonesia
E-Mail: coelite@upi.edu
Website: https://ejournal.upi.edu/index.php/COELITE



Journal of Computer Engineering, Electronics and Information Technology (COELITE)

Journal homepage: https://ejournal.upi.edu/index.php/COELITE

Application of the K-Nearest Neighbor Method in Web-Based Applications to Determine Early Childhood Development

*Rahmat Hidayat**

Department of Information System, Universitas Nusa Putra, Indonesia.
Correspondence: E-mail: rahmat.hidayat@nusaputra.ac.id

ABSTRACT

Childhood is a period of rapid growth and development, so it is often termed as the golden period as well as the critical period. The golden period can be realized if at this time infants and children obtain suitable nutritional intake for optimal growth of flowers. But during this time Posyandu process is still done manually, so in providing the nutritional status information children must wait long. System that is able to assist in solving the problem is the Decision Support System (SPK) with the method K-Nearest Neighbor (KNN) that can produce information on the nutritional status of children that is the status of poor nutrition obtained from the calculation of weight /Age, very short status is obtained from the calculation of height/age, and very skinny status obtained from calculation of weight/height. And hopefully this system can be used in Posyandu to monitor the growth of children so that children with less nutritional status and poor nutrition get better and faster handling. The design of this decision support system uses a Web-based system with PHP programming language and a MySQL database for the web to be easily accessed.

© 2023 Universitas Pendidikan Indonesia

ARTICLE INFO

Article History:
Submitted/Received 20 Apr 2023
First Revised 07 May 2023
Accepted 04 Jun 2023
First Available online 21 Jun 2023
Publication Date 01 Oct 2023

Keyword:
Childhood,
K-Nearest Neighbor,
MySQL,
PHP,
Toddlers,
Web-based application.

1. INTRODUCTION

Adolescence is a period of very rapid growth and development. so that it is often termed a golden period as well as a critical period. The golden period can be realized if at this time infants and children receive appropriate nutritional intake for optimal growth and development [1-4]. According to data disclosed by the Director of Community Nutrition, the Ministry of Health, "The 2017 Nutrition Status Monitoring (PSG) results obtained 3.8% of children under five in Indonesia experiencing malnutrition. This **Figure 1.** increased from the previous year which was 3.4%. This shows that the problem of malnutrition needs to be handled more seriously. This PSG is an effort made by the government for planning, determining policies and monitoring and taking intervention actions in terms of improving toddler nutrition [5][6].

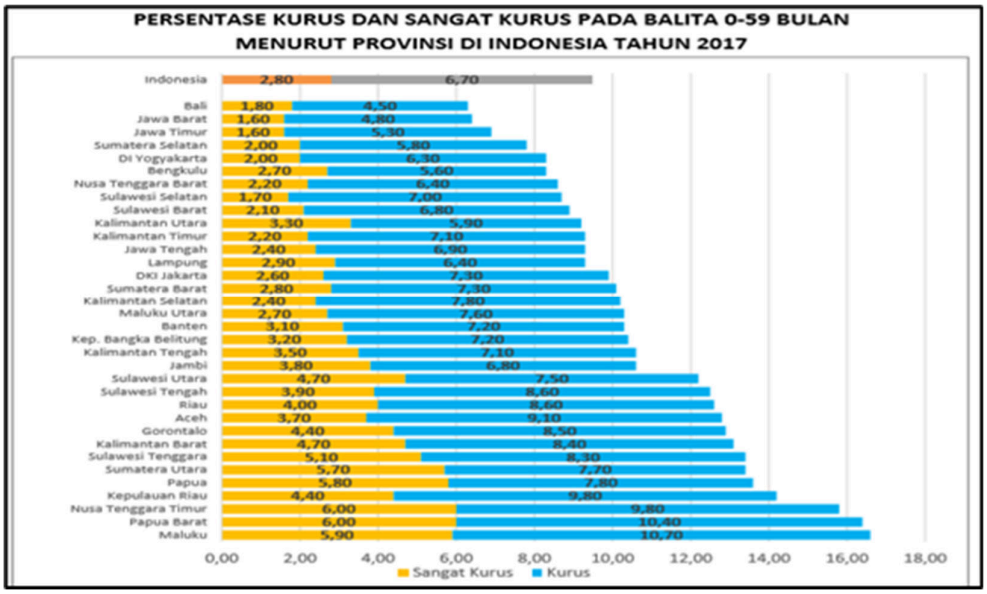


Figure 1. Graph of undernutrition and malnutrition in toddlers

The graph in **Figure 1.** shows cases of malnutrition and malnutrition by region in 2017. These cases occurred a lot in the East Nusa Tenggara region reaching 7.40 malnutrition and 20.90 malnutrition [7]. *Posyandu* is a form of community-based health effort (UKBM) that is managed and organized from, by, for and with the community in carrying out health development in order to empower the community and provide convenience to the community in obtaining basic/basic social health services to accelerate the reduction of mortality rates[8][9].

The people of Mangkalaya village - Sukabumi, in determining the nutritional status of toddlers are still manual, so that in providing information on the nutritional status of toddlers they have to wait a long time. In addition to the lack of control over the nutritional status of children, weighing data is also stored in a ledger, making it easier for data to be lost. The writer conducted this research using the K-Nearest Neighbor (KNN) method and compared several references such as K-Nearest Neighbor with Naïve Bayes and K-Nearest Neighbor with Backpropagation and K-Nearest Neighbor with Decision Trees[10-16].

From the results of the K-Nearest Neighbor with Naïve Bayes resulting in the results of classification of teak tree data from the forestry company, it was found that the K-Nearest Neighbor method has a higher level of accuracy, which is equal to 96.66%. Whereas for Naïve Bayes it is only 82.63% in terms of the accuracy of the results of the system and its calculations

DOI: https://doi.org/10.17509/coelite.v2i2.56919
p- ISSN 2829-4157 e- ISSN 2829-4149

[11]. While the comparison of K-Nearest Neighbor, if the classification can be done in a relatively faster time using the right algorithm is the K-Nearest Neighbor method, because using the K-Nearest Neighbor method with several training functions has an accuracy of 100%, whereas using the K-Nearest Neighbor method Backpropagation has a difference of 90% . And while the results of the K-Nearest Neighbor using the Decision Tree method produce an accuracy value of 98.18%, while the accuracy for the Decision Tree method is 96.06%. Therefore, to predict the nutritional status of toddlers, the K-Nearest Neighbor method is the best method [17-20].

From the comparison results above the K-Nearest Neighbor method in making decisions is much more accurate compared to other methods, and it is hoped that this research will provide a solution in determining the growth and development of early childhood from calculations with the K-Nearest Neighbor method according to existing criteria. This study provides an assessment of the growth and development of children at the *posyandu* starting from monitoring growth, nutrition and others, also helping the performance of *posyandu* cadres in calculating the nutritional status of children so that there are no delays in knowing the status of the child's growth and development.

2. METHODS

The research method is a scientific way to obtain data with a specific purpose. The scientific way means that research activities are based on scientific characteristics, namely rational, empirical, and systematic. From the definition above it can be concluded that the research method is a scientific way of obtaining data for certain uses.

2.1. Place and Research Object

The place of this research is the M.Toha *Posyandu* which is located in the village of Mkalaya – Sukabumi 43156, and the object of this research is the development and development of early childhood at the M.Toha Mkalaya *Posyandu*

2.2. Data Type

The study collected two types of data: quantitative data and qualitative data. Quantitative data was obtained from *Posyandu*, which included documents containing numerical information about children's data. These documents played a crucial role in supporting the research process related to child development. On the other hand, qualitative data was gathered through interviews conducted with relevant authorities or decision-makers involved in child growth and development. This data provided valuable insights into the current system used for calculating nutritional status.

2.3. Data Collection Stages

In this study, the researchers employed multiple approaches to gather comprehensive and reliable data. These approaches included literature study, interviews, and observation, each serving a unique purpose in data collection. The first method, literature study, involved an extensive review of relevant sources such as books, magazines, and online materials. This process allowed the researchers to acquire a solid theoretical foundation and gather information related to the research topic. Specifically, the literature study aimed to explore the K-Nearest Neighbor (K-NN) method and its applicability in the context of the study.

Moving on to the second method, interviews were conducted to obtain qualitative data. The researchers engaged in one-on-one interviews with individuals who possessed valuable

DOI: https://doi.org/10.17509/coelite.v2i2.56919
p- ISSN 2829-4157 e- ISSN 2829-4149

insights and expertise, such as the *Posyandu* cadres. By structuring the interviews around specific topics, such as performance, regulations, and monthly visit data, the researchers aimed to gain a deeper understanding of the current system for calculating nutritional status. Interviews provide an opportunity to capture subjective perspectives, personal experiences, and nuanced information that cannot be obtained through other means.

Lastly, the researchers employed observation as a direct data collection method. Through careful and systematic observation, they aimed to capture the actual behaviors, processes, and workflows taking place within the studied context. By immersing themselves in the field and witnessing the actions and interactions firsthand, the researchers gained valuable insights into the object of study. The observations were recorded meticulously to ensure accuracy and completeness.

Overall, this comprehensive data collection approach enabled the researchers to gather diverse types of data, including quantitative data from the literature study and qualitative data from interviews and observation. This multi-faceted approach provides a robust foundation for the subsequent analysis and interpretation of the findings.

2.4. System Planning

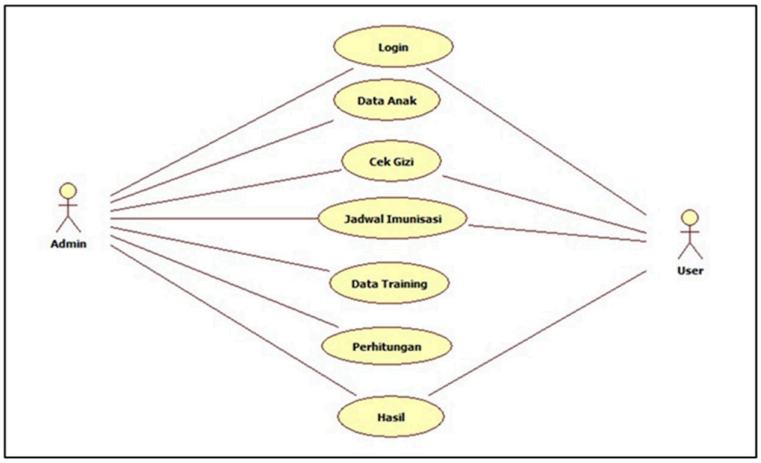


Figure 2. Use case diagram of child growth and development.

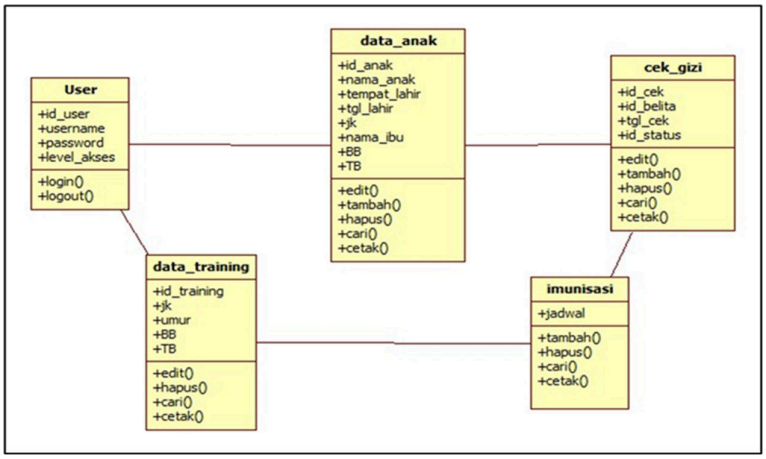


Figure 3. Class diagram of child growth and development.

DOI: https://doi.org/10.17509/coelite.v2i2.56919
p- ISSN 2829-4157 e- ISSN 2829-4149

3. RESULTS AND DISCUSSION

Testing Data Testing which totaled 25 data and Data Training 75 samples of early childhood data in **Figure 4**.

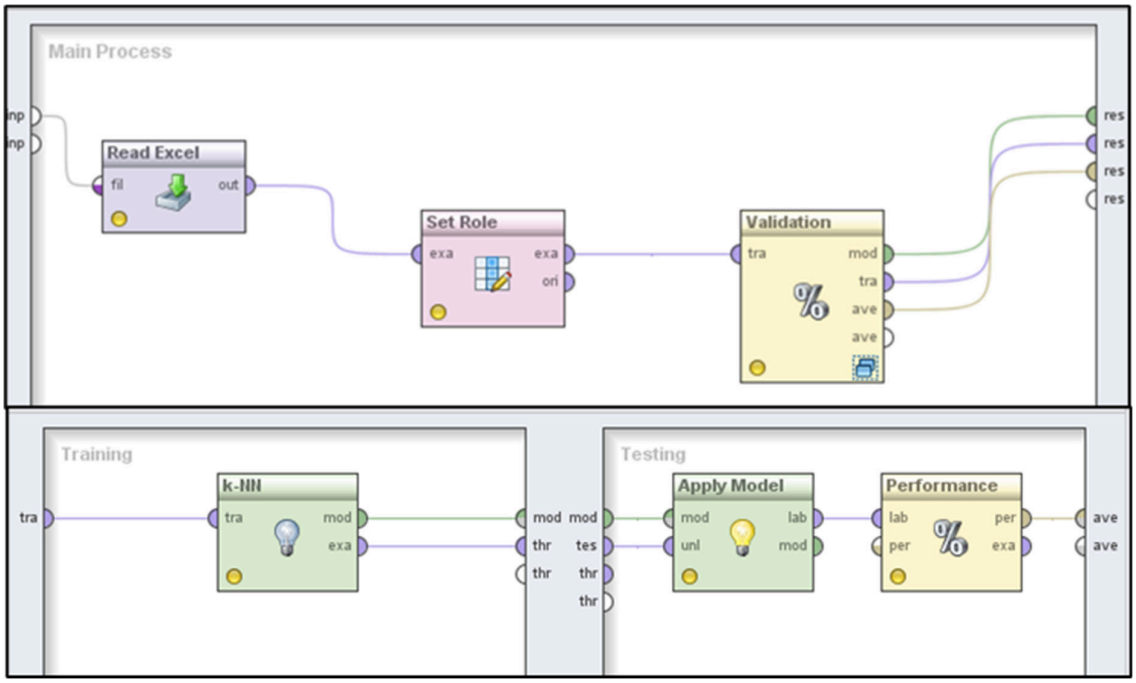


Figure 4. The K-Nearest Neighbor Testing Process from Testing Data.

The picture above shows that the testing process is added by the roll set operator, validation, apply model and performance.

The set roll operator functions as a pattern of behavior to channel two lines from one input data testing to validation, and from validation it continues to the k-nearest neighbor to be processed by the apply model operator without changing the input data testing earlier, and the results of the apply model are forwarded to the performance operator. to test the level of accuracy.

After the data testing process is complete, the results of the data accuracy are shown in **Table 1**.

Table 1. Testing KNN Accuracy Results from Data Testing

Accuracy.81.67%+/-32.02% (mikro:84.00%)			
	True LACK NUTRITION	**True GOOD NUTRITION**	**Class Precision**
Pred. MALNUTRITION	13	1	92.86%
Pred. GOOD NUTRITION	3	8	72.73%
Class Recall	81.25%	88.89%	

In the **Table 1.**, it is noted that the results of the accuracy of the K-Nearest Neighbor from the testing data are 81.67%, with a presentation of undernourishment of 81.25%, and good nutrition of 88.89%, with respective grade precision values of 92.86% for undernourished and 72.73% for good nutrition.

DOI: https://doi.org/10.17509/coelite.v2i2.56919
p- ISSN 2829-4157 e- ISSN 2829-4149

Table 2. KNN Accuracy Results from Training Data

Accuracy.77.08%+/-12.59% (mikro:76.71%)			
	True LACK NUTRITION	**True GOOD NUTRITION**	**Class Precision**
Pred. MALNUTRITION	13	1	92.86%
Pred. GOOD NUTRITION	3	8	72.73%
Class Recall	81.25%	88.89%	

In the **Table 2.**, it is noted that the results of the accuracy of the K-Nearest Neighbor from the testing data are 77.08%, with a normal presentation of 73.08%, and thin 78.72%, with each class precision value of 65.52% for normal and 84.09% for thin.

Table 3. Comparison of training and testing data accuracy values

No	**Scheme**	**KNN**
1	Training Data	77.08%
2	Data Testing	81.67%

In the **Table 3.**, it is noted that the KNN algorithm with training data has an accuracy value of 77.08% while data testing has a value of 81.67%, indicating that the accuracy value of testing data is greater than training data.

4. CONCLUSION

Based on the results of research and discussion related to determining the nutritional status of toddlers at *Posyandu* M.Toha Mangkalaya, the researchers draw the following conclusions:

1. The Decision Support System for the nutritional status of toddlers using the K-Nearest Neighbor is able to determine the nutritional status based on the parameters of the toddlers who will be checked using the Z-score formula to normalize the values for each parameter before calculating the distance using the Euclidian distance formula. Classification is carried out if the value of the distance between the training data and the data being tested has been sorted based on the shortest distance which is then drawn a conclusion based on the predetermined K value.
2. 2.The creation of this system was built based on the results of the needs analysis and system design described in the previous chapter in the form of a UML (Unified Modeling Language) model. At the analysis stage, system development requirements are obtained from the results of observations and interviews with related parties. In the following stages, system development is carried out based on the results of system analysis and implementation.
3. At the time of classification using the K-Nearest Neighbor (KNN) method, after the process of forming the classification accuracy value of the ROC and AUC curves shows that the K-Nearest Neighbor (KNN) algorithm with training data has an accuracy value of 81.67% while data testing has an accuracy value 77.08%.
4. Testing the feasibility of the system was carried out by 6 testers including 3 IT experts (IT practitioners) and 3 users (midwives, cadres and parents of children), testing by the IT field got a score of 72% which got a score which was included in the agree

DOI: https://doi.org/10.17509/coelite.v2i2.56919
p- ISSN 2829-4157 e- ISSN 2829-4149

category, while testing by users get a score of 73% included in the agreed category for the system to assist in the process of determining the status of child development.

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

6. REFERENCES

[1] Lockyer, S., White, A., and Buttriss, J. L. (2018). Biofortified crops for tackling micronutrient deficiencies–what impact are these having in developing countries and could they be of relevance within europe?. *Nutrition Bulletin*, *43*(4), 319-357.

[2] Jaman, J., Martić, K., Rasic, N., Markulin, H., and Haberle, S. (2021). Is the use of specific time cut-off or "golden period" for primary closure of acute traumatic wounds evidence based? a systematic review. *Croatian Medical Journal*, *62*(6), 614-622.

[3] Helm, S., Harmon, P. C., Noe, C., Calodney, A. K., Abd-Elsayed, A., Knezevic, N. N., and Racz, G. B. (2021). Transforaminal epidural steroid injections: a systematic review and meta-analysis of efficacy and safety. *Pain Physician*, *24*(S1), S209-S232.

[4] Lian, Z., and Van Ryzin, G. (2021). Optimal growth in two-sided markets. *Management Science*, *67*(11), 6862-6879.

