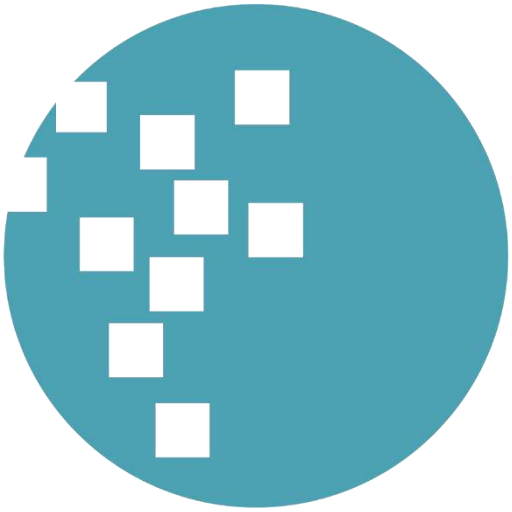
ISSN 2085-4552

ULTIMATICS

Jurnal Teknik Informatika

Engineering | Vol. 16 | No. 1 | Hal. 1 - 75 | June 2024 | ISSN 2085-4552



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2085-4552

EDITORIAL BOARD

EDITORIAL ADDRESS

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

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

Call for Papers

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

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

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

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

ISSN 2085-4552

FOREWORD

ULTIMA Greetings!

Ultimatics : Jurnal Teknik Informatika is the Journal of the Informatics Study Program at Universitas Multimedia Nusantara which presents scientific research articles in the fields of Computer Science and Informatics, as well as the latest theoretical and practical issues, including Analysis and Design of Algorithm, Software Engineering, System and Network Security, Ubiquitous and Mobile Computing, Artificial Intelligence and Machine Learning, Algorithm Theory, World Wide Web, Cryptography, as well as other topics in the field of Informatics. Ultimatics : Jurnal Teknik Informatika is published regularly twice a year (June and December) and is published by the Faculty of Engineering and Informatics at Universitas Multimedia Nusantara.

In this June 2024 edition, Ultimatics enters the 1st Edition of Volume 16. In this edition there are ten scientific papers from researchers, academics and practitioners in the fields of Computer Science and Informatics. Some of the topics raised in this journal are: A Comparative Study of Body Motion Recognition Methods for Elderly Fall Detection: A Review, Application of Convolutional Neural Network Using TensorFlow as a Learning Medium for Spice Classification, Comparing Karate Framework with Others for Automated Regression Testing: A Case Study of PT Fliptech Lentera Inspirasi Pertiwi, Educational Game Design For Carbon Emission Using Game Development Life Cycle Method, Designing a QR Code Attendance System Using BYOD (Bring Your Own Device), U-TAPIS Sal-Tik : Typing Error Detection Using Random Forest Algorithm, Recommendation System Coffee Shop using AHP and TOPSIS Methods, Sentiment Analysis of IMDB Movie Reviews Using Recurrent Neural Network Algorithm, Comparison of Fine-tuned CNN Architectures for COVID-19 Infection Diagnosis, Public Sentiment Analysis on the Transition from Analog to Digital Television Using the Random Forest Classifier Algorithm.

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

Finally, we would like to thank all contributors to this December 2023 Edition of Ultimatics. We hope that scientific articles from research in this journal can be useful and contribute to the development of research and science in Indonesia.

June 2024,

Alexander Waworuntu, S.Kom., M.T.I.
Editor-in-Chief

ISSN 2085-4552

TABLE OF CONTENT

ISSN 2085-4552

A Comparative Study of Body Motion Recognition Methods for Elderly Fall Detection: A Review

Roni Apriantoro[1], Muhammad Adriano Khairur Rizky Setyawan[2], Eri Eli Lavindi[3]
[1,2,3] Dept. of Electrical Engineering, Politeknik Negeri Semarang, Semarang, Indonesia
[1]roni.apriantoro@polines.ac.id, [2]muh.adriano76@gmail.com, [3]erielilavindi@polines.ac.id

Accepted 23 August 2023
Approved 17 January 2024

***Abstract*— To maintain the welfare of the elderly, intensive and effective monitoring is needed to ensure their safety. Conventional elderly activity monitoring has several limitations (i.e., space and time) due to human abilities. This problem can be overcome by applying real-time monitoring methods using Wireless Body Area Networks (WBAN) and Artificial Intelligence (AI). Several methods have been used and tested, including artificial intelligence implementations from sensor data-based to computer vision-based pattern recognition for body motion classification. Several methods that have been studied show accurate results in classifying elderly body motions/gestures. However, the Human Activity Recognition (HAR) method performs better for elderly activity monitoring applications and makes fall classification more accurate.**