[5] Ahluwalia, N. (2020). Nutrition monitoring of children aged birth to 24 mo (B-24): data collection and findings from the NHANES. *Advances in Nutrition*, *11*(1), 113-127.

[6] Khalatbari-Soltani, S., and Marques-Vidal, P. (2018). Adherence to hospital nutritional status monitoring and reporting guidelines. *PloS One*, *13*(9), 1-13.

[7] Besora-Moreno, M., Llauradó, E., Tarro, L., and Solà, R. (2020). Social and economic factors and malnutrition or the risk of malnutrition in the elderly: a systematic review and meta-analysis of observational studies. *Nutrients*, *12*(3), 1-16.

[8] Majid, M. I., dan Linuwih, S. (2019). Pengembangan unit kegiatan belajar mandiri (UKBM) materi usaha dan energi berbasis pembelajaran kontekstual untuk meningkatkan pemahaman konsep siswa. *UPEJ Unnes Physics Education Journal*, *8*(3), 228-238.

[9] Robertson, A. G., Easter, M. M., Lin, H., Frisman, L. K., Swanson, J. W., and Swartz, M. S. (2018). Medication-assisted treatment for alcohol-dependent adults with serious mental illness and criminal justice involvement: effects on treatment utilization and outcomes. *American Journal of Psychiatry*, *175*(7), 665-673.

[10] Bell, J., and Strang, J. (2020). Medication treatment of opioid use disorder. *Biological Psychiatry*, *87*(1), 82-88.

[11] Chen, L., Wang, C., Chen, J., Xiang, Z., and Hu, X. (2021). Voice disorder identification by using hilbert-huang transform (HHT) and k-nearest neighbor (KNN). *Journal of Voice*, *35*(6), 1-11.

[12] Anggoro, D. A., and Kurnia, N. D. (2020). Comparison of accuracy level of support vector machine (SVM) and k-nearest neighbors (KNN) algorithms in predicting heart disease. *International Journal*, *8*(5), 1689-1694.

DOI: https://doi.org/10.17509/coelite.v2i2.56919
p- ISSN 2829-4157 e- ISSN 2829-4149

[13] Muhathir, M., Sibarani, T. T. S., and Al-Khowarizmi, A. K. (2020). Analysis k-nearest neighbors (KNN) in identifying tuberculosis disease (TB) by utilizing hog feature extraction. *Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal*, *1*(1), 226-235.

[14] Lu, J., Qian, W., Li, S., and Cui, R. (2021). Enhanced k-nearest neighbor for intelligent fault diagnosis of rotating machinery. *Applied Sciences*, *11*(3), 1-14.

[15] Rajaguru, H., and SR, S. C. (2019). Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer. *Asian Pacific Journal of Cancer Prevention: APJCP*, *20*(12), 3777–3781.

[16] Himeur, Y., Alsalemi, A., Bensaali, F., and Amira, A. (2021). Smart power consumption abnormality detection in buildings using micromoments and improved k-nearest neighbors. *International Journal of Intelligent Systems*, *36*(6), 2865-2894.

[17] Ula, M., Ulva, A. F., Saputra, I., Mauliza, M., and Maulana, I. (2022). Implementation of machine learning using the k-nearest neighbor classification model in diagnosing malnutrition in children. *Multica Science and Technology (MST) Journal*, *2*(1), 94-99.

[18] Marniati, M., Putri, E. S., Sriwahyuni, S., Khairunnas, K., and Duana, M. (2020). Knowledge study, income level and socio-culture of the nutritional status of toddler. *Journal of Nutrition Science*, *1*(2), 65-71.

[19] Low, D. W., Jamil, A., Md Nor, N., Kader Ibrahim, S. B., and Poh, B. K. (2020). Food restriction, nutrition status, and growth in toddlers with atopic dermatitis. *Pediatric Dermatology*, *37*(1), 69-77.

[20] Riley, L. K., Rupert, J., and Boucher, O. (2018). Nutrition in toddlers. *American Family Physician*, *98*(4), 227-233.

DOI: https://doi.org/10.17509/coelite.v2i2.56919
p- ISSN 2829-4157 e- ISSN 2829-4149

Journal of Computer Engineering, Electronics and Information Technology (COELITE)

Journal homepage: https://ejournal.upi.edu/index.php/COELITE

Sentiment Analysis of Twitter User Opinions Data Regarding the Use of ChatGPT in Education

Jezzy Putra Munggaran[*], *Ahmad Ali Alhafidz, Maulana Taqy, Devi Aprianti Rimadhani Agustini, M. Munawir*

Department of Computer Engineering, Universitas Pendidikan Indonesia, Indonesia
Correspondence: E-mail: jezzy.putra.munggaran@upi.edu

ABSTRACT

This article presents a sentiment analysis of Twitter users' opinions regarding the use of ChatGPT in education. ChatGPT, an AI chatbot developed by OpenAI, has gained significant attention for its ability to provide detailed responses across various knowledge domains. However, concerns have been raised about its occasional inclusion of inaccurate information. This study aims to analyze the sentiment of Twitter users' opinions towards ChatGPT in education and evaluate its accuracy. The sentiment analysis process involves data crawling, labelling, preprocessing, sentiment analysis, and evaluation. Data is collected from Twitter using the RapidMiner Studio tool and labelled as positive or negative sentiment based on the presence of positive or negative words. Preprocessing techniques are applied to standardize and reduce the volume of words in the data. The sentiment analysis classification is performed using machine learning algorithms, specifically Naive Bayes and Support Vector Machine (SVM). The accuracy, precision, and recall of the classification models are evaluated. The sentiment analysis results provide insights into Twitter users' overall sentiment towards ChatGPT in education. This study contributes to understanding Twitter users' opinions and sentiments regarding using ChatGPT in education. The findings can be valuable for educators and policymakers in assessing the potential impact of ChatGPT on academic integrity and the educational landscape.

ARTICLE INFO

Article History:
Submitted/Received 30 Apr 2023
First Revised 10 May 2023
Accepted 7 Jun 2023
First Available online 21 Jun 2023
Publication Date 01 Oct 2023

Keyword:
ChatGPT,
Education,
RapidMiner,
Sentiment analysis,
Twitter.

© 2023 Universitas Pendidikan Indonesia

1. INTRODUCTION

A few months back, in December 2022, an Artificial Intelligence (AI) tool called ChatGPT got viral just a month after its release. It is getting recognized by many people because of its ability to solve almost any questions users give. Many people see ChatGPT as a massive breakthrough in the age of AI [1].

ChatGPT is an AI chatbot created by OpenAI and made public in November 2022. The name "Chat" signifies its purpose as a conversational bot, while "GPT" refers to its foundation as a large language model known as a generative pre-trained transformer (LLM). ChatGPT is built upon OpenAI's GPT-3.5 and GPT-4 models and has been fine-tuned for conversational applications using supervised and reinforcement learning methods [2].

ChatGPT was initially introduced as a prototype on November 30th, 2022, and gained attention due to its ability to provide detailed and coherent responses across various knowledge domains [3]. However, it has been noted that the chatbot occasionally includes inaccurate information, which is seen as a significant drawback. As a result of the release of ChatGPT, OpenAI's estimated valuation reached US$29 billion in 2023. This introduction of the chatbot has intensified competition in the field, leading to the development of Google's chatbot Bard (based on LaMDA and PaLM) and Meta's LLaMA foundation model, which is being utilized by others to create their chatbots [4-7].

The initial version of ChatGPT was based on GPT-3.5, but a version utilizing GPT-4, the latest OpenAI model, was made available to limited paid subscribers on March 14th, 2023 [8]. The emergence of ChatGPT and GPT-4 has raised concerns among computer scientists such as Geoffrey Hinton and Yoshua Bengio. These concerns include the potential for future AI systems to surpass human intelligence, pursue misaligned goals, and pose existential risks [9].

Regarding ChatGPT, according to a recent survey conducted by Study.com, an online learning platform, 43% of educators express concerns that the program will create additional professional challenges. Conversely, nearly six out of ten teachers anticipate that it will enhance their professional lives [10].

One aspect of concern among educators is the potential threat to academic integrity posed by the program, as it may promote novel methods of cheating and plagiarism. Due to its user-friendly interface, accessibility, and convenience, students have resorted to utilizing the program to generate responses for homework assignments and even complete essays, falsely claiming authorship of the chatbot-generated content as their work [11].

However, on the other hand, some educators find the straightforward design and brainstorming capabilities of ChatGPT appealing, as they recognize its potential to enhance the educational landscape. These teachers argue that, in the long run, the program's impact will not primarily increase academic dishonesty but revitalize lesson plans and classroom instruction [12-16].

2. METHODS

To do some sentiment analysis and determine its accuracy, several stages are involved to obtain the best results. The following steps are performed: data crawling, labelling, preprocessing, sentiment analysis, and evaluation to determine accuracy, precision, and recall [17-19].

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

2.1. Data Crawling

The first stage in sentiment analysis is data collection through data crawling. Data is collected from Twitter using the Twitter Application Programming Interface (API) with RapidMiner Studio version 10.1 as the tool. Tweets containing the keyword "chatgpt in education" are retrieved. The crawled data is then manually labelled as either positive or negative sentiment. The model used for crawling Twitter data is illustrated in **Figure 1**.

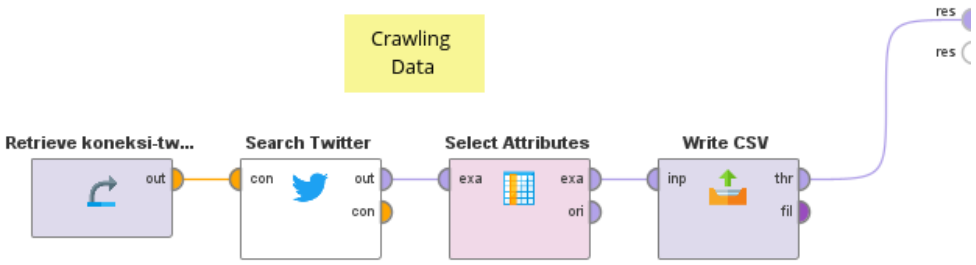


Figure 1. The data crawling process.

Figure 1. depicts the Twitter data crawling using the "twitter-v3-connection" operator, which is connected to the Twitter API, and the "Search Twitter" operator. The tweets are retrieved by selecting the "Text" attribute using the "Select Attributes" operator and saved in a .csv file format using the "Write CSV" operator.

2.2. Data Labelling

The next stage involves data labelling, where the data is divided into two sentiment classes: positive and negative. The criteria for labelling tweets are based on the presence of positive or negative words. Neutral sentences agreeing that ChatGPT positively impacts education are categorized as positive. In contrast, tweets containing negative words and expressing disagreement regarding the positive impact of ChatGPT on education are classified as negative. **Table 1.** below shows an example of 4 labelled datasets out of the total 804 datasets.

Table 1. Data labeling table.

Text	Sentiment
From now on, New York City Public Schools will overturn the ChatGPT ban, embracing the potential benefits of AI in education, and preparing students for the future in this transformative move. https://t.co/kS4759aOyZ	Positive
ChatGPT is revolutionizing education! From personalized instruction to assignment assistance, this AI is transforming the way teachers teach and students learn. Link https://t.co/FzkB2D6QEh	Positive
@TRKShady @Xx17965797N AI will not end well for us. ChatGPT has already taken over education. Nearly all high schoolers and college students use it for homework. The workforce is about to get chopped and all I hear are praises and cheers on LinkedIn. https://t.co/3Z325sTfyj	Negative
The huge potential problem with chatgpt and other generative ML in vital institutions like medicine, science, education? Law? is as some people have said, there is no clear vetting of information and peer review from a giant dataset already poisoned with misinformation.	Negative

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

2.3. Preprocessing Data

After labelling the data, the next stage is preprocessing. Both the training and testing data need to undergo preprocessing before the classification process to reduce the dimension of the vector space model and make the classification process faster. The goal of preprocessing is to standardize and reduce the volume of words. This stage is performed to prepare the data for analysis. The preprocessing steps are divided into two parts: preprocessing subprocess before splitting the data into training and testing sets and preprocessing in the document process. The cleansing steps in the subprocess include Replace RT, Replace URL 1, Replace URL 2, Replace Hashtag, Replace Mention, Replace Symbol, and Trim. The first preprocessing operator is shown in **Figure 2.** below.

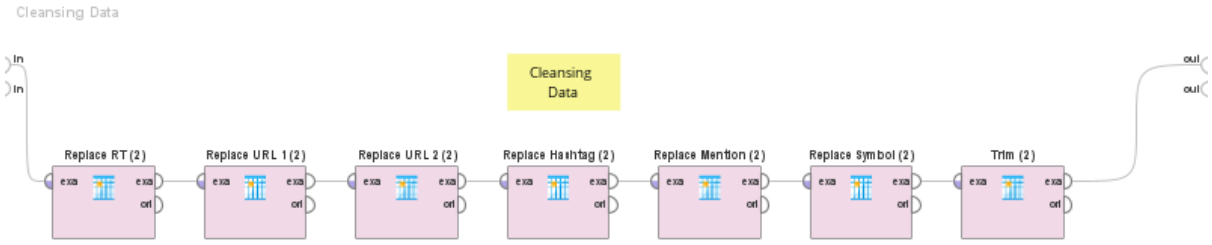


Figure 2. The data cleansing process.

The second preprocessing stage in the "Document Process" uses the operators "Transform Case", "Tokenize", "Filter Token by Length", "Filter Stopword (English)", and "Stem (Porter)" as shown in **Figure 3.** below.

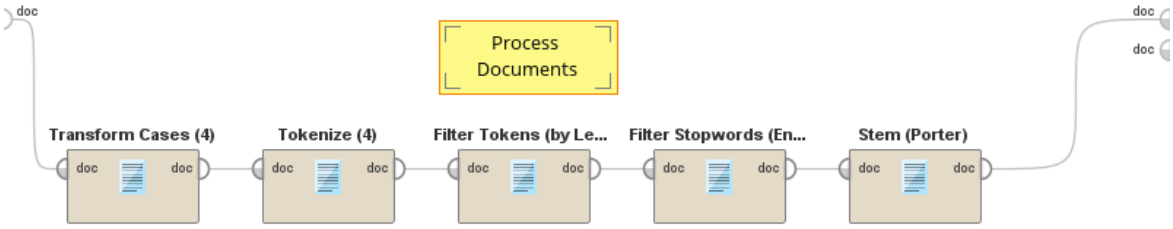


Figure 3. The preprocessing process in document process.

The following is an explanation of each preprocessing stage:

2.3.1. Cleansing

Cleansing is the stage where unnecessary characters and punctuation marks are removed from the text. The purpose of cleansing is to reduce noise or disturbances in the dataset. The cleansing steps include Replace RT, Replace URL 1, Replace URL 2, Replace Hashtag, Replace Mention, Replace Symbol, and Trim. These processes utilize regular expressions (Regrex) as parameters in the RapidMiner operators. Examples of characters removed include URLs, hashtags (#), usernames (@), RT, as well as punctuation marks such as periods (.), and commas (,).

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

2.3.2. Case Folding

The case folding process involves converting all characters in a document to either upper case or lower case. In this study, lowercase is used to standardize the text. Case folding is performed to facilitate search operations as text documents may have inconsistent capitalization. For example, "ChatGPT in Education" is transformed into "chatgpt in education".

2.3.3. Tokenization

Tokenization is the process of breaking sentences into meaningful and significant words. It separates each word that forms a sentence. For example, "chatgpt in education" after tokenization becomes "chatgpt", "in", and "education".

2.3.4. Filter Tokens by Length

After tokenization, the Filter Tokens by Length stage selects tokens based on a minimum and maximum character requirement. In this study, only tokens with 2-25 characters are included. Thus, only words with 2 to 25 characters are extracted in this process.

2.3.5. Stopword Removal

Stopword Removal is the process of eliminating words that do not contribute to the intended meaning. These words are commonly found in a dictionary and are known as stopwords. Examples of stopwords used in this study include "or," "there," "is," "to," "back," "then," "by," and "that."

2.3.6. Stemming

Stemming involves transforming inflected or derived words into their base form to capture the underlying meaning and achieve more specific categorization. For example, the word "improper" stems from its base form, "proper".

2.3.7. Word Weighting

Word weighting is a mechanism to assign scores to the frequency of word occurrences in a text document.

2.3.8. Term Frequency – Inverse Document Frequency (TF-IDF)

Term Frequency – Inverse Document Frequency (TF-IDF) is a powerful and widely used weighted scoring algorithm in the field of information retrieval and natural language processing. It combines the strength of two distinct weighting algorithms, Term Frequency (TF) and Inverse Document Frequency (IDF), to provide a comprehensive measure of word importance within and across a corpus of documents.

TF-IDF operates by calculating the relative frequency of a word in a specific document, considering its inverse proportion across the entire collection of documents. This process helps to discern the relevance and significance of a word within the context of a particular document, as it takes into account not only its occurrence but also its scarcity or prevalence in the overall document collection.

The Term Frequency component of TF-IDF measures the frequency of a word within a given document. It assigns higher weights to words that appear more frequently, assuming that these words are more relevant and informative in conveying the document's content. By

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

emphasizing frequently occurring words, TF captures the local importance of terms within a document, enabling a more fine-grained analysis of its textual content [20].

On the other hand, Inverse Document Frequency addresses the global importance of a word across the entire collection of documents. It computes a weight that diminishes the significance of words that are commonly found in many documents, as these words tend to be less discriminative in distinguishing between documents. By inversely scaling the weights of such common words, IDF allows for highlighting words that are more unique and specific to individual documents, thereby enhancing their importance in the ranking and retrieval processes [21].

Integrating TF and IDF in the TF-IDF algorithm facilitates a robust method for identifying essential words in a document and differentiating them from less informative terms. This technique has been widely applied in various applications, including text classification, information retrieval, and document clustering, enabling efficient and accurate analyses of large textual datasets [22].

TF-IDF plays a crucial role in information retrieval and natural language processing tasks by capturing words' relative frequency and significance within individual documents and across a corpus. By combining Term Frequency and Inverse Document Frequency, this algorithm provides a comprehensive and nuanced representation of word importance, contributing to enhanced document understanding and effective text-based analysis.

2.4. Wordcloud

After preprocessing the data, wordcloud generation is performed, which is not crucial to the subsequent processes. Wordclouds are created to visualise frequently occurring words during data crawling on Twitter using the keyword or query "chatgpt in education." The RapidMiner operators used for wordcloud generation are illustrated in **Figure 4.**

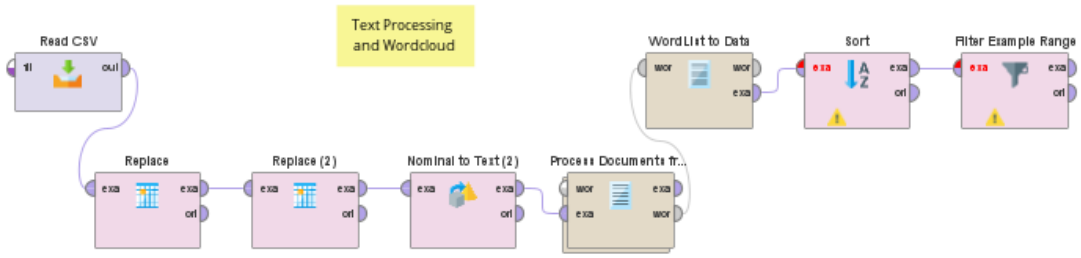


Figure 4. Wordcloud generation process.

2.5. Data Classification Techniques

After data preprocessing, the next stage involves sentiment analysis classification. This stage involves training and implementing various machine learning algorithms. **Figure 5.** displays the contents of the "Cross Validation" operator in RapidMiner. Two different classification operators, Naive Bayes and SVM (Support Vector Machine), are used for comparison purposes. After creating the proposed models through experimentation, further experiments are conducted to test the existing models using pre-grouped datasets as training and testing data. The "Performance" operator is used to display accuracy, precision, and recall results.

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

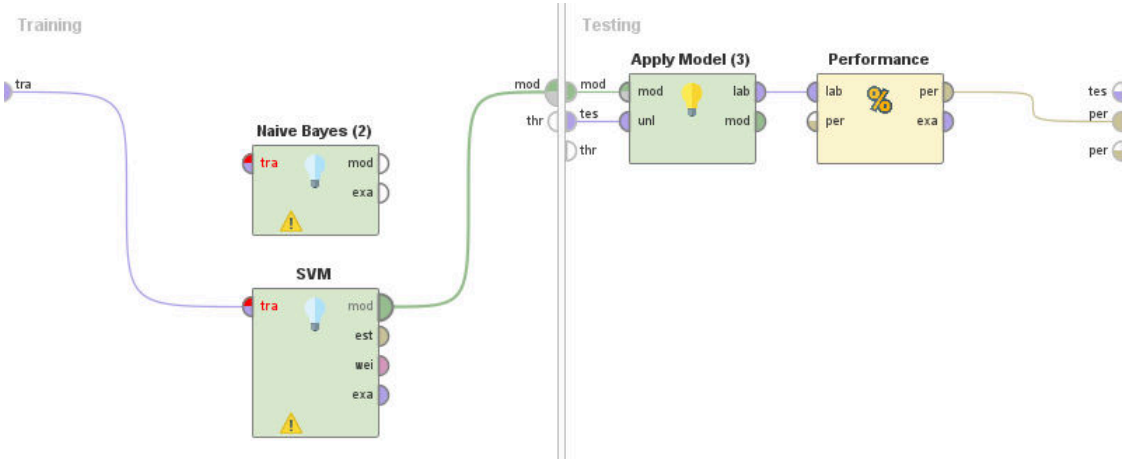


Figure 5. Sentiment analysis classification.

2.5.1. Naïve Bayes

Naive Bayes is a machine learning algorithm that utilizes probability calculations based on Bayesian inference. The application of Bayes' theorem in the Naive Bayes algorithm involves combining prior probabilities and conditional probabilities in a formula that can be used to compute the probability of each possible classification. The formula can be seen in Equation [1].

$$P(X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

- X: Data with an unknown class
- H: Hypothesis that data X belongs to a specific class
- P(H|X): Probability of hypothesis H based on condition X
- P(H): Probability of hypothesis H
- P(X|H): Probability of data X based on condition H
- P(X): Probability of X

2.5.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful and widely used method for analyzing data and identifying intricate patterns that can be effectively employed in classification tasks. In the realm of binary classification, where data is divided into positive and negative classes, SVM stands out as an exceptional approach. This method operates by skillfully constructing hyperplanes in multidimensional input space, and strategically optimizing their positioning to best segregate the different classes. By precisely measuring the boundaries between class memberships, SVM achieves unparalleled accuracy in classification tasks.

Originally, SVM gained prominence for its prowess in classifying numerical data. However, its remarkable capabilities and versatility have transcended beyond numerical domains, making it an invaluable tool for tackling complex text data problems as well. Text data is particularly well-suited for classification using the SVM algorithm due to its inherent high-dimensional nature. Textual information is often characterized by a plethora of features that may not all be relevant, yet tend to exhibit correlations with one another. SVM adeptly handles this challenge by effectively organizing the features into separate linear categories, allowing for efficient and accurate classification of text-based data.

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

The ability of SVM to excel in classifying text data can be attributed to its unique strengths. By skillfully constructing hyperplanes, SVM can discern subtle linguistic nuances and identify patterns that may not be immediately apparent to human observers. This capability is crucial when dealing with text, as it allows SVM to capture intricate relationships between words, phrases, and concepts, thereby facilitating precise and meaningful classification. Additionally, SVM's robustness to noise and its ability to handle large feature spaces further contribute to its effectiveness in text classification tasks.

Support Vector Machine (SVM) is a sophisticated and versatile method that revolutionizes the field of classification. Its ability to construct hyperplanes in multidimensional spaces enables accurate discrimination between classes, making it an invaluable tool for binary classification problems. Moreover, SVM's exceptional adaptability to text data allows it to efficiently handle the high-dimensional and complex nature of textual information. By effectively organizing features into separate linear categories, SVM unlocks valuable insights hidden within text, paving the way for advanced applications in natural language processing, sentiment analysis, and document classification, among others [23].

2.6. Sentiment Analysis Evaluation

After the classification process, evaluation is necessary to determine the quality of the performed process. In this evaluation stage, the performance of the classification process is tested using accuracy, precision, and recall as evaluation parameters. The evaluation in this system analysis is conducted by calculating the accuracy level of a method in analyzing opinions. K-fold cross-validation is chosen to test the accuracy of the Naive Bayes and Support Vector Machine methods. Cross-validation is based on the principle of dividing data into two parts: training data and testing data. In this study, 10-fold cross-validation is used. The data is divided into 10 parts, where Part 1 becomes the training data and Parts 2-10 becomes the testing data. Subsequently, a cross-validation process is performed, where the training data becomes the testing data and vice versa. This process is repeated 10 times.

In the 10-fold cross-validation process, a confusion matrix is created, which consists of True Positives (TP) representing the number of positive class data classified correctly as positive, and True Negatives (TN) representing the number of negative class data classified correctly as negative. False Positives (FP) represents the number of negative class data classified incorrectly as positive, and False Negatives (FN) represents the number of positive class data classified incorrectly as negative. The average accuracy, precision, and recall values can be obtained from the confusion matrix table as shown in **Table 2.** below.

Table 2. Confusion matrix table.

	True Yes	**True No**
Prediction Yes	TP	FP
Prediction No	FN	TN

Accuracy is the number of documents correctly classified when comparing the predicted results of the model with the predetermined sentiment, including both True Positives and True Negatives. The accuracy can be calculated using the formula in Equation [2].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (2)$$

Precision is the ratio of True Positives to the total number of data predicted as positive. Precision can be calculated using the formula in Equation [3].

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (3)$$

Recall represents how many relevant documents are retrieved from the total number of documents that are actually positive, including False Negatives. Recall can be calculated using the formula in Equation [4].

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

3. RESULTS AND DISCUSSION

This section presents the findings of the conducted research. This study aimed to determine the sentiment analysis of Twitter users' opinions regarding the use of ChatGPT in education and compare the classification results between the Naive Bayes and Support Vector Machine (SVM) methods. The research utilized a dataset consisting of 803 tweets, with 300 tweets allocated for training and 503 tweets for testing purposes.

Table 3. provides the percentage distribution of sentiment polarity in the obtained tweets. The sentiment analysis of the collected tweets revealed 635 positive tweets and 168 negative tweets. The percentage of positive opinions was 79.07%, while the percentage of negative opinions was 20.93%. Based on these results, it can be concluded that Twitter users' opinions regarding the use of ChatGPT in education were predominantly positive. However, some netizens expressed reservations about the utilization of ChatGPT in the educational context, as it is not their field of expertise. The overall research findings are elaborated in **Table 3.**, **Figure 6.**, and **Figure 7.** below.

Table 3. Sentiment polarity of tweets.

	Sum of Tweets	Percentage
Positive Tweets	635	79.07%
Negative Tweets	168	20.93%

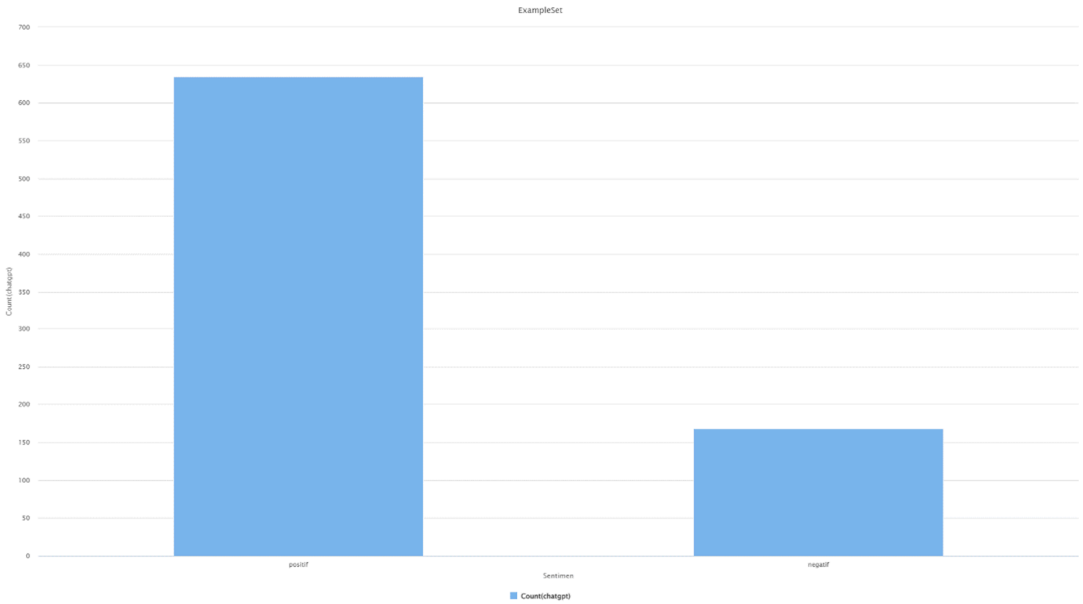


Figure 6. Histogram of positive and negative tweets.

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

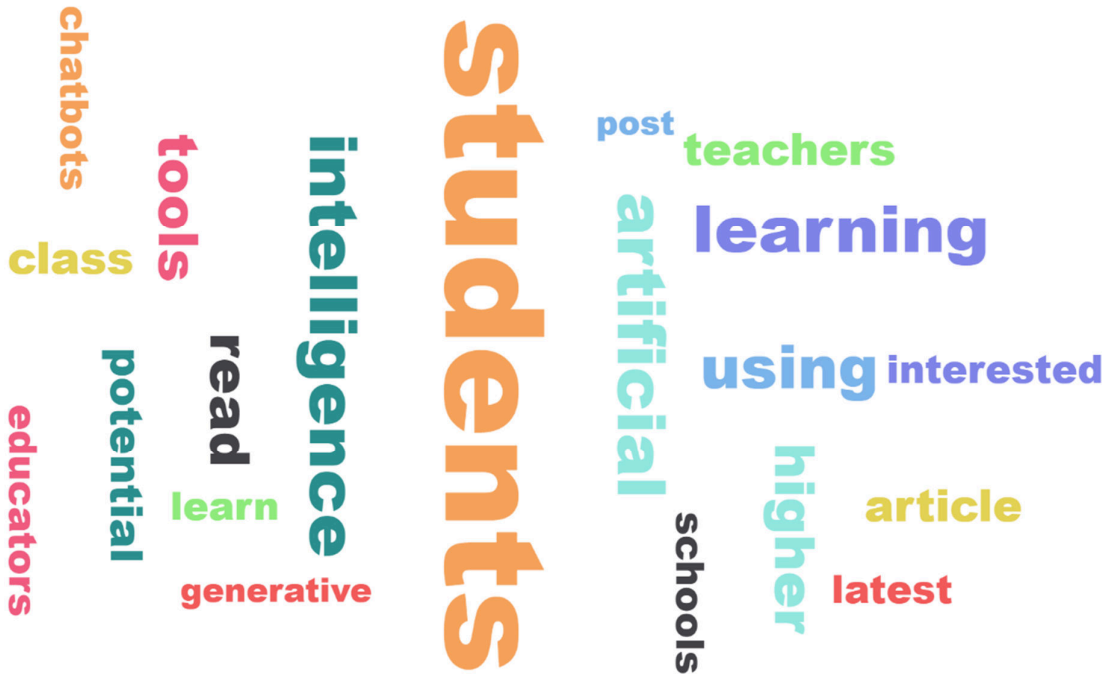


Figure 7. Wordcloud generated based on positive and negative tweets.

Figure 8. illustrates the primary process in the RapidMiner application. The "Dataset" operator represents the "Read Excel" operator used to read the dataset from an Excel file. The "Subprocess" operator signifies the initial stage of preprocessing, where data cleansing takes place. The "Training Process" and "Testing Process" operators denote the "Document Process" operators employed for preprocessing and data weighting. The "Cross Validation" operator is utilized for sentiment analysis classification and evaluation, with the experiment conducted ten times using 10-fold cross-validation.



Figure 8. Main Process in RapidMiner

Figure 9. displays the sentiment prediction results obtained from the RapidMiner application. A comparison is made between the predetermined and predicted sentiment labels generated during the process. The following **Table 4.** presents the confusion matrix results for each algorithm.

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

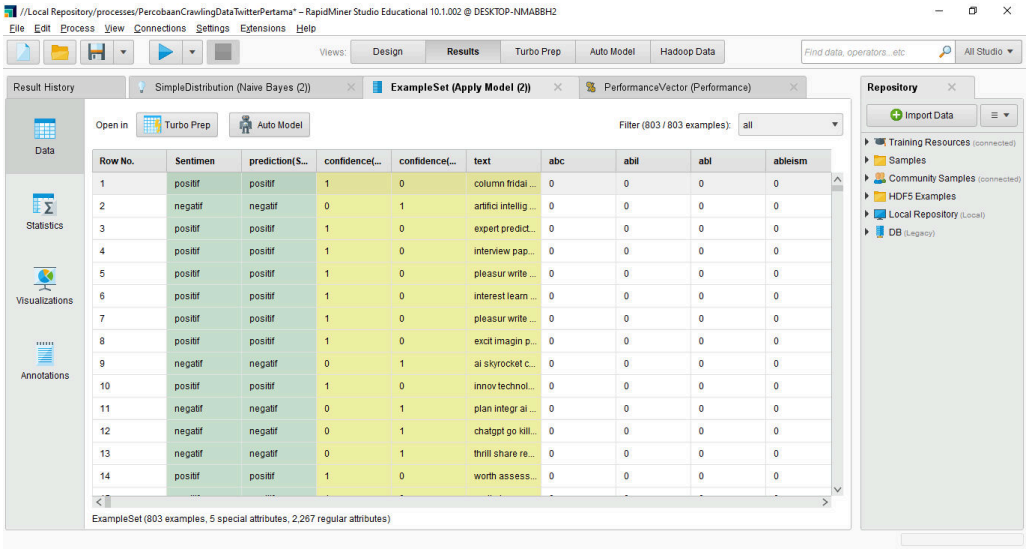


Figure 9. Prediction Results from RapidMiner

Table 4. Confusion matrix for every method.

Metode	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)
Naïve Bayes	614	53	115	21
SVM	628	99	69	7

Based on the confusion matrix table in **Table 4.**, the average values of accuracy, precision, and recall in RapidMiner are shown in **Table 5**.

Table 5. Accuracy, precision, and recall results.

Metode	Accuracy	Precision	Recall
Naïve Bayes	90.77%	84.98%	68.42%
SVM	86.79%	90.08%	41.10%

Table 5. presents the accuracy results of Naive Bayes as 90.77% and SVM as 86.79%. The precision results for Naive Bayes are 84.98%, while for SVM, they are 90.08%. Furthermore, the recall values for Naive Bayes and SVM are 68.42% and 41.10%, respectively. It can be observed that Naive Bayes performed better as a classifier in sentiment analysis using the Twitter dataset, providing more accurate and precise predictions.

The disparity in results can be attributed to the suitability of dataset characteristics and the different parameters of each algorithm. In this case, Naive Bayes operates on independent variables, specifically text data from tweets.

Due to the compatibility between the data type and variables, Naive Bayes outperforms Support Vector Machine. This difference in polarity results may also stem from SVM separating data based on predetermined linear patterns through word weighting rather than the frequency-based probabilities of the words.

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

4. CONCLUSION

In this study, a sentiment analysis classification process was conducted on various opinions regarding the use of ChatGPT in the field of education. The conclusion drawn from the sentiment polarity of netizens' opinions indicates a dominant positive sentiment, with positive opinions accounting for 69.72% and negative opinions accounting for 30.28%. Based on these results, this sentiment can be utilized to evaluate the performance of ChatGPT's utilization in education until the completion of this research, as well as to assess and consider future performance.

The research employed two different classifiers, namely Naive Bayes and SVM. These classifiers predicted labels within the dataset based on the data patterns present. The results revealed that the accuracy of Naive Bayes was 90.77%, while SVM achieved an accuracy of 86.79%. The precision results for Naive Bayes were 84.98%, compared to 90.08% for SVM. Moreover, the recall values for Naive Bayes and SVM were 68.42% and 41.10%, respectively. Therefore, it can be concluded that Naive Bayes outperforms SVM by providing more accurate and precise predictions. This difference in performance can be attributed to the suitability of dataset characteristics and the differing parameters of each algorithm. In this study, Naive Bayes operated on independent variables, specifically text data from tweets. Due to the compatibility between the data type and variables, Naive Bayes exhibited superior performance compared to SVM.

Based on the findings of this research, it is recommended to continue exploring the application of ChatGPT in the field of education. The dominant positive sentiment expressed by netizens suggests a favorable perception of its usage. Furthermore, future studies can further focus on refining and enhancing the Naive Bayes classifier to improve accuracy and precision in sentiment analysis tasks. Additionally, it may be beneficial to investigate other machine learning algorithms or hybrid approaches to sentiment analysis to explore their potential to capture the nuances of user opinions more effectively.

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

6. REFERENCES

[1] Surameery, N. M. S., and Shakor, M. Y. (2023). Use chatGPT to solve programming bugs. *International Journal of Information Technology and Computer Engineering (IJITC)*, *3*(01), 17-22.

[2] Naumova, E. N. (2023). A mistake-find exercise: a teacher's tool to engage with information innovations, chatGPT, and their analogs. *Journal of Public Health Policy*, *44*(2), 173-178.

[3] George, A. S., George, A. H., and Martin, A. G. (2023). The environmental impact of AI: a case study of water consumption by chatGPT. *Partners Universal International Innovation Journal*, *1*(2), 97-104.

[4] Cooper, G. (2023). Examining science education in chatGPT: an exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, *32*(3), 444-452.