***Index Terms*— wireless body area networks; artificial intelligence; fall classification; pattern recognition; computer vision.**

I. INTRODUCTION

A fall is an accidental event that occurs at a fast pace that can result in a person's body lying on the ground or floor, or other lower levels [1], [2]. Fall-related injuries are categorized as fatal and non-fatal [3], [4]. Based on information released by the World Health Organization (WHO), it is estimated that around 684,000 fatal falls occur worldwide every year. The majority of victims of such incidents are individuals over the age of 60. The estimated number of cases makes it the leading cause of death from accidental injuries after traffic accidents. Falls are a significant public health problem for the elderly worldwide, with more than 80% occurring in low- and middle-income countries. Injuries sustained by older adults from falls have far-reaching repercussions for their families, healthcare institutions, and society [2].

One way to reduce the risk of fatal injuries due to falls in the elderly is to conduct intensive supervision [5]. Intensive supervision makes detecting falls and medical enforcement faster [6]. Elderly supervision is still carried out using conventional methods by human labor (volunteers/nurses) [7]. Manual supervision requires more time and limited human resources, while consistent monitoring must continue to detect emergency conditions and provide timely responses [8]. Therefore, new approaches are needed to improve the efficiency of elderly monitoring, provide early warning of changes in behavior or suspicious movements (risk fatal conditions), and reduce reliance on limited human supervision.

In recent years, the development of the Internet of Things (IoT), WBAN, and AI technologies has provided rapid progress in the health sector. Many studies have discussed methods of utilizing these technologies to automate monitoring in the elderly [9]–[11]. An example is the application of sensors installed on the body (around the neck, wrists, or waist) connected to the internet. The device is called a sensor node, a system for identifying and providing user notifications. If a user or individual falls, then the system will quickly activate an alarm to notify a supervisor [12].

In this paper, the authors present a comparative study of methods for detecting falls in the elderly. In addition, challenges and prospects for problems that can be solved through research in the future are also discussed. The contribution to this paper is to provide an accurate description of human activity classification algorithms to be further developed and applied to the real environment for fall recognition.

II. METHODOLOGY

Data collection in this paper comes from journals and proceeding articles between 2018 and 2022. The authors searched for references from the IEEE Explore, ScienceDirect, MDPI, and Google Scholar websites using the keywords "IoT", "elderly", "older adults", "motion recognition", "fall detection", and "machine learning". All journals used by the authors are based on relevance to the development and application of motion recognition technology for monitoring elderly activities, especially in fall conditions. Data collection methods can be seen in Fig. 1.

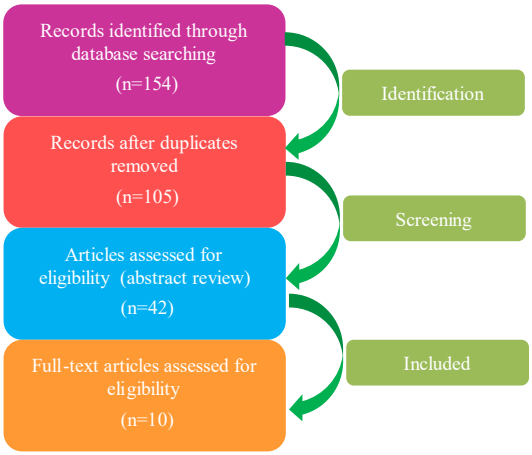


Fig 1. Literature Search Process Diagram

III. Result and General Overview

A. Literature Search Result

This study includes ten journal and conference articles published between 2018 and 2023. These studies come from nine countries, including Thailand [13], Germany [14], Canada [15], Malaysia [16], Spain [17], Portugal [18], Japan [19], India [20], and Indonesia [21]. Most of the research is the innovation and development of motion recognition methods using pattern recognition.

B. Body Motion Recognition Systems and Techniques

Motion recognition in humans can be done by installing sensor nodes on the human body combined with artificial intelligence algorithms or using computer vision method approaches. The method of installing sensor nodes connected to the network (WBAN) is described in subchapter 1. The artificial intelligence algorithm used to perform the fall classification is discussed in Subchapter 2.