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

[5] Rospigliosi, P. A. (2023). Artificial intelligence in teaching and learning: what questions should we ask of chatGPT?. *Interactive Learning Environments*, *31*(1), 1-3.
[6] George, A. S., and George, A. H. (2023). A review of chatGPT AI's impact on several business sectors. *Partners Universal International Innovation Journal*, *1*(1), 9-23.
[7] Alhumaid, K., Naqbi, S., Elsori, D., and Mansoori, M. (2023). The adoption of artificial intelligence applications in education. *International Journal of Data and Network Science*, *7*(1), 457-466.
[8] Lee, P., Bubeck, S., and Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *New England Journal of Medicine*, *388*(13), 1233-1239.
[9] Ouyang, F., Wu, M., Zheng, L., Zhang, L., and Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, *20*(1), 1-23.
[10] Guerrero, D. T., Asaad, M., Rajesh, A., Hassan, A., and Butler, C. E. (2023). Advancing surgical education: the use of artificial intelligence in surgical training. *The American Surgeon*, *89*(1), 49-54.
[11] Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., and Hatta, H. R. (2023). Analysing the role of chatGPT in improving student productivity in higher education. *Journal on Education*, *5*(4), 14886-14891.
[12] Ahmad, N., Murugesan, S., and Kshetri, N. (2023). Generative artificial intelligence and the education sector. *Computer*, *56*(6), 72-76.
[13] Giuggioli, G., and Pellegrini, M. M. (2023). Artificial intelligence as an enabler for entrepreneurs: a systematic literature review and an agenda for future research. *International Journal of Entrepreneurial Behavior and Research*, *29*(4), 816-837.
[14] Rahman, M. M., and Watanobe, Y. (2023). ChatGPT for education and research: opportunities, threats, and strategies. *Applied Sciences*, *13*(9), 5783.
[15] Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., and Agyemang, B. (2023). What if the devil is my guardian angel: chatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 15.
[16] Yang, T. C. (2023). Application of artificial intelligence techniques in analysis and assessment of digital competence in university courses. *Educational Technology and Society*, *26*(1), 232-243.
[17] Munawir, M., Rendy, P. K., Lia, N. F., and Bart, D. (2020). Place branding reputation by online reviewers' perceptions using sentiment analysis of thematic park in Bandung city, Indonesia. *JAILCD*, *1*(1), 211-216.
[18] Munawir, B. D., and Fritz, A. N. (2017). Sustainable urban planning based on sentiment analysis on twitter. *JAILCD*, *1*(1), 281-284.
[19] Munawir, Koerniawan, M. D., and Dewancker, B. J. (2019). Visitor perceptions and effectiveness of place branding strategies in thematic parks in Bandung city using text mining based on google maps user reviews. *Sustainability*, *11*(7), 1-20.
[20] Suryaningrum, K. M. (2023). Comparison of the TF-IDF method with the count vectorizer to classify hate speech. *Engineering, Mathematics and Computer Science (EMACS) Journal*, *5*(2), 79-83.
[21] Kandhro, I. A., Jumani, S. Z., Lashari, A. A., Nangraj, S. S., Lakhan, Q. A., Baig, M. T., and Guriro, S. (2019). Classification of sindhi headline news documents based on TF-IDF text analysis scheme. *Indian Journal of Science and Technology*, *12*(33), 1-10.

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

[22] Luthfi, M. F., and Lhaksamana, K. M. (2020). Implementation of TF-IDF method and support vector machine algorithm for job applicants text classification. *Jurnal Media Informatika Budidarma*, *4*(4), 1181-1186.

[23] Hendriyana, H., Karo, I. M. K., dan Dewi, S. (2022). Analisis perbandingan algoritma support vector machine, naive bayes dan regresi logistik untuk memprediksi donor darah. *Jurnal Teknologi Terpadu*, *8*(2), 121-126.

DOI: https://doi.org/10.17509/coelite.v2i2.59645
p- ISSN 2829-4157 e- ISSN 2829-4149

Journal of Computer Engineering, Electronics and Information Technology (COELITE)

Journal homepage: https://ejournal.upi.edu/index.php/COELITE

Comparison of Machine Learning and Deep Learning Algorithms for Classification of Breast Cancer

*Puji Ayuningtyas[1, *], R. Rahmawati[2], Akhmad Miftahusalam[3]*

[1]Department of Information Engineering, Institut Teknologi Telkom Purwokerto, Indonesia
[2]Department of AI Research, PT Bisa Artificial Indonesia, Indonesia
[3]Department of AI Hacker, PT Bisa Artificial Indonesia, Indonesia
Correspondence: E-mail: 20102122@ittelkom-pwt.ac.id

ABSTRACT

Statistical data from the American Cancer Society which shows that breast cancer ranks first with the highest number of cases of all types of cases of malignant tumors (cancer) worldwide. through a data mining process that is used to extract information and data analysis, a classification process can be carried out to carry out further analysis of the pattern of a data. The dataset used in this study is the Breast Cancer Wisconsin (Diagnostic) Dataset obtained from UCI Machine Learning. The purpose of this study is to compare five algorithms, namely Logistic Regression, K Neighbors Classifier (KNN), Decision Tree Classifier, Deep Neural Network, Genetic Algorithm. The results showed that deep neural network algorithms and multilayer perceptron-genetic algorithms get 96% accuracy, logistic regression algorithms have 96% accuracy, then KNN with 94%, and decision tree classifier with 92%.

ARTICLE INFO

Article History:
Submitted/Received 27 Mar 2023
First Revised 12 Apr 2023
Accepted 19 May 2023
First Available online 22 Jun 2023
Publication Date 01 Oct 2023

Keyword:
Decision Tree Classifier,
Deep Neural Network,
Genetic Algorithm,
KNN,
Logistic Regression.

© 2023 Universitas Pendidikan Indonesia

1 INTRODUCTION

Cancer is a disease that occurs due to the uncontrolled growth of abnormal cells. Breast cancer is the most common cancer in women [1]. Based on data from GLOBOCAN (Global Burden of Cancer), the International Agency for Research on Cancer (IARC) it is known that in 2018 there were 18.1 million new cases of cancer and 9.6 million deaths from cancer worldwide. It is estimated that annual cancer cases will increase from 18.1 million to 22 million in the next two decades. WHO estimates that by 2030 the incidence of cancer will reach 26 million people and 17 million of them will die from cancer [2]. This makes breast cancer the most common type of cancer in women after cervical cancer [3]. This is supported by statistical data from the American Cancer Society which shows that breast cancer ranks first with the highest number of cases of all types of malignant tumors (cancer) worldwide [4].

Data mining is defined as the process of extracting or mining the required knowledge from large amounts of data. In the process, data mining will extract valuable information by analyzing the existence of certain patterns or relationships from large data. Data mining is related to other fields of science, such as Database Systems, Data Warehousing, Statistics, Machine Learning, Information Retrieval, and High Level Computing. In addition, data mining is supported by other sciences such as Neural Networks, Pattern Recognition, Spatial Data Analysis, Image Databases, Signal Processing [5].

Classification is included in supervised learning because it uses a set of data to be analyzed first, then the pattern from the results of the analysis is used to classify the test data. The data classification process consists of learning and classification. In learning training data is analyzed using a classification algorithm then in classification data testing is used to ensure the level of accuracy of the classification rules used. Classification techniques are divided into five categories based on differences in mathematical concepts, namely statistical-based, distance-based, decision tree-based, neural network-based, and rule-based [6].

As a series of processes, data mining has several data processing, one of which is classification. Based on research conducted by Pangaribuan, et al regarding a comparison between algorithms C.50, SVM, KNN, Logistic Regression, and neural networks for detecting heart disease, the accuracy rate for KNN was 86.50% while logistic regression was 83.7% [7]. Research on breast cancer using data mining and machine learning has been carried out by Higa using a Decision Tree and Artificial Neural Network, the two algorithms successfully classify more than 92% of cases correctly in 10 trials. However, the Neural Network algorithm has an average level of predictive accuracy that is better (the correct classification rate is up to 95.9%) [8]. Research conducted by Derisma also uses a Neural Network with optimization using a Genetic Algorithm to obtain an accuracy of 97.24% [9].

The purpose of this study is to compare five algorithms, namely Logistic Regression, K Neighbors Classifier (KNN), Decision Tree Classifier, Deep Neural Network, Genetic Algorithm. Based on these algorithms, it will be determined which algorithm produces a better accuracy value.

2 METHODS

2.1 Dataset

The dataset used in this study is Breast Cancer Wisconsin (Diagnostic) Dataset obtained from UCI Machine Learning (see https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data). The dataset consists of 33 features and 569 data. This dataset contains patient diagnosis information and the value of breast cancer image computation results. Diagnosis consists of two classes, namely benign and malignant. A total of 357 patients were

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

diagnosed as having benign breast cancer and 212 patients were diagnosed as having malignant breast cancer. It can be seen if the target has imbalanced data.

2.2 Classification Algoritma

2.2.1 Logistic Regression

Logistic Regression is often used in medical research, especially since this procedure is readily available in statistical software packages. Logistic Regression is a statistical model that describes the relationship between the qualitative dependent variable and the independent variable [10][11][12]. Where μ is the location of the parameter (the midpoint of the curve where $p(\mu) = \frac{1}{2}$) and s is the scale of the parameter, see Equation [1].

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}} \quad (1)$$

2.2.2 K Neighbors Classifier (KNN)

KNN clusters each test sample based on the nearest neighbor k value. The cluster whose center has the minimum distance from the test sample is selected as the appropriate reduced training dataset [13][14]. The quality of the KNN classification depends on how well the closest neighbors are found. The chances of finding the exact k nearest neighbors also depend on how well the large dataset has been trimmed. There are two functions in determining the closest neighbor distance between them, see Equation [2] and Equation [3].

Euclidean:

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

Manhattan:

$$d(x,y) \sum_{i=1}^{m}|x_i - y_i| \quad (3)$$

2.2.3 Decision Tree Classifier

Decision Tree is a machine learning algorithm where each decision is described as a node [15][16]. Each decision tree is built on several branches and nodes. Each node represents a feature in the category to be classified and each subset defines a value that can be taken by the node. The way this algorithm works is by comparing the numerical features with the threshold values in each test. Roots, branches, and nodes in the decision tree are determined by the entropy value and the gini index value, see Equation [4] and Equation [5].

$$Entropy(t) = -\sum_{i=1}^{c-1} p(i|t)\ log_2 p(i|t) \quad (4)$$

$$Gini(t) = 1 - \sum_{i=0}^{c-1} [p(i|t]^2 \quad (5)$$

2.2.4 Deep Neural Network

Deep Neural Network (DNN) is one of the Deep Learning algorithms which is inspired by the human nervous system. Like the human brain, the DNN has neurons that are

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

interconnected with each other. Deep Learning consists of several layers such as input layer, hidden layer, and output layer [17][18]. This algorithm consists of 2 stages, namely Forward Pass and Backward Pass. Forward pass is the stage where the input data will pass through each neuron at each layer, starting from the input layer to the output layer. At this stage, the error value will be calculated, see Equation (6) and Equation (7).

$$dot_j = \sum_{i}^{3} w_{ji} x_i + b_j \tag{6}$$

$$h_j = \sigma(dot_j) = max(0, dot_j) \tag{7}$$

Second, there is the Backward Pass, which is the stage where the weight and bias values will be updated using the error values that have been obtained at the Forward pass stage.

2.2.5 Genetic Algorithm

Genetic Algorithm (GA) or in Indonesian it is called Genetic Algorithm is one of the algorithms whose way of working is inspired by Darwin's Theory of Evolution. GA population-based algorithm. Each solution corresponds to a chromosome and each parameter represents a gene [19][20].

2.3 Confusion Matrix

Measurement of the classification model can be seen through the confusion matrix. The confusion matrix is a special table that visualizes the performance of the classification algorithm. The confusion matrix table can be seen in **Table 1**.

Table 1. Confusion Matrix.

Prediction	Actual		
	Positive	**Negative**	**Total**
Positive	TP	FP	TP + FP
Negative	FN	FN	FN + TN
Total	TP + FN	FP + TN	TP + FP + FN + TN

From the confusion matrix, model accuracy can be obtained. Accuracy is the most frequently used indicator to measure the classification performance of a model [21]. Model accuracy can be obtained through the Equation [8]:

$$accuration = \frac{TP + TN}{(TP + FP + TN + FN)} \tag{8}$$

Where,

TP: True Positive, positive is classified as positive
TN: True Negative, negative is classified as negative
FP: False Positive, negative is classified as positive
FN: False Negative, positive is classified as negative

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

3 RESULTS AND DISCUSSION

Comparison of the accuracy of each of the algorithms, as follows in **Table 2.**:

Table 2. Accuracy comparison

Model	Accuracy
Logistic Regression	0.96
K Neighbors Classifier (KNN)	0.94
Decision Tree Classifier	0.92
Deep Neural Network	0.96
Genetic Algorithm	0.96

The computing process for each model uses Google Collab. Tests are carried out by comparing simple machine learning algorithms with deep learning algorithms. Simple machine learning algorithms include Logistic Regression, KNN, and Decision Tree. The test results show that the Logistic Regression model has a training accuracy value of 98% and a testing accuracy value of 95% as shown in **Figure 1**. Matrix evaluation shows that this model has an accuracy of 96% with a precision value of 97%, a recall of 92%, and an f1-score of 94% for patients diagnosed with breast cancer.

	precision	recall	f1-score	support
0	0.95	0.98	0.97	108
1	0.97	0.92	0.94	63
accuracy			0.96	171
macro avg	0.96	0.95	0.96	171
weighted avg	0.96	0.96	0.96	171

Figure 1. Logistic Regression model result.

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

The K Neighbors Classifier (KNN) model has a training accuracy value of 96% and a testing accuracy value of 93% as shown in **Figure 2**. Matrix evaluation shows that this model has an accuracy of 94% with a precision value of 95%, a recall of 87%, and an f1-score of 91% for patients diagnosed with breast cancer.

	precision	recall	f1-score	support
0	0.93	0.97	0.95	108
1	0.95	0.87	0.91	63
accuracy			0.94	171
macro avg	0.94	0.92	0.93	171
weighted avg	0.94	0.94	0.94	171

Figure 2. K Neighbors Classifier (KNN) model result.

The Decision Tree Classifier model has a training accuracy value of 96% and a testing accuracy value of 92% as shown in **Figure 3**. Matrix evaluation shows that this model has an accuracy of 92% with a precision value of 87%, a recall of 94%, and an f1-score of 90% for patients diagnosed with breast cancer.

	precision	recall	f1-score	support
0	0.96	0.92	0.94	108
1	0.87	0.94	0.90	63
accuracy			0.92	171
macro avg	0.91	0.93	0.92	171
weighted avg	0.93	0.92	0.92	171

Figure 3. Decision Tree Classifier model result.

Furthermore, for algorithms based on deep learning, there are Deep Neural Networks (DNN) and Genetic Algorithm (GA). The DNN algorithm uses 5 layers where 2 of them are input and output layers and the other 3 are hidden layers. ReLU is used for the activation function in each hidden layer. In addition, the output layer uses the sigmoid activation function because the dataset is a binary class. The comparison between the train size and the test size is 7:3. The training process was carried out for 100 epochs with Adam as the optimizer.

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

The results show that the DNN model has an accuracy value of 96% with a precision value of 90%, 98% recall, and 94% f1-score for patients diagnosed with breast cancer as shown in **Figure 4.**, **Figure 5.**, and **Figure 6**.

	precision	recall	f1-score	support
0	0.99	0.95	0.97	113
1	0.90	0.98	0.94	58
accuracy			0.96	171
macro avg	0.95	0.96	0.96	171
weighted avg	0.96	0.96	0.96	171

Figure 4. Deep Neural Networks (DNN) model result.

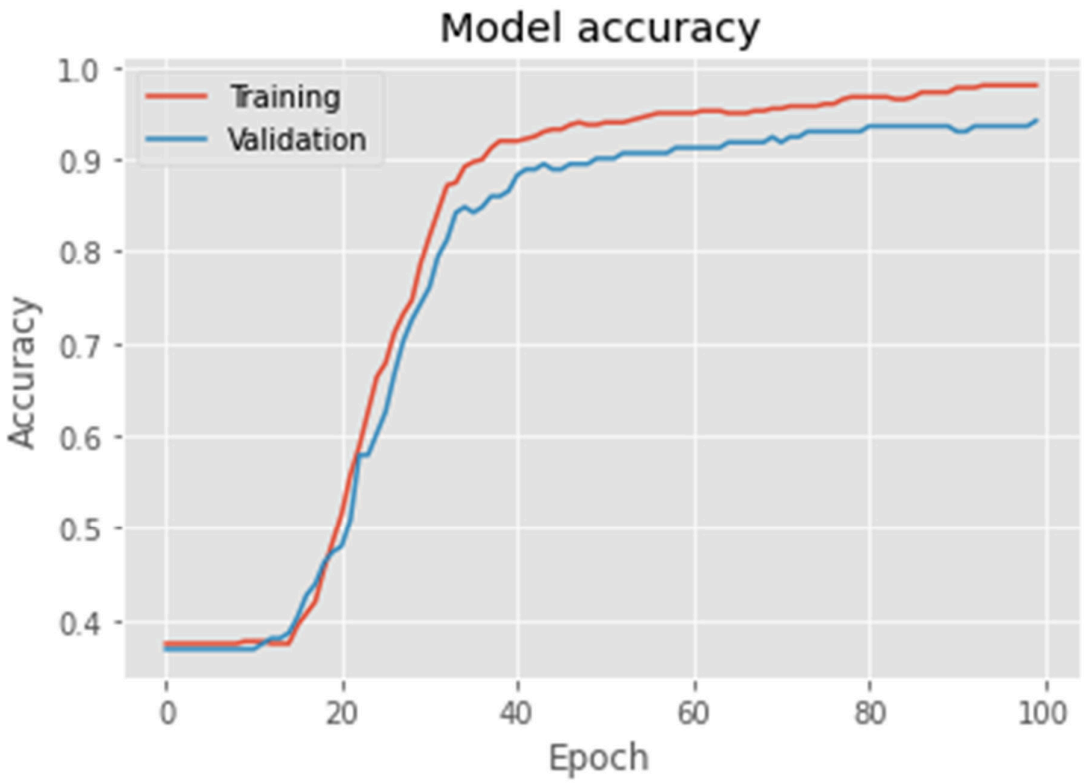


Figure 5. Graph of the accuracy of the Deep Neural Networks (DNN) model.

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

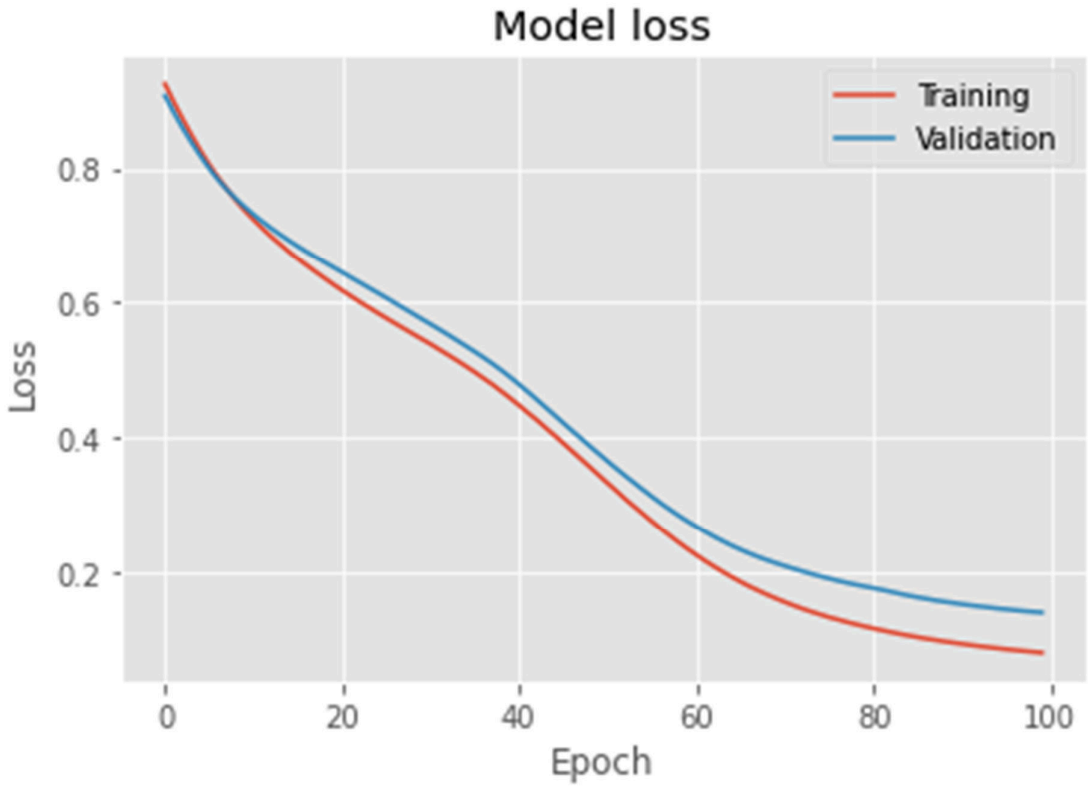


Figure 6. Graph of the loss of the Deep Neural Networks (DNN) model.

In the Genetic Algorithm, the existing data will first be classified using the MLP Classifier and Decision Tree algorithms. The classification results show that in the MLP classifier, the training process has an accuracy value of 89% and 90% in model testing. Whereas in the Decision Tree algorithm, the training process has an accuracy value of 100% and 94% when testing the model. After doing the classification, then implement the GA. The specified population size is 100, meaning that the generation to be used is half of the population, namely 50. After retraining, it was found that when adding GA as the optimizer of MLP and Decision Tree, the accuracy rate will increase to 96%.

4 CONCLUSION

The conclusions obtained in this study are that 5 (five) algorithms are used, namely logistic regression, KNN, Decision tree classifier, deep neural network, and genetic algorithm using the Breast Cancer Wisconsin (Diagnostic) Dataset obtained from UCI Machine Learning. The research results obtained that deep learning classification algorithms have a tendency to obtain high accuracy. The deep learning algorithms in question are deep neural networks and multilayer perceptron-genetic algorithms which get 96% accuracy. Meanwhile, the machine learning algorithm that has the highest accuracy is the logistic regression algorithm, which is 96%, then KNN with 94%, and decision tree classifier with 92%.

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

5 AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

6 REFERENCES

[1] Cahyanti, D., Rahmayani, A., dan Husniar, S. A. (2020). Analisis performa metode KNN pada dataset pasien pengidap kanker payudara. *Indonesian Journal of Data and Science*, *1*(2), 39-43.

[2] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209-249.

[3] Ma'arif, F., dan Arifin, T. (2017). Optimasi fitur menggunakan backward elimination dan algoritma SVM untuk klasifikasi kanker payudara. *Jurnal Informatika*, *4*(1), 46-53.

[4] Mattiuzzi, C., and Lippi, G. (2019). Current cancer epidemiology. *Journal of Epidemiology and Global Health*, *9*(4), 217-222.

[5] Ikotun, A. M., Almutari, M. S., and Ezugwu, A. E. (2021). K-means-based nature-inspired metaheuristic algorithms for automatic data clustering problems: recent advances and future directions. *Applied Sciences*, *11*(23), 1-61.

[6] Sartika, D., dan Indra, D. (2017). Perbandingan algoritma klasifikasi naive bayes, nearest neighbour, dan decision tree pada studi kasus pengambilan keputusan pemilihan pola pakaian. *Jurnal Teknik Informatika dan Sistem Informasi*, *1*(2), 151–161.

[7] Pangaribuan, J. J., Tanjaya, H., dan Kenichi, K. (2021). Mendeteksi penyakit jantung menggunakan machine learning dengan algoritma logistic regression. *Journal Information System Development (ISD)*, *6*(2), 1-10.

[8] Higa, A. (2018). Diagnosis of breast cancer using decision tree and artificial neural network algorithms. *International Journal of Computer Applications Technology and Research*, *7*(1), 23–27.

[9] Rane, N., Sunny, J., Kanade, R., and Devi, S. (2020). Breast cancer classification and prediction using machine learning. *International Journal of Engineering Research and Technology*, *9*(2), 576-580.

[10] Wang, Q. Q., S. C. Yu, Xiao Qi, Y. H. Hu, W. J. Zheng, J. X. Shi, and H. Y. Yao. (2019). Overview of logistic regression model analysis and application. *Chinese Journal of Preventive Medicine, 53*(9), 955-960.

[11] Silva, J. C. F., Teixeira, R. M., Silva, F. F., Brommonschenkel, S. H., and Fontes, E. P. (2019). Machine learning approaches and their current application in plant molecular biology: a systematic review. *Plant Science*, *284*(1), 37-47.

[12] Senaviratna, N. A. M. R., and A Cooray, T. M. J. (2019). Diagnosing multicollinearity of logistic regression model. *Asian Journal of Probability and Statistics*, *5*(2), 1-9.

[13] Saadatfar, H., Khosravi, S., Joloudari, J. H., Mosavi, A., and Shamshirband, S. (2020). A new K-nearest neighbors classifier for big data based on efficient data pruning. *Mathematics*, *8*(2), 1-12.

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

[14] Alfeilat, H. A. A., Hassanat, A. B., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Salman, H. S. E., and Prasath, V. S. (2019). Effects of distance measure choice on k-nearest neighbor classifier performance: a review. *Big Data*, *7*(4), 221-248.

[15] Charbuty, B., and Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends, 2*(01), 20-28.

[16] Cichy, R. M., and Kaiser, D. (2019). Deep neural networks as scientific models. *Trends in Cognitive Sciences*, *23*(4), 305-317.

[17] Montavon, G., Samek, W., and Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*(1), 1-15.

[18] Patra, T. K., Meenakshisundaram, V., Hung, J. H., and Simmons, D. S. (2017). Neural-network-biased genetic algorithms for materials design: evolutionary algorithms that learn. *ACS Combinatorial Science*, *19*(2), 96-107.

[19] Katoch, S., Chauhan, S. S., and Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, *80(1)*, 8091-8126.

[20] Purwa, T. (2019). Perbandingan metode regresi logistik dan random forest untuk klasifikasi data imbalanced (studi kasus: klasifikasi rumah tangga miskin di kabupaten Karangasem, Bali tahun 2017). *Jurnal Matematika, Statistika dan Komputasi*, *16*(1), 58-73.

DOI: https://doi.org/10.17509/coelite.v2i2.59717
p- ISSN 2829-4157 e- ISSN 2829-4149

Journal of Computer Engineering, Electronics and Information Technology (COELITE)

Journal homepage: https://ejournal.upi.edu/index.php/COELITE

Decision Support System for Selection of Productive at Sahabat Sampulo Foundation Using the Profile Matching Method

Nia Nandara *, *R. Rahman, Adhy Rizaldy*

Department of Information System, Universitas Islam Negeri Alauddin Makassar, Indonesia
Correspondence: E-mail: 60900117037@uin-alaudin.ac.id

ABSTRACT

The Sahabat Sampulo Foundation is dedicated to supporting and empowering underprivileged individuals by providing them with opportunities to engage in productive activities. However, the process of selecting the most suitable individuals for various productive projects can be challenging due to the large number of applicants and the diverse range of skills required. To address this issue, a Decision Support System (DSS) utilizing the Profile Matching Method is proposed. Making decisions is one of the most basic things in everyday life, in the decision-making process humans are often faced with many alternatives to choose from, so that in a problem, several decision makers can make different decisions. The advancement of this technology has also been put to good use by the Sahabat Sampulo Foundation to determine the selection of productive and unproductive employees. Because it still uses traditional methods subjectively and manually, from these problems a system is needed that can help determine employee data decisions for the Sahabat Sampulo Foundation. This Profile Matching method compares the value and actual data of a profile that will be assessed with the profile value that is applied.

© 2023 Universitas Pendidikan Indonesia

ARTICLE INFO

Article History:
Submitted/Received 8 Apr 2023
First Revised 18 Apr 2023
Accepted 22 May 2023
First Available online 19 Jun 2023
Publication Date 01 Oct 2023

Keyword:
Decision Making System,
Employee,
Profile Matching,
Selection,
Waterfall.

1. INTRODUCTION

The discipline of human resources or employees is one of the many factors that can affect the performance of a company or foundation. Human resource discipline is formed by recording experiences and lessons from various repeated deficiencies in the process of training employee discipline [1].

At the moment, the Sahabat Sampulo Foundation has various problems experienced by several companies and the Sahabat Sampulo Foundation in determining the best employees, the determination is made not only by direct appointment by management and not by means of subjective assessment, but a company must evaluate employee performance in certain period [2].

In addition, the condition that is a problem at the Sahabat Sampulo Foundation in determining the best employee lies in the evaluation process which is still carried out manually. According to [3]. This makes it difficult for companies to determine the eligibility of the best employee [3]. Furthermore, the problem of subjectivity in the selection of productive and unproductive employees is an almost unavoidable problem. Management and employees need a performance appraisal process that is accurate and fast and can provide direct feedback and improvements in the work environment [4]. One of the factors that need serious attention for managers of a company or agency is decision making. To make a decision, careful analysis and calculations are needed, depending on the number of criteria that affect the problems that require a decision to be interrelated [5].

Along with the times, system development efforts have increased in terms of complexity and integration. The system can not only save 10%–20% of the project budget, but also prevent half of all failures [6]. In assessing human resources or employees, companies are starting to follow Information Technology (IT) such as Decision Support Systems (SPK) with the aim that the assessment carried out by the company will suit the company's needs [7].

To overcome some of the problems that have been described by researchers, this study aims to use a Decision Support System (DSS) with the Profile Matching method. The use of this method aims to make it easier for the Sampulo Friends Foundation (YSS) to select productive and unproductive employees.

The reason for using the Profile Matching method is because Profile Matching is an appropriate method for making decisions that is closely related to the value of job performance and competence because the calculations are carried out by weighting and calculating competency gaps [9]. Researchers chose the Profile Matching method compared to other methods in decision support systems because the Profile Matching method considers logical consistency in the assessment used to determine priorities so as to produce alternatives that are not many at the same time have a short time and have better accuracy compared to other methods [9].

2. METHODS

This chapter explains the approach and steps used in the research to develop a Decision Support System for the Selection of Productive Individuals at Sahabat Sampulo Foundation using the Profile Matching method. This section consists of the following sub-sections: types and research location, research approach, data sources, data collection method, data processing and analysis methods, system testing techniques, and system development method.

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

2.1. Types and Research Locations

This research was conducted using this type of research. This research implemented a solution to the problem of selecting employees with a decision support system, namely using the Profile Matching Method. This research was conducted at the Sampulo Friends Foundation (YSS) which is located at Jl. Sultan Dg. No. 4 Malimongan Baru, Bontoala District, Makassar City, South Sulawesi.

2.2. Research Approach

This study employed a comprehensive research approach by combining quantitative and qualitative methodologies. The quantitative approach involved collecting and analyzing numerical data to examine the impact of decision support system variables on identifying productive and unproductive employees at the Sahabat Sampulo Foundation (YSS) using the Profile Matching method. The qualitative approach involved gathering and analyzing non-numerical data, such as interviews and observations, to gain deeper insights into contextual factors and individual perspectives. By utilizing both approaches, this study aimed to provide a holistic understanding of the relationship between decision support system variables and employee productivity at YSS.

2.3. Data Source

The source of the data in this study is using library research which is a way of collecting data from several books, journals, theses, which are related to the intent of the management information system which can be used as a reference for discussing this issue. Linkage to online or internet data sources as well as results from previous research as reference material for future researchers.

2.4. Method of Collecting Data

1) Interview: The interview process was conducted to obtain information in the form of: conducting interviews with the managers of the Sahabat Sampulo Foundation to obtain employee information.
2) Literature Study: In this study, the literature study was conducted by collecting data sourced from reference books, journals, papers, and theses related to information systems and website decision support systems.
3) Observation: Make direct observations of the object of research to the employees of the Sahabat Sampulo Foundation.

2.5. Data Processing and Analysis Methods

1) Data Processing: Data processing in question is by assigning an identity code and presenting in a tabulation table each data to be analyzed such as processing report data, report formats and efficiency and effectiveness data obtained from user observations in using the application.
2) Data Analysis: The analysis used is the analysis of quantitative and qualitative approaches, where the quantitative approach emphasizes testing theory through measuring variables using statistical procedures.

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

2.6. System Testing Techniques

In this study, the researchers have adopted a rigorous testing approach to evaluate the functionality and performance of each released feature. Specifically, they employed black box testing techniques to systematically assess the system's behavior and ensure its adherence to the specified requirements. Through black box testing, the researchers aimed to uncover any potential issues or bugs that may arise during user interaction with the system. The testing process involved simulating various scenarios and inputs to validate the expected outputs and identify any discrepancies or anomalies.

Furthermore, to evaluate the effectiveness and efficiency of the system, the researchers conducted detailed observations on the tasks assigned to the users. These observations aimed to gauge the user's ability to successfully perform the assigned tasks using the system's features. Additionally, the researchers recorded the time taken by the users to complete each given task, providing insights into the system's efficiency in supporting task completion. By closely observing user interactions and task performances, the researchers could identify any usability challenges, bottlenecks, or areas where further improvement was required.

The combination of black box testing and task observations allowed the researchers to gather comprehensive data on the system's performance, usability, and overall effectiveness. This data-driven approach provided valuable insights into the system's functionality, user experience, and potential areas for enhancement. Ultimately, by meticulously testing and observing the system's features, this study aimed to ensure the reliability and usability of the developed system, contributing to its successful implementation and user satisfaction.

2.7. System Development Method

In In this study, the researcher opted to utilize the Waterfall methodology as the system development method. The Waterfall model is a widely recognized and established approach that follows a sequential development process. It is characterized by distinct phases that are completed in a linear fashion, with each phase building upon the completion of the previous one.

The Waterfall model typically consists of several key stages, including requirements gathering, system design, implementation, testing, deployment, and maintenance. In the requirements gathering phase, the researcher carefully documented the system's functional and non-functional requirements, ensuring a clear understanding of the project's objectives and scope. Subsequently, the system design phase involved creating a comprehensive blueprint that outlined the system's architecture, components, and interrelationships.

Once the design phase was completed, the implementation stage commenced, where the researcher translated the system design into actual code and developed the necessary software components. Following the implementation, rigorous testing was conducted to validate the system's functionality, performance, and reliability. The testing phase involved various testing techniques, such as unit testing, integration testing, and system testing, to ensure the system met the predefined requirements.

Upon successful completion of testing, the system was ready for deployment. This phase involved the installation and configuration of the system in the intended environment, ensuring smooth integration and compatibility with existing infrastructure. Finally, the maintenance phase focused on ongoing support, bug fixing, and system updates to ensure the system's longevity and optimal performance.

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

By adopting the Waterfall model in this study, the researcher aimed to establish a structured and systematic approach to system development. This methodology allowed for a clear progression through each stage, enabling comprehensive documentation, thorough testing, and effective project management. Additionally, the Waterfall model facilitated better control over project timelines, costs, and deliverables, contributing to the successful development and implementation of the proposed system.

3. RESULTS AND DISCUSSION

This chapter provides a comprehensive discussion on the findings and analysis derived from the successful implementation of the Decision Support System (DSS) for the Selection of Productive Individuals at Sahabat Sampulo Foundation, utilizing the Profile Matching method. It delves into various sub-sections that shed light on different aspects of the implementation process. Firstly, the waterfall approach is explored, which entails a systematic and phased progression through the project stages, ensuring a well-organized and structured development process. Next, the chapter delves into the system implementation phase, highlighting the intricate steps taken to integrate the DSS into the existing infrastructure of Sahabat Sampulo Foundation, resulting in a seamless and efficient operation. Subsequently, the blackbox testing is discussed, which involved meticulously examining the functionality of the system from an end-user perspective, ensuring that it met the desired requirements and delivered the intended outcomes. Furthermore, the chapter addresses the evaluation of the system's usability through the employment of the System Usability Scale (SUS), enabling a comprehensive assessment of the user experience and satisfaction levels. Lastly, the chapter delves into the Profile Matching testing, providing an in-depth analysis of this method's effectiveness in selecting productive individuals for Sahabat Sampulo Foundation, unveiling valuable insights into the decision-making process. By delving into these sub-sections, this chapter offers a comprehensive overview of the implementation process and sheds light on the effectiveness and performance of the DSS, thus providing valuable insights for future improvements and enhancements.

3.1. Waterfall

In the development process using the waterfall methodology, multiple stages were followed to ensure a systematic and sequential approach [11]. One of the key aspects of the waterfall methodology is the release management process, which involves dividing the development into distinct releases. Throughout the development process, two releases were carried out, each representing a milestone in the project. These releases included the implementation and deployment of a functioning system, as depicted in **Figure 1**. The first release focused on the initial development and testing of core functionalities, while the second release involved further refinements and enhancements based on feedback and requirements. Each release was carefully planned and executed, with thorough testing and quality assurance measures in place [12-15]. The waterfall approach ensured a step-by-step progression in the development process, allowing for systematic and manageable releases of the system, such as running system as seen in **Figure 1.**

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

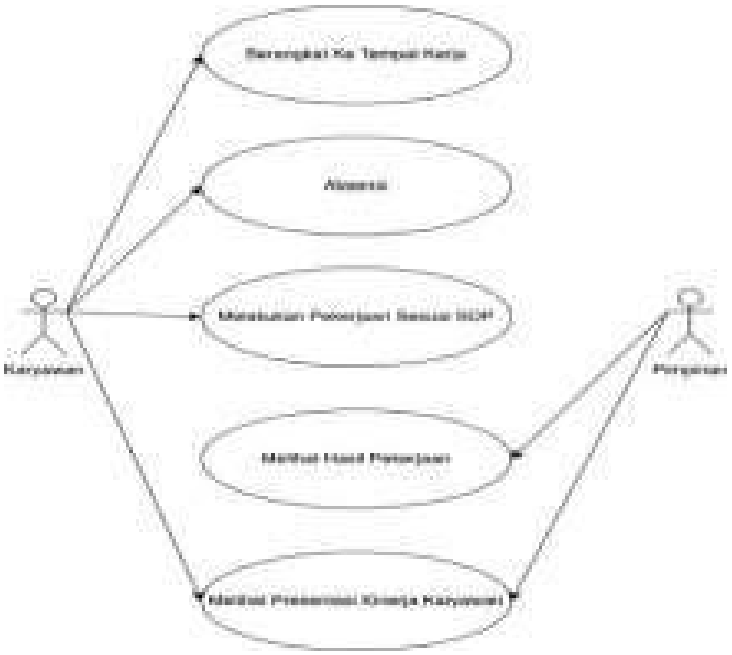


Figure 1. Running system analysis.