1) Wireless Body Area Network

Wireless Body Area Network (WBAN) is a specific type of sensor network that uses wireless sensor nodes on a person's body to measure physiological parameters such as body temperature [22], blood pressure [23], blood glucose [24], heartbeat [25], and other parameters, which allows the patient's health to be monitored remotely [26]. WBAN can be wearable or implanted in a person's body [27]. WBAN aims to ensure individual health by intensively monitoring physiological information using sensors and sending the data to the server. It can help doctors continuously understand patients' health. In addition, WBAN can also significantly reduce patient care costs by monitoring data related to patients' vital signs for an extended period [28], [29].

Almost the same as wireless sensor networks in general, WBAN uses a star topology architecture but with a slightly different approach. In this topology, sinkholes are located in the body to collect information from sensor nodes [10], [30]. Details on the WBAN network architecture can be seen in Fig. 2. In this network, sensor nodes have access to limited energy resources. In physical sensor networks, sensors (such as motion sensors) are installed on the patient's body to observe the patient's vital signs or detect motions/gestures in real-time [31]. By doing this, WBAN can provide an instant response to users (medical personnel, volunteers, families), and the user can follow the patient's disease progression and take the necessary precautions relatively quickly [32]. The use of sensor energy in this network is crucial because if the energy source runs out, the duration of network life will be shorter [33].

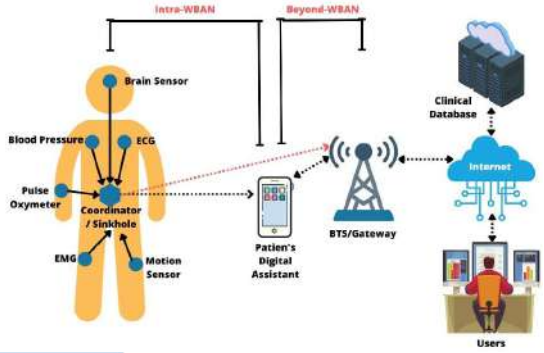


Fig 2. Architecture of WBAN

2) Artificial Intelligence Algorithm for Fall Classification

In detecting the motions or activities of the elderly, the authors found several artificial intelligence methods that can be used to classify human motions/gestures. Further explanation can be seen in Table 1.

C. Implementation of Body Motion Recognition Techniques in Elderly Monitoring

Pattern recognition is an approach to identifying specific patterns or characteristics in data to classify or identify objects or phenomena being observed. In body motion recognition, pattern recognition methods are used to classify body motions/gestures based on patterns identified in sensor or camera data [34]. Pattern recognition methods in body motion recognition involve training and testing using machine learning algorithms [35]. The training process involves stages to identify, analyze, and classify motions performed by objects or individuals [36]. Body motion recognition involves several steps, from collecting data to making classification decisions [37]. The training process of a machine learning algorithm model can be seen in Fig. 3. After the training process is complete, the moved pattern recognition model will be used to classify body motions in the

TABLE I. SUMMARY OF MOTION/GESTURE-RECOGNITION TECHNIQUES.