3.1. System Implementation

During the development process of this system, one of the crucial steps involved the creation of an intuitive and user-friendly interface. The interface design, depicted in **Figure 2 and Figure 3.**, was carefully crafted to ensure optimal usability and seamless interaction for users. Extensive attention was given to the visual elements, layout, and navigation of the interface to provide a smooth and efficient user experience. By incorporating user-centered design principles, such as clear labelling, intuitive icons, and logical information hierarchy, the interface aimed to enhance user engagement and facilitate effortless system exploration. Moreover, the interface design was informed by extensive user feedback and iterative testing, enabling the identification and resolution of any usability issues or challenges. Overall, the meticulous development of the interface played a pivotal role in ensuring the effectiveness and user satisfaction of the system. More details can be seen from **Figure 2** and **Figure 3.**



Figure 2. System interface.

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

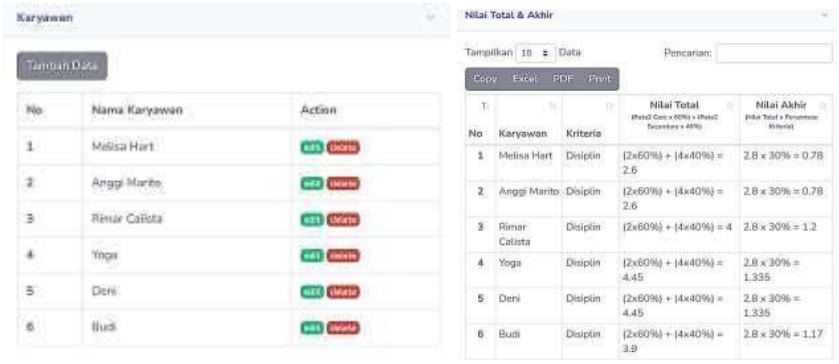


Figure 3. System interface.

3.2. BlackBox Testing

The blackbox testing method is commonly employed in software testing. To document the testing results, a blackbox table is used, facilitating systematic and structured testing. This table captures various test cases, along with their detailed information such as test scenarios, steps taken, inputs, and expected outputs (refer to **Table 1.**). By utilizing the blackbox table, the testing team can effectively track and analyze the testing outcomes, ensuring the application or system meets the predetermined requirements and specifications [16-20]. More details can be seen from **Table 1.**

Table 1. Login test.

Input Data	Expectation	Observation	Conclusion
System accessed. Edit text and login button Pressed	A login form will appear consisting of a username and password after that enter the user database validation.	Login interface. Displays a login form consisting of a username and password and can perform user validation	Accepted
Information	Display data in the form of employee information.	Displays data in the form of employee information	Accepted
Employee	Display an employee's data	Displays data in the form of employee information	Accepted
Choose to add data	Show details of adding employee data.	Successfully displays detailed data on the addition of employees	Accepted
Change analysis module.	It shows that the analysis module has registered	Shows the analysis module has registered	Accepted
Choose the delete button	Delete the selected sub criteria data.	Deletes the selected sub criteria data	Accepted
Process information	Display data in the form of process information	Displays data in the form of process information	Accepted
Select the Details button.	Show data	Displays detailed data	Accepted

3.3. SUS Testing

The decision support system software is tested through usability testing. To assess usability, the author utilizes the system usability scale (SUS), which is used to evaluate the overall usability of a system. The SUS test involves presenting users with ten questions and five alternative answer scales. These answers range from "strongly agree" to "strongly disagree" [10]. This is depicted in the following image, as shown in **Figure 4**.

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

Responden	Skor Hasil Hitung (Data Contoh)										Jumlah	Nilai (Jumlah x 2.5)
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10		
1	3	3	3	2	3	2	3	3	3	2	27	68
2	3	3	3	4	3	3	3	3	3	3	31	78
3	3	4	4	3	4	4	3	3	4	3	35	88
4	4	4	4	4	4	4	4	4	4	4	40	100
5	3	3	4	4	3	3	4	3	3	3	33	83
6	3	4	3	3	3	3	3	4	3	4	33	83
7	4	4	4	3	3	4	3	3	3	4	35	88
8	3	3	3	3	3	3	3	3	3	4	31	78
9	4	4	3	2	3	3	3	3	3	3	31	78
10	3	3	4	3	4	3	4	3	3	0	30	75
11	4	4	4	4	4	3	4	4	4	3	38	95
12	4	3	3	4	4	4	3	3	4	4	36	90
13	4	4	4	4	4	4	4	4	4	4	40	100
14	4	4	4	4	4	3	3	4	4	4	38	95
15	3	3	4	4	4	3	4	4	4	2	35	88
16	4	4	4	4	4	4	4	4	4	4	40	100
17	4	3	4	3	3	4	4	3	4	4	36	90
18	4	4	4	3	4	3	3	4	4	2	35	88
19	3	3	4	4	4	4	4	4	4	4	38	95
20	3	4	4	3	4	3	4	4	4	4	37	93
Skor Rata-rata (Hasil Akhir)												1748

Figure 4. Running system analysis.

The results of the SUS (System Usability Scale) test showed 87% and received an "Excellent" ive rating and received Grade "A". This means that in terms of usability or usability, this has been appropriate and acceptable to Una. And after being tested functional, this is also functioning properly and can produce Profile Matching calculations that are the same as manual calculations.

3.4. Profile Matching Test

The Profile Matching method is as follows:

1) Determine the criteria for the data needed.
2) Determine the aspects used for assessment

The following are the stages in using the Profile Matching Method:

1) Gap Profile Mapping. The gap referred to here is the difference/difference in the value of each aspect with the target value. The following is the profile gap mapping formula: Gap = Applicant Profile – Profile Value.
2) Weighting. After calculating the gaps for each applicant, then each employee profile is given a weighted value based on the gap value weight table.
3) Calculation and grouping of core factors and secondary factors. After determining the weight of the gap value, then each aspect is further divided into two groups, namely the core factor group and the secondary factor group. Core factors, criteria (competencies) that are most important or prominent or most needed by an assessment that is expected to obtain optimal results. Secondary factors are items other than those in the core factor.
4) Calculation of Total Value. From the calculation of the core factor and secondary factor of each aspect, the total value of each aspect is calculated which is estimated to affect

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

the performance of each profile. Calculation of the total value can be shown in the equation [1].

$$N = (X)\%\ NCF + (X)\%NSF \quad (1)$$

Information :
N : The total value of each aspect
NCF : The average value of the core factor
NSF : Average value of secondary factor
(X)% : The percentage value that is inputted

5) Rank Determination Calculation The end result of the Profile Matching process is ranking. Determination of ranking refers to the results of certain calculations. Calculation of the total value can be shown in the equation [2].

$$Ranking = (x)\%\ Npd + (x)\%Nu + (x)\%Npg \quad (2)$$

Information :
Ranking : Final Result
Npd : Value of Educational Aspects
Nu : Age Aspect Value
Npg : Value Aspects of Experience
(X)% : The percentage value entered.

For example :
Unsatisfactory = 0
Needs repair = 1
Meet expectations = 2
Exceeded expectations = 3
Awesome = 4

Nama	CF 20% Loyalitas			SF 20% Kreatifitas			SF 20% Prestasi			CF 20% Komitmen			CF 20% Disiplin		
A	0	3	5.0	0	3	5.0	0	3	5.0	0	3	5.0	-1	4	4.0
B	1	2	4.5	1	2	4.5	0	3	5.0	0	3	5.0	0	3	5.0
C	-1	4	4.0	-1	4	4.0	0	3	5.0	0	3	5.0	1	2	4.5
D	1	2	4.5	-1	4	4.0	1	2	4.5	1	2	4.5	1	4	4.5
E	0	3	5.0	1	2	4.5	0	3	5.0	0	3	5.0	1	4	4.5
Rata-Rata		3			3			3			3			3	

Nilai Konversi GAP: *Core Factor*

Figure 5. Core factor.

GAP Conversion Value: Core Factor
A = (4.0 + 5.0 + 5.0) / 3 = 4.6
B = (5.0 + 5.0 + 4.5) / 3 = 4.8
C = (4.5 + 5.0 + 4.0) / 3 = 4.5
D = (4.5 + 5.0 + 4.5) / 3 = 4.6

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

E = (4.5 + 5.0 + 5.0) / 3 = 4.8

GAP Conversion Value: Secondary Factor
A = (5.0 + 5.0) / 2 = 5
B = (5.0 + 4.5) / 2 = 4.75
C = (5.0 + 4.0) / 2 = 4.5
D = (4.5 + 4.0) / 2 = 4.25
E = (5.0 + 4.5) / 2 = 4.75

Table 2. Secondary factor.

Nama	CF + SF	Nilai Akhir	Hasil	Peringkat
A	(4.6*60%) * (5.0*40%)	2.76 + 2	4.76	2
B	(4.8*60%) * (4.75*40%)	2.88 + 1.9	4.78	1
C	(4.5*60%) * (4.5*40%)	2.7 + 1.8	4.5	3
D	(4.6*60%) * (4.25*40%)	2.76 + 1.68	4.44	4
E	(4.8*60%) * (4.75*40%)	2.88 + 1.9	4.78	1

4. CONCLUSION

The conclusions of the thesis research entitled "Decision Support System for Selection of Productive and Non-productive Employees at Sahabat Sampulo Foundation Using the Profile Matching Method" are as follows:

1) The researcher conducted an interview, the researcher concluded the user's needs based on the running system problems, the expected interface, and the desired output or output by the Sahabat Sampulo Foundation (YSS) to facilitate the selection of productive and unproductive employees. The needs of the Sahabat Sampulo Foundation include: The system that will be created is expected to have an interface that is easily understood by users or user experience. The system that will be created can provide the results of selecting productive and unproductive employees effectively and efficiently. The system that will be created can provide results that can be relied upon by Managers in selecting productive and unproductive employees.
2) Based on the results of the questionnaire, it can be concluded that 80% of the presentation of concise and direct information is achieved with an interface that is easy to use and understand and predictable with a percentage of 60%. The navigation mechanism runs well with a percentage rate of 40-80%, and with the application function running properly with a percentage result of 80%.
3) In the Profile Matching Method, all criteria have a specified weight. So that it is likely to have an influence on the assessment therein.
4) The test results of the System Usability Scale (SUS) on this system are 87.4 with grade "A", this shows that this system is feasible to use and can facilitate solving the problem of selecting productive and unproductive employees at the Sahabat Sampulo Foundation.

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

6. REFERENCES

[1] Wang, Z. (2021). Human resource management mode based on big data. *Journal of Physics: Conference Series, 1852*(2), 01-08.

[2] Sari, J. P., dan Yusa, M. (2020). Penentuan karyawan terbaik pada collection PT. Panin Bank menggunakan metode smart. *seudocode*, *7*(2), 157-164.

[3] Munawir, M., dan Ardiansyah, A. (2017). Decision support system pemilihan karyawan berprestasi dengan pendekatan analisa gap profile matching di kantor perwakilan bank indonesia provinsi aceh. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *1*(1), 7-14.

[4] Sudrajat, B. (2018). Pemilihan pegawai berprestasi dengan menggunakan metode profile mathcing. *Jurnal dan Penelitian Teknik Informatika, 3*(1), 202–210.

[5] Vargas, I. G., and Braga, R. T. V. (2020). Understanding system of systems management: A systematic review and key concepts. *IEEE Systems Journal*, *16*(1), 510-519.

[6] Prasetiyo, B., Baroroh, N., and Rufiyanti, D. E. (2016). Fuzzy simple additive weighting method in the decision making of human resource recruitment. *Lontar Komput. J. Ilm. Teknol. Inf*, *7*(3), 174-181.

[7] Purwanto, H. (2017). Penerapan metode profile matching dalam sistem pendukung keputusan penilaian kinerja karyawan pada Pt. Hyundai mobil indonesia cabang kalimalang. *J. Techno Nusa Mandiri, 16*(1), 34–40.

[8] Muhammad, M., Safriadi, N., dan Prihartini, N. (2017). Implementasi metode simple additive weighting (SAW) pada sistem pendukung keputusan dalam menentukan prioritas perbaikan jalan. *Jurnal Sistem dan Teknologi Informasi (JUSTIN), 5*(4), 223-228.

[9] Nioga, A., Brata, K. C., dan Fanani, L. (2019). Evaluasi usability aplikasi mobile KAI access menggunakan metode system usability scale (SUS) dan discovery prototyping (Studi Kasus PT KAI). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *3*(2), 1396-1402.

[10] Kuswanto, J. (2020). Penerimaan karyawan baru menggunakan metode profile matching. *Jurnal Processor, 15*(2), 85-97.

[11] Lucitasari, D. R., and Khannan, M. S. A. (2019). Designing mobile alumni tracer study system using waterfall method: an android based. *International Journal of Computer Networks and Communications Security, 7*(9), 196-202.

[12] Saputro, H., Wahyudin, I., and Iskandar, A. (2020). Development of a web-based online library information system at the national university using the waterfall method: development of a web-based online library information system at the national university using the waterfall method. *Jurnal Mantik*, *4*(1), 300-310.

[13] Suhirman, S., Hidayat, A. T., Saputra, W. A., and Saifullah, S. (2021). Website-based e-pharmacy application development to improve sales services using waterfall method. *International Journal of Advances in Data and Information Systems*, *2*(2), 114-129.

[14] Aroral, H. K. (2021). Waterfall process operations in the fast-paced world: project management exploratory analysis. *International Journal of Applied Business and Management Studies*, *6*(1), 91-99.

[15] Mahadevan, L., Kettinger, W. J., and Meservy, T. O. (2015). Running on hybrid: control changes when introducing an agile methodology in a traditional “waterfall” system development environment. *Communications of the Association for Information Systems, 36*(1), 5.

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

[16] Jaya, T. S. (2018). Pengujian aplikasi dengan metode blackbox testing boundary value analysis (studi kasus: kantor digital Politeknik Negeri Lampung). *Jurnal Informatika: Jurnal Pengembangan IT*, *3*(1), 45-48.

[17] Febiharsa, D., Sudana, I. M., dan Hudallah, N. (2018). Uji fungsionalitas (blackbox testing) sistem informasi lembaga sertifikasi profesi (SILSP) batik dengan appperfect web test dan uji pengguna. *Joined Journal (Journal of Informatics Education)*, *1*(2), 117-126.

[18] Hidayat, T., dan Muttaqin, M. (2018). Pengujian sistem informasi pendaftaran dan pembayaran wisuda online menggunakan black box testing dengan metode equivalence partitioning dan boundary value analysis. *Jutis (Jurnal Teknik Informatika)*, *6*(1), 25-29.

[19] Supriyono, S. (2020). Software testing with the approach of blackbox testing on the academic information system. *IJISTECH (International Journal of Information System and Technology)*, *3*(2), 227-233.

[20] Munthe, I. R., Rambe, B. H., Pane, R., Irmayani, D., and Nasution, M. (2020). UML Modeling and Black Box Testing Methods in the School Payment Information System. *Jurnal Mantik*, *4*(3), 1634-1640.

DOI: https://doi.org/10.17509/coelite.v2i1.57092
p- ISSN 2829-4157 e- ISSN 2829-4149

Journal of Computer Engineering, Electronics and Information Technology (COELITE)

Journal homepage: https://ejournal.upi.edu/index.php/COELITE

Teacher and Student Attendance System at Noor Faqih Usman Foundation Based on RFID Integrated with Raspberry Pi

Dhimaz Purnama Adjhi [*], Mohamad Rizal Hanafi, Rastra Wardana Nanditama, Rifqi Alamsyah, Hafidz Rizki Fahriza, Anugrah Adiwilaga

Departement of Computer Engineering, Universitas Pendidikan Indonesia, Indonesia
Correspondence: E-mail: dhimaz@upi.edu

ABSTRACT

In educational institutions, tracking attendance is crucial for ensuring effective administration and student engagement. This abstract presents the development of a Teacher and Student Attendance System at the Noor Faqih Usman Foundation, leveraging Radio Frequency Identification (RFID) technology integrated with Raspberry Pi. The proposed system aims to automate the attendance recording process, streamline administrative tasks, and enhance the overall efficiency of attendance management. This can create gaps in the development of data manipulation, data frameworks, and rigid systems. The purpose of this study was to create an RFID-Based Teacher and Student Attendance System integrated with Raspberry Pi. This study was conducted using a quantitative method of study and development study. This RFID-based attendance technology will later replace the role of paper to record the attendance of teachers and students using cards / keychains to make it easier to report attendance, data can be stored in digital form such as excel, can be accessed via wireless smartphones, interested parties can monitor through the website, and the educational institution will be touched by at least technology. The results showed that the system was able to control the attendance process and succeeded well.

ARTICLE INFO

Article History:
Submitted/Received 8 Apr 2023
First Revised 18 Apr 2023
Accepted 22 May 2023
First Available online 16 Jun 2023
Publication Date 01 Oct 2023

Keyword:
Attendance System,
RFID,
Raspberry Pi,
Efficiency,
Web Server.

© 2023 Universitas Pendidikan Indonesia

1. INTRODUCTION

Alongside the fourth industrial revolution, the education system in Indonesia must inevitably adapt to this continuously evolving digital era. It is not only the education system that needs to adapt, but also teachers and students must utilize existing technology to the best of their abilities in order to facilitate their tasks. One of the tasks that many teachers still perform traditionally is attendance, which is mandatory before starting a lesson at school.

The attendance system used in the Indonesian education world, especially in schools located far from urban areas, still relies heavily on paper sheets. This paper-based attendance system has many shortcomings [1]. One of the issues with this system is the excessive use of paper, which indirectly impacts the environment and may contribute to global warming since paper is made from trees [2]. Additionally, excessive paper usage can lead to a pile-up of documents, making it difficult to find specific files when needed. Furthermore, a paper-based system is prone to data redundancy, resulting in more paper usage and excessive costs to purchase paper [1]. This system is also highly vulnerable to data manipulation by unauthorized individuals [3]. Additionally, data storage in a paper-based system is susceptible to loss and damage.

In the midst of current technological advancements, there are already many ways to store data without relying on paper. A paperless document management system that exists today can simplify the documentation process itself, including the attendance system that is still paper-based. With a paperless system, student attendance data can be better organized, improved, cost-effective, and indirectly reduce paper usage, which may have a positive impact on the environment.

One way to digitize the attendance system is by utilizing Radio Frequency Identification (RFID) sensor technology. RFID is a technique used to identify objects using radio waves. RFID involves two devices in its process: a tag and a reader. The usage of RFID is straightforward—users simply attach the tag to the reader, and the reader will identify the tag. Each tag has a unique ID number, ensuring that no tags have the same ID as others. RFID technology is suitable for fast and automated operational activities.