Method	Description	Pros	Cons
Artificial Neural Network (ANN) [3], [38]	• Computational models inspired by the workings of the biological nervous system in the human brain • ANN consists of a large number of simple processing units called neurons or nodes • There are several commonly used types of ANN architectures, including Feedforward Neural Networks (FNN), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Temporal Convolutional Networks (TCN), and Long Short-Term Memory (LSTM).	• Can handle problems of high complexity, including pattern recognition, classification, and prediction. • It can learn from training data and adapt itself. • Can solve the problem of imperfect data (noise) • Can process data simultaneously.	• Requires a considerable amount of training data to produce accurate results. • Large amounts of data affect long computation times. • Lack of clear interpretation in making decisions. • Prone to overfitting.
Decision Trees [38]	• A pattern recognition method that uses a set of tree-based decision rules to classify data. • Decision trees generate a predictive model in the form of a tree structure, where each node in the tree represents a decision rule, and each branch represents the possible outcome of that decision.	• Humans can easily interpret and explain it due to simple tree-based decision rules. • Can overcome incomplete data or have missing values of various types and scales.	• Prone to overfitting if the tree is too complex and fits too much into the training data. • Sensitive to small changes in training data • Must discretize continuous attributes.
Support Vector Machines (SVM) [3], [38]	• Machine learning methods are used for classification and regression by separating data into two classes using a hyperplane (Separation space between 2 dimensions).	• Effective handling of high-dimensional data, even when the sample size is small. • Can separate data that cannot be separated linearly.	• Less efficient for large datasets. • Requires proper selection of parameters, such as kernel parameters, which can affect the performance and stability of the model.
Random Forest [3], [38]	• It uses a combination of several decision trees to perform classification or regression. • Each decision tree in a random forest is generated randomly and independently. Finally, the classification results are taken through voting or averaging all existing decision trees.	• Have good performance in terms of classification or prediction accuracy. • It can handle large datasets with different features, including numerical and categorical ones. • Can solve overfitting problems by using the ensemble concept	• When doing training, it takes relatively longer than simple models such as a single decision tree.
k-Nearest Neighbor (k-NN) [3], [13], [38]	• Non-parametric methods used for classification and regression based on proximity to training data known to the class • k-NN is an unsupervised learning method that falls into the category of lazy learning algorithms, which means that it does not perform complex training processes at an early stage. • Instead, k-NN stores all training data and performs the classification process directly when new data is classified.	• Easy to understand and implement. • It does not require complex training processes. • Effective in cases with high-dimensional data. • Suitable for data that does not have complex patterns	• Sensitive to irrelevant data or noise • Sensitive to different data scales.

testing process. Testing data not used in the training process will be provided to the model, and the model will identify patterns that match the observed body motions [37].

Several studies apply the gyroscopes and accelerometers (IMU) sensor that aim to analyze changes in acceleration vectors during falls and create fall recognition algorithms based on specific stages (such as the beginning of the fall, shock, aftershock, and body position), such as research in [15]. Research by [17] also discussed the development of pattern recognition methods by integrating an accelerometer into a smartwatch to recognize falls. When the detector device detects a fall event, the data will be sent via Bluetooth to the smartphone as a gateway. Then smartphone forwards (transmits) the data to the cloud server, then the cloud server sends a notification [17].

Another study by [16] has also conducted research on pattern recognition methods by applying IMU sensors, which are a combination of Gyroscope and Accelerometer sensors (MPU6050) installed on elderly bodies, and Bluetooth Low Energy (BLE) Beacon location sensors scanned using the Raspberry Pi module. The data generated by the IMU sensor is then

acquired and analyzed using machine learning. Using Random Forest and SVM classification techniques, the accuracy results obtained were 97.69% in detecting motion and 97.25% in detecting the location of users or individuals. On the other hand, research by [39] had an average of 73.7% accuracy and 81.1% precision for older women's fall detection using the RF classification method. Furthermore, research by [40] used SVM to classify the fall events from IMU data and got a 99% accuracy rate.

Meanwhile, research by [15] proposes a configurable real-time motion pattern recognition framework using the same sensor (MPU6050). The classification method applied is the single-hidden-layer Feedforward Neural Network (FNN). This technique's training results showed an accuracy rate of 98.23% from as many trials (~207 thousand data points) or datasets. Similar trials have also been conducted using the k-NN classification technique. By selecting k = 3 with the same histogram feature and training test data points, the accuracy obtained was 97.66%. FNN slightly outperformed k-NN. These results can be improved by involving more subjects and (or) more IMU nodes. While FNN and k-NN provide similar results for large data sets, k-NN requires significant memory usage and experiences long latency, which makes it unfeasible for real-time execution.

Another study by [14] proposes classifying patterns using Temporal Convolutional Network (TCN). The matrix used for performance analysis is accuracy and F1 score using the proposed approach, then compared with baseline LSTM, bidirectional LSTM, residual LSTM, and deep residual bidirectional LSTM. The results showed that the proposed approach with TCN performed better than others, with 94.2% accuracy on the F1 score matrix. A high F1 score indicates the efficiency of the model.

Not only using IMU sensors, the study in [41] also uses RFID to detect fall events. The various classifier algorithms include XGBoost, Gated Recurrent Units (GRUs), RF, KNN, and Logistic Regression (LGR). The results show that the GRUs exhibit a 44% accuracy rate, the RF algorithm achieves a 43% accuracy rate, and XGBoost achieves a 33% accuracy rate. Meanwhile, KNN outperforms the others with a 99% accuracy rate.