With RFID technology, the attendance system becomes much easier and faster. Each student and teacher will have a tag with a unique ID representing their identity. Students and teachers can independently record their attendance by attaching their respective tags to the readers located outside the classroom or in designated areas [4][5]. The inputted data from the tags will be processed and stored in a database, ensuring more accurate and well-organized data compared to manual attendance, which is prone to human errors [4]. With RFID technology, teachers no longer need to call out each student's name for attendance, and relevant staff can easily monitor teacher attendance.

The RFID-based attendance system will have many positive effects, benefiting both teachers and students in schools [6]. Firstly, since all data will be digitally stored in a database, it reduces the foundation's expenses for purchasing paper and/or books solely used for student attendance records. Additionally, attendance can be taken much faster as it is done by the students themselves, and indirectly, the RFID-based attendance system introduces the extent of current technological advancements to underage students [7]. Since the data is stored in a database, data storage becomes more structured, and data security is significantly improved compared to manual attendance records on paper [8][9].

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

2. METHODS

This research was conducted through direct observation at Noor Faqih Usman Foundation using a quantitative survey method. The research involved data collection at the relevant institution and interviews with the Chairman of the Foundation. The data obtained from the observation and interview at Noor Faqih Usman Foundation are as follows:

1. Redundancy or overlapping of attendance data has occurred.
2. The foundation's head cannot control which teachers enter.
3. Attendance data needs to be stored in a digital format such as Excel.
4. Teachers or staff at the institution do not have laptops or computers.
5. There are three busy session schedules: early childhood education in the morning, tutoring in the afternoon, and children's religious study in the evening.

After the data was collected, the research proceeded to the next stage, which is the development phase. This research focuses on the goal of developing, expanding, and further exploring the attendance system based on a theory within the discipline of technology. The flowchart of this research method is shown in **Figure 1**.

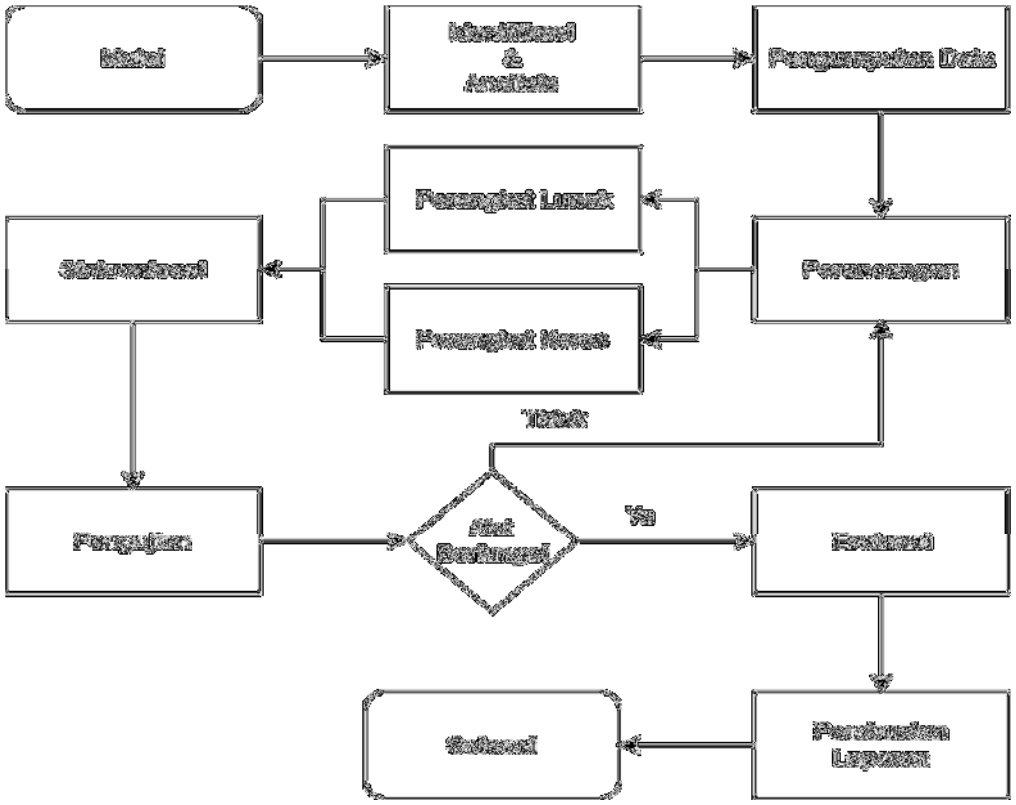


Figure 1. Research Method Flowchart.

In the design of the teacher and student attendance system at Noor Faqih Usman Foundation, this institution utilizes a hardware system consisting of Arduino Uno microcontroller, NodeMCU ESP8266 microcontroller [10], Raspberry Pi 2, RFID Module, LCD I2C, Access Point, Mini Buzzer Alarm, and two LEDs (green and red). The architectural diagram illustrating the hardware components in this system is shown in **Figure 2**.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

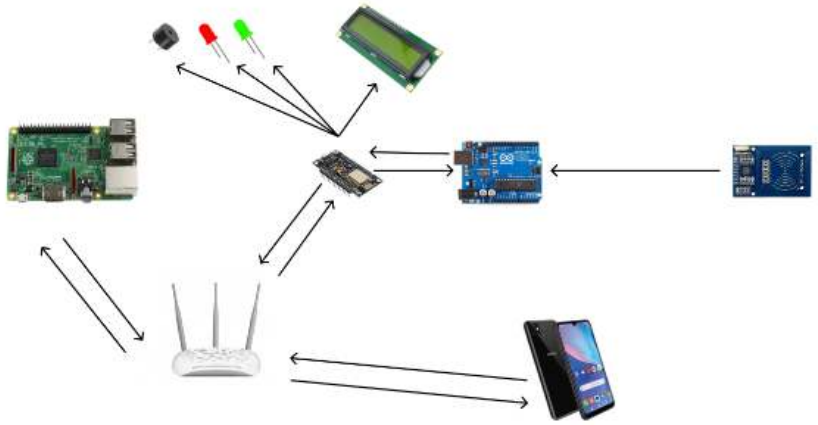


Figure 2. System Architecture Diagram.

This system consists of several components, each with its own function, and there are relationships between these components. The relationships between the components can be depicted through a system block diagram, with each component represented by a block. The block diagram of the teacher and student attendance system can be illustrated as shown in **Figure 3**.

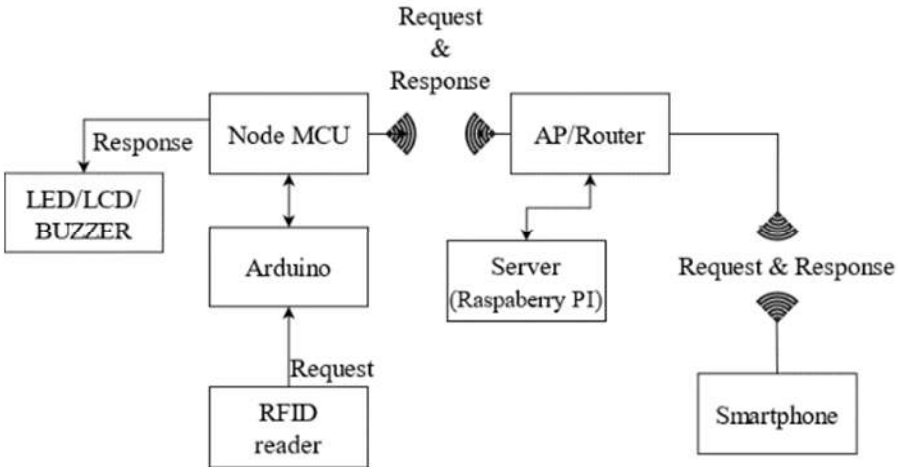


Figure 3. System Block Diagram.

The block diagram shown in **Figure 3.** represents the basic overview of the system design. A detailed explanation of the above Block Diagram is as follows:

2.1 LED/LCD/BUZZER

LED, LCD, and buzzer are devices that display and provide notifications to indicate interactions between the system and users.

2.2 RFID Reader

The RFID reader is responsible for reading information from RFID cards. This information will be used to fill in parameters or queries when making requests to the server. The RFID reader module plays a crucial role when users interact with the system.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

2.3 Arduino Microcontroller

Arduino is used to receive data read by the RFID reader module. The received string data will be processed or assembled to make requests through NodeMCU.

2.4 NodeMCU Microcontroller

NodeMCU is used to identify or retrieve data from Arduino for processing. NodeMCU is also used to transfer GET requests to the server, and because it has a WiFi module, it can be used for GET requests to the Raspberry Pi server.

2.5 AP/Router

Since the built system is a local system, which means the attendance system is not connected to the internet, an AP/Router is used as a provider of local network access. Data can be forwarded from NodeMCU to the Raspberry Pi server.

2.6 Server (Raspberry Pi)

In this attendance system, Raspberry Pi is used as the server [11], which serves as the database storage used by users [12]. Requests sent from the NodeMCU microcontroller or smartphone will be received by the server. After the server receives the request, the response will be sent back to the requester. This response contains the requested components or information.

2.7 Smartphone

Users need a device to communicate with the system, and one of the devices is a smartphone. This device is used to make requests to the server, allowing users to interact with the system. The smartphone functions as a platform to view user data and attendance data through the browser. Since the system runs on a local network, the smartphone needs to be connected to the same network to use all the features of the attendance system web app.

The system flow diagram in **Figure 4.** shows how the interaction is performed with the system, starting from the card being read by the RFID reader, entering the predefined process request of the Arduino microcontroller. After processing, the request will be forwarded through NodeMCU to the predefined process request of the Raspberry Pi. The request received will generate a response, which will be sent back as an answer from the server. The response will be sent back through NodeMCU and processed to display the obtained answer from the server through the buzzer, LCD, and LED.

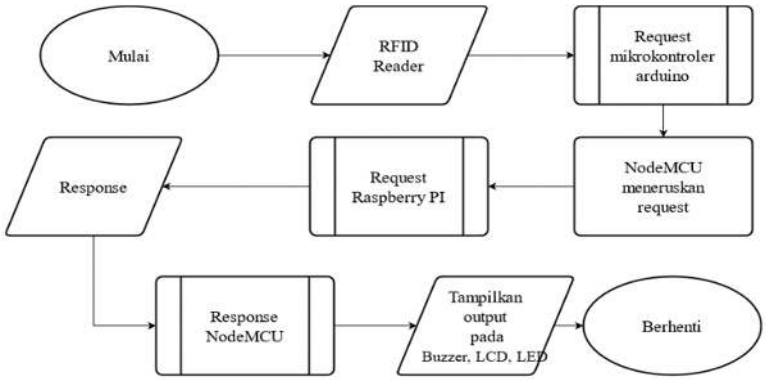


Figure 4. Flowchart of attendance system.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

As mentioned earlier, the Arduino microcontroller is the component where the predefined process request takes place. This process involves data processing, where the information obtained from the card read by the RFID reader is assembled with the Arduino microcontroller to make a request. The execution of this sequence is then passed on by the NodeMCU. The flowchart illustrating this process can be seen in **Figure 5**.

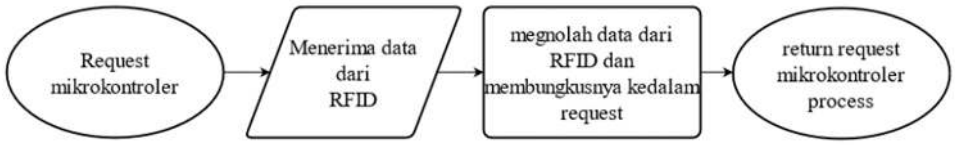


Figure 5. Flowchart of predefined process request microcontroller.

After the data from Arduino is forwarded to NodeMCU, NodeMCU will then make a request to Raspberry Pi. The data from NodeMCU's request will be processed in the predefined process on Raspberry Pi to obtain a response. The response will be sent back as a reply, containing the information requested. If the requested information is found, the response will indicate success along with the answer. However, if the requested information is not found, the response will still be given but will contain information that the request cannot be processed. The flowchart illustrating this process can be seen in **Figure 6**.



Figure 6. Flowchart of predefined process Rasspberry PI.

After obtaining the response from the request by the predefined process response microcontroller, the NodeMCU microcontroller will process the response further in order to display it. The process response of the microcontroller can be illustrated in **Figure 7**.

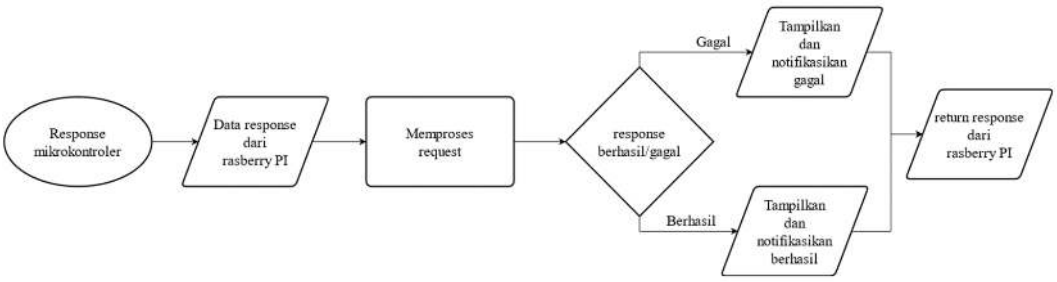


Figure 7. Flowchart of predefined process response microcontroller.

If the sent response is successful, the successful display and notification will be returned. However, if the received response fails (the requested data is not available), the failed display and notification will be returned.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

Since this system is web-based, users can access their attendance history through the browser on their smartphones. In order to access the attendance history, the smartphone's browser needs to send a request to the server. The process from the smartphone to the server to retrieve attendance data from the server can be illustrated with a flowchart in **Figure 8**.

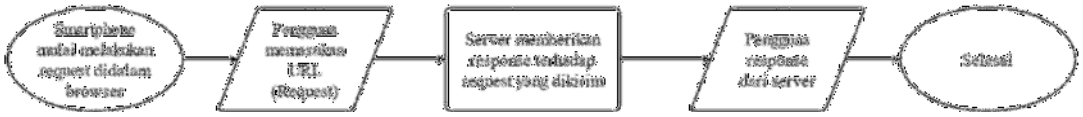


Figure 8. Flowchart diagram *smartphone* to *server*.

Smartphones and devices can communicate directly if the smartphone makes a request through the browser on the smartphone. Therefore, the requests made through the browser can be received directly by the server, and the server can also send responses directly.

The wiring diagram for the hardware components of the teacher and student attendance system can be illustrated in **Figure 9**.

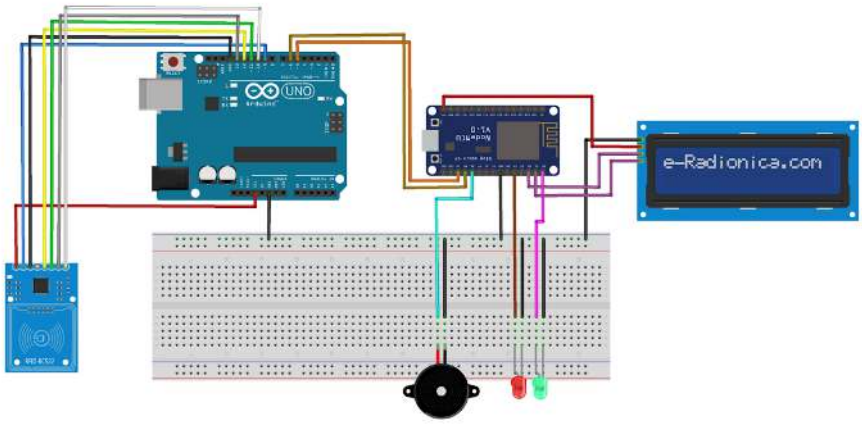


Figure 9. Hardware *Wiring* Diagram.

The conceptual design of the Entity Relationship Diagram (ERD) for the software of the teacher and student attendance system can be illustrated in **Figure 10**.

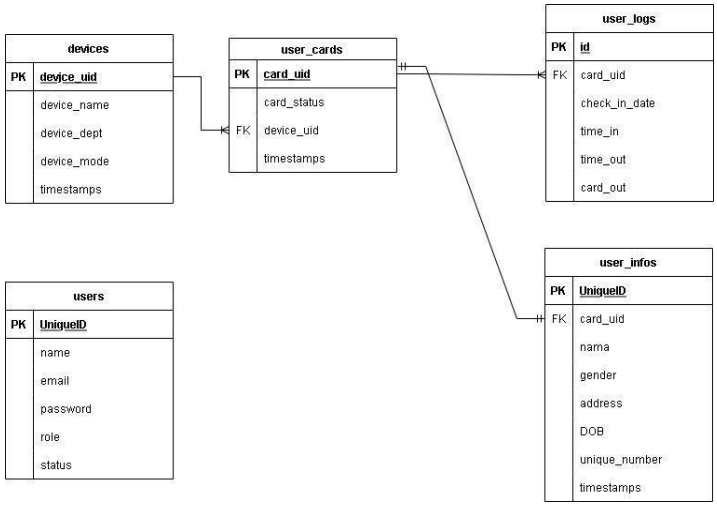


Figure 10. *Entity Relationship Diagram* (ERD).

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

The last one is the conceptual design of the RFID card/keychain and the junction box as the casing for the hardware of the teacher and student attendance system at Noor Faqih Usman Foundation. These designs can be illustrated in **Figure 11**. and **Figure 12.**



Figure 11. *card/keychain design*.

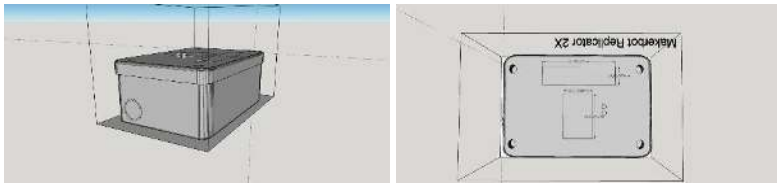


Figure 12. *junction box design as casing*.

3. RESULTS AND DISCUSSION

For the implementation, this system is divided into several menus. Each menu has different functional features and actors, which will affect the process. The use case actors in this system include:

1. User, with read-only access, can only view attendance history.
2. Staf, with access rights, can view attendance history and user card data.
3. Actor Admin, with the highest access rights, can view attendance history, user card data, and WebApp user data. They can add user cards and WebApp accounts, modify user card information and WebApp credentials, delete user cards and WebApp users, as well as create, modify, and delete RFID devices for attendance purposes.

In **Figure 13.**, it shows the interface of the login page of the WebApp. This login page is used by various actors to access the main page of the attendance system. There is a feature to create an account in **Figure 14.**, specifically designed for new users who will have their account creation requests accepted or rejected by the admin actor [13][14].



Figure 13. *Login Page WebApp*.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149



Figure 14. Creating Account Page *WebApp.*

Table 1. Login Page Description

Object	Usage	Description
Login Authentication Account	Enter the registered email and password credentials, then press "Login".	Usage rights for a valid account
Account Creation	Press "Create Account!", then fill in the account information and credentials. After that, wait for the request to be accepted and the actor's role to be assigned by the admin.	Creating a new WebApp Account

If logged in as an admin, the main menu of the web application specifically designed for admins will appear. It can be seen in **Figure 15**.

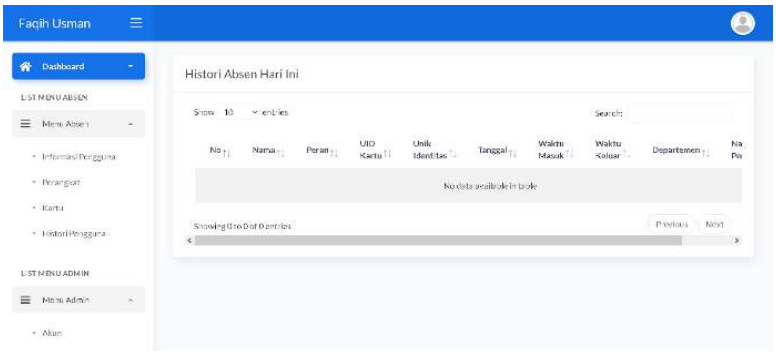


Figure 15. Main Page WebApp Admin.

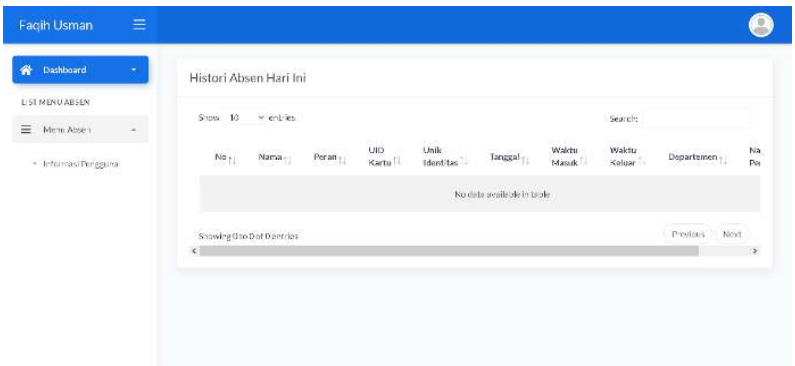


Figure 16. Main Page WebApp Staff.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

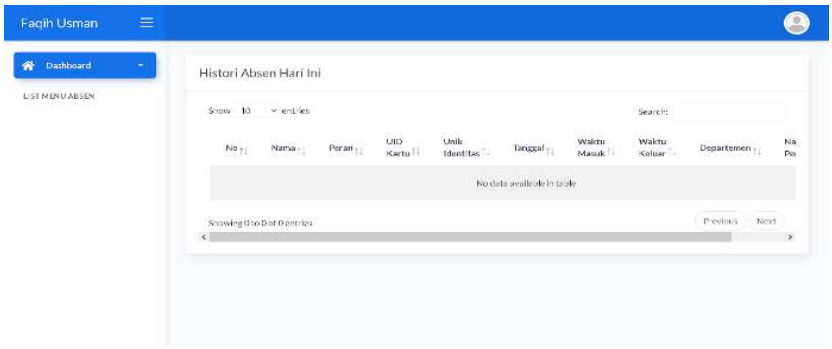


Figure 17. Main Page WebApp User.

On the homepage of the WebApp, can be seen in **Figure 16**. and **Figure 17.**, regardless of the user's access level (admin, staff, or user), a table displaying the attendance history for the current day (Dashboard) will be shown [3][8]. The main functions of the navbar are primarily for navigation to other features. To access other WebApp features, there is a dropdown menu titled “Menu Absen” (Attendance Menu) and a specific dropdown menu for admin actors titled “Menu Admin” can be sen in **Figure 18.** and **Figure 19.** Within the “Menu Absen”, several sub-dropdowns will appear to access different system features, such as “Informasi Pengguna” (User Information), “Perangkat” (Devices), “Kartu” (Cards), and “Histori Pengguna” (User History) can be seen in **Figure 20**. Meanwhile, within the “Menu Admin”, there is one sub-dropdown to access the “Akun” (Account) feature in **Figure 21**. The description of these menu shown in those figure can be seen in **Table 2., Table 3., Table 4., Table 5.** However, the usage of these features is restricted based on the user's role for security reasons [9][17][18].

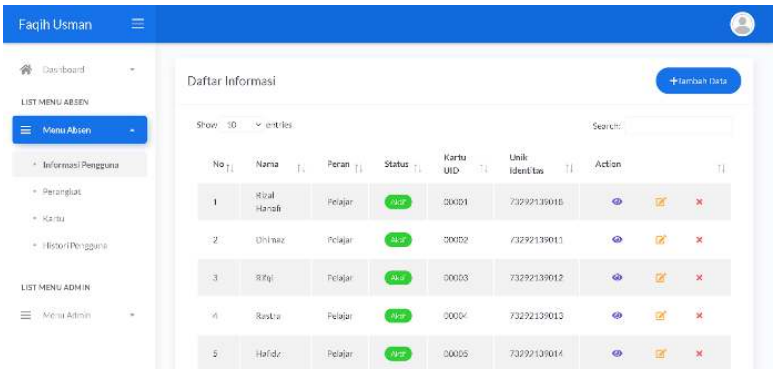


Figure 18. Information Page User on Admin.

Table 2. Description of User information page for Admin.

Object	Usage	Description
Add user data	Press the "+Add Data" button, then fill in the data for the new user and click the save button.	Register and add a new user
Edit User Data	Click the pencil icon in the "action" column and the row of the selected user.	Modify the identity of the selected user
Delete User Data	Click the cross icon in the "action" column and the row of the selected user.	Delete the selected user

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

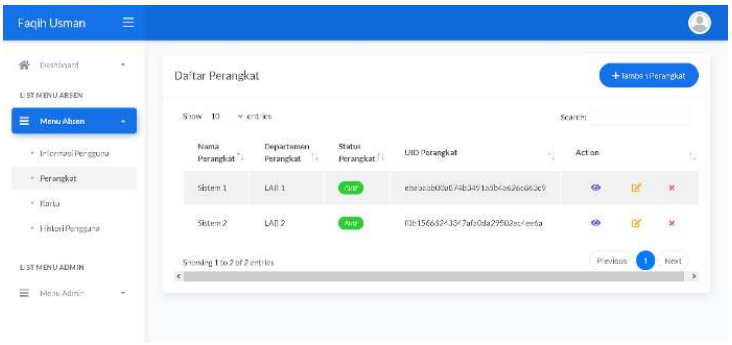


Figure 19. Page Device for Admin.

Table 3. Description of the Device Page for Actor

Object	Usage	Description
Add Device	Press the "+Add Device" button, then fill in the device name and department, and press the add button.	Register and add a new attendance device.
Edit Device	Click on the pencil icon in the "action" column and the selected device row to edit the device.	Modify the identity of the selected device.
Delete Device	Click on the cross icon in the "action" column and the selected device row to delete the device.	Delete the selected device.

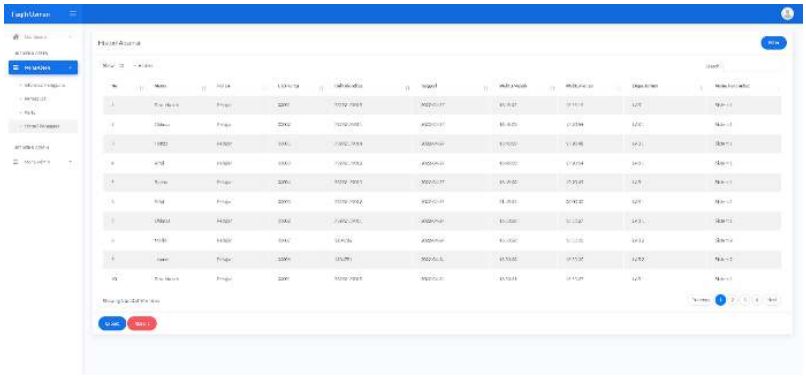


Figure 20. User History Page on Admin.

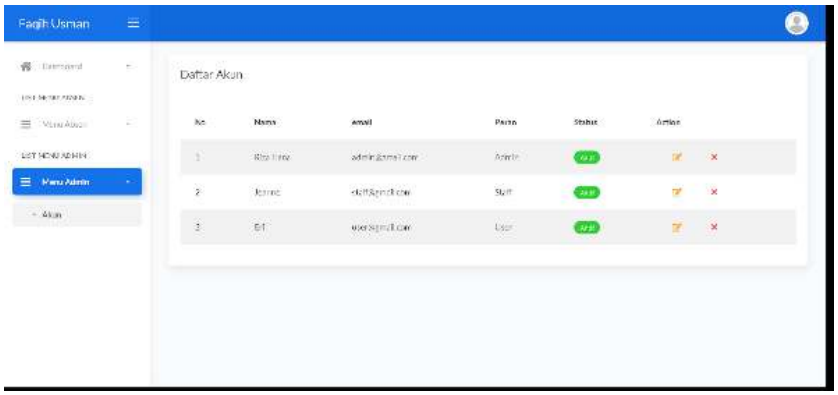


Figure 21. Account Page on Admin.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

Table 4. Description of User *History Page* on Admin.

Object	Usage	Description
Filter	Press the "Filter" button above the "search" text field and specify the parameters to filter the data.	Filter the data based on the parameters input by the user.
Search	Click on the text field above the right side of the table and enter the name to be searched.	Search for a name based on what is entered in the "search" text field.
Export	*Click on the "Export" button next to the "RESET!" button and specify the parameters to limit the data to be exported.*	Export the data according to the parameters specified by the user in .xlsx format.
Reset	Click on the red button with the title "RESET!" next to the "export" button to reset the data.	Delete all history data.

Tabel 5. Description of Account Page on Admin.

Object	Usage	Description
Delete WebApp Account	Click on the cross icon in the "action" column and the row of the account name to be deleted.	Delete the selected account.
Edit WebApp Account	Click on the pencil icon in the "action" column and the row of the selected account name.	Modify the identity of the selected account.

In the software testing of the WebApp for the Teacher and Student Attendance System, black box testing with fuzzing technique and boundary value analysis technique were employed [15][16]. Black box testing involves providing a set of inputs to the program, while fuzzing is a technique for searching for bugs in software by injecting imperfect data, and boundary value analysis focuses on identifying errors from the external or internal boundaries of the software. The specific testing scenarios in this research are divided into two parts: front-end and back-end testing, and hardware client and server architecture testing [17][18]. The testing scenarios are as follows:

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

Table 6. Results and Testing of the Attendance System

Tested Objects	Actor Action Descriptions	Results
	1. User History Page	
Tapping History Time In dan Time Out	The user attaches a card/keychain to the RFID, provided that the devices mode has been activated.	Success
Log Filter	An actor with admin access performs Log Filtering from a presence history. The filters can be based on date, user, time, and RFID device in the department room.	Success
Export Excel	*An actor with admin access performs an export of the existing presence history data from the WebApp to a digital format such as Excel.*	Success
Search	An actor with admin access performs a search of the existing presence history data on the WebApp.	Success
Reset	An actor with admin access performs a reset of the presence history data.	Success
	2. Device Page	
List Devices RFID	An actor with admin access views the list of RFID devices distributed in the rooms per department within the institution or organization.	Success
Add Device	An actor with admin access adds an RFID device for a new room. They then input the device name along with its corresponding department.	Success
Delete Device	An actor with admin access deletes the desired RFID device.	Success
Edit Device	An actor with admin access modifies a device, including the device name, room department name, and RFID device mode status. The active mode status indicates that the device is in presence history mode, while the inactive mode status indicates that the device is in user card/keychain registration mode.	Success
	3. Card Page	
Search	An actor with admin access performs a search for card/keychain user data on the WebApp.	Success
List of Registered Users	An actor with admin access views the list of registered card/keychain users, marked with an active status.	Success
List of Unregistered Users	An actor with admin access taps a card/keychain to the RFID device with the device mode set to passive, intending to register a user's card/keychain. Then, the admin actor can view the UID of the card user, which is still unregistered, marked with a passive status.	Success

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

4. User Information Page		
Add User Data	After the actor with admin access taps a card/keychain to the RFID device with the device mode set to passive, intending to register a user's card/keychain, they then navigate to the User Information page. They click on the Add User Data button, select the desired Card UID, change the status to active, and input user information including name, unique number, date of birth, and address. Finally, they assign a role to the user (student or teacher) and click the save button.	Success
Edit User	An actor with admin access performs changes to a desired card/keychain user.	Success
Delete User	An actor with admin access performs changes to a desired card/keychain user.	Success
5. Account Page		
List of Registered WebApp Accounts	An actor with admin access views the list of registered WebApp accounts marked with an active status.	Success
List of Unregistered WebApp Accounts	A user who wants to create a WebApp account first performs a sign-up process on the WebApp login page. Then, the admin actor can see the registered WebApp accounts marked with a passive status. If the admin approves the creation of the WebApp account, they will assign a role (admin, staff, user) and change the status to active.	Success
Edit Account	An actor with admin access performs changes to a desired WebApp account.	Success
Delete Account	An actor with admin access deletes a desired WebApp account.	Success
6. Login Page & Object Logout		
Login as *User*, Staff, and Admin	Please enter your email and password credentials to log in. Once entered successfully, you will be directed to the main page.	Success
Creating Account *WebApp*	To create a WebApp account, please provide the desired credentials such as name, email, and password. Once done, please wait for the admin actor to approve the account creation.	Success
Logout	Select the Logout button on the navbar. Upon doing so, the actor with any access rights will be automatically logged out from the system and redirected to the initial login page.	Success

4. CONCLUSION

Based on the testing results, the system is functioning well. The RFID reader module in the system successfully reads data from tags, cards, and E-KTP (Electronic Identity Card) at a distance of approximately 2-4 cm between the card and the RFID reader. The obtained data consists of unique codes that are different for each card. After a successful card tap, each unique code is automatically stored in the database, representing the identity of each card,

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

including the number, name, date, and time of entry, as well as the date and time of exit for both teachers and students who attend classes. By implementing this attendance system for teachers and students, the effectiveness of teaching and learning activities at Yayasan Noor Faqih Usman will be maximized. In the beginning of the class, teachers no longer need to manually take attendance by calling out each student's name, thus allowing more time for actual learning in the classroom without interruption.

However, it is important to note that the system is not created to be perfect. There is a need for further in-depth research and development in managing the attendance system by utilizing the latest technologies to ensure the effective and efficient management of student and teacher attendance. For example, utilizing biometric face recognition as an input method for student and teacher attendance.

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

6. REFERENCES

[1] E. Erlangga, Y. Oktavia, A. Cucus, R. Y. Endra dan F. Ariani (2020), Digitalisasi presensi kelas offline berbasis radio frequency identification (rfid), *Jurnal Sistem Informasi dan Telematika EXPLORE, 2*(1), 141-149.

[2] Y. A. Paskawati, Susyana, Antaresti dan E. S. Retnoningyas (2010), Pemanfaatan sabut kelapa sebagai bahan baku pembuatan kertas komposit alternatif, *Widya Teknik, 9*(1), 12-21.

[3] A. Wantoro (2016), Pengembangan sistem presensi dan kedisiplinan dosen, *Jurnal TEKNOINFO*, *10*, 1-5.

[4] R. A. Setyawan (2015), Pengembangan sistem presensi mahasiswa elektronik berbasis RFID, *Jurnal EECCIS, 9*(1), 130-137.

[5] R. Susanto, A. Ananta, A. Santoso dan M. Trianto (2009), Sistem absensi berbasis rfid, *Jurnal Teknik Komputer, 17*(1), 67-74.

[6] K. P. Aji, U. Darusalam dan N. D. Nathasia (2020), Perancangan sistem presensi untuk pegawai dengan RFID berbasis iot, *JOINTECHS (Journal of Information Technology),* 5(1), 25-32.

[7] D. Naufalrochman, A. Muslim dan S. B. Kembaren (2021), Rancang bangun sistem absensi perkuliahan menggunakan kartu rfid dengan website berbasis php, *Jurnal Ilmiah Komputasi, 20*(2), 165-172.

[8] Sudarso, A. (2022). Pemanfaatan basis data, perangkat lunak dan mesin industri dalam meningkatkan produksi perusahaan (literature review executive support system ESS for business). *Jurnal Manajemen Pendidikan dan Ilmu Sosial*, 3(1), 1-14.

[9] Khotimah, T., dan Evanita, E. (2021). Aplikasi Basis Data Perpustakaan Untuk Revitalisasi Perpustakaan Desa Puyoh. *Muria Jurnal Layanan Masyarakat, 3*(1), 46-50.

[10] H. Yuliansyah (2016), Uji kinerja pengiriman data secara wireless menggunakan modul esp8266 berbasis rest architecture, *ELECTRICIAN – Jurnal Rekayasa dan Teknologi Elektro, 10*,(2), 68-77.

[11] R. Dawood, S. F. Qiana dan S. Muchallil (2014), Kelayakan raspberry pi sebagai web server: perbandingan kinerja nginx, apache, dan lighttpd pada platform raspberry pi, *Jurnal Rekayasa Elektrika, 11*(1), 25-29.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149

[12] B. Purnama (2017), Arsitektur teknologi webserver berbasis mini pc dengan raspberry pi, *Jurnal Ilmiah Meida Processor, 9*(3), 281-285.
[13] E. W. Fridayanthie dan T. Mahdiati (2016), Rancang bangun sistem informasi permintaan atk berbasis intranet, *Jurnal Khatulistiwa Informatika, 4*(1), 126-138.
[14] Renaldi, R., Santoso, B. C., dan Natasya, Y. (2020). Tinjauan Pustaka Sistematis terhadap basis data MongoDB. *Jurnal Inovasi Informatika, 5*(2), 132-142.
[15] Fikry, M. (2019). Rancangan basis data kependudukan berdasarkan aspek-aspek kualitas schema database. *TECHSI-Jurnal Teknik Informatika, 8*(2), 28-43.
[16] Waruwu, T. S. (2019). Implementasi postgreSQL sebagai sistem manajemen basis data pada pendaftaran mahasiswa baru berbasis web. *Jurnal Mahajana Informasi, 4*(1), 57-61.
[17] Hamdallah, F., Husain, A., Wijaya, L. A., dan Widianto, S. R. (2020). Sistem Manajemen Basis Data pada Sistem Perpustakaan (Studi Kasus: SMK Al-Wafa). Seminar Nasional Teknologi Komputer dan Sains (SAINTEKS) 1(1), 30-32.
[18] Pratama, G. B., Nurani, T. W., dan Wahju, R. I. (2020). Rancang Bangun Sistem Basis Data untuk Menunjang Pengelolaan Perikanan Tuna Cakalang dan Tongkol di Perairan Natuna. *Akuatika Indonesia, 5*(2), 77-85.
[19] Irawan, R., dan Kaestria, R. (2020). Pemodelan basis data dengan pendekatan model data berorientasi objek pada native apps lokasi tempat ibadah di kota palangka raya. *Jurnal Sains Komputer dan Teknologi Informasi, 2*(2), 36-43.
[20] Fahlevi, F. R., Syafi'i, A. M., Abi Restianto, A., dan Athiyah, U. (2021). *Desain basis data terhadap kebutuhan petani bawang merah. klik-kumpulan jurnal ilmu komputer, 8*(3), 254-264.

DOI: https://doi.org/10.17509/coelite.v2i2.59723
p- ISSN 2829-4157 e- ISSN 2829-4149