In the context of accuracy, research using IMU sensors installed on individual bodies has proven effective and high-quality in monitoring the elderly. However, the research requires a relatively large number of sensor nodes, so other methods are needed to support or complement the performance of devices used to monitor the elderly more efficiently in terms of equipment quantity. In this case, the authors found a solution in a recent study, where visual sensors in the form of cameras were used to identify body motion [19]–[21]. This method can monitor the activities of the elderly using only cameras so that the quantity of equipment and installation will be more efficient.

The development of alternative WBAN methods to monitor the elderly has been carried out using visual sensors to identify motions combined with computer vision-based artificial intelligence technology. This method aims to produce accurate and detailed visual representations of body motions in three-dimensional space. In this method, Artificial Neural Network (ANN) is the most commonly used classification technique [36]. The reconstruction process begins with data capture from 3D camera sensors that can produce information about the position and depth of objects in space. Once the 3D data is collected, the next step is to build a 3D body motion model. This model can be a digital representation consisting of dots or a mesh (network) that describes the shape and position of the body in space. Such 3D models can be used to perform further analysis, such as feature extraction, motion tracking, or animation. This method is commonly called Human Activity Recognition (HAR) [19]–[21].

Research by [20] proposed motion recognition and fall detection systems using deep convolutional long short-term memory (ConvLSTM) networks, which involve the merging of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. This study used geometric and kinematic features on the Depth Camera Kinect sensor to build features. Meanwhile, cross-entropy and softmax activation are used to obtain performance models and measures. This proposed model was evaluated on a video dataset (KinectHAR dataset), and results were obtained with an accuracy rate of 98.89%.

On the other hand, research by [21] also proposed a new motion recognition method involving motion feature extraction techniques to form windows of frames representing motion in a time series, which were then provided as input to the CNN model. This model was trained and tested using the Florence public dataset with 3D data from the positions of fifteen joints comprising 215 videos. Using the Kinect 3D camera sensor, the accuracy results obtained by the model were 99.3421%, the precision value was 100%, the recall value was 94.73%, and the F1 score was 97.29%.

In addition to using ANN to classify data from 3D camera sensors, some studies use SVM as a classification method. Research conducted by [19] proposed a method, namely motion recognition using the SVM technique in depth differences between two consecutive frames to determine whether motion has been detected in a video frame. Using the Depth Camera sensor, this feature extraction system shows that sitting/standing/lying down has a classification accuracy of 96.60%, 96.41%, and 95.35%, respectively. In addition, the overall accuracy of the system is 91.82%.

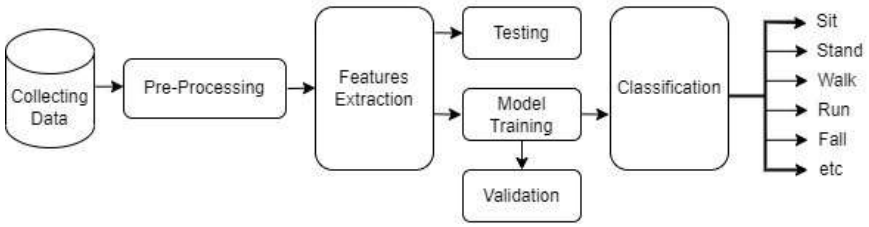


Fig 3. The learning process of the machine learning algorithm

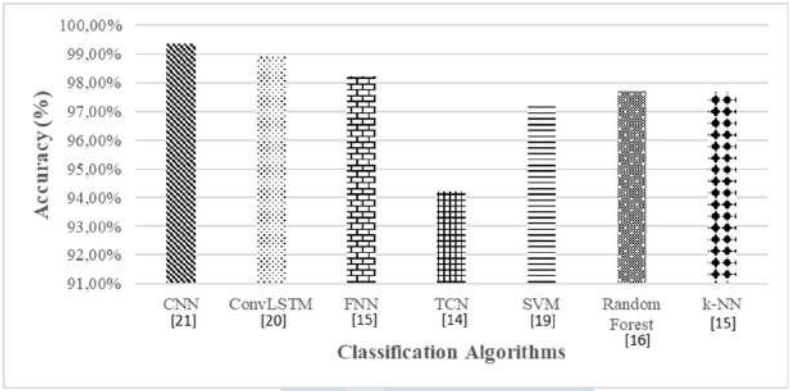


Fig 4. Accuracy comparison of body motion/gesture recognition techniques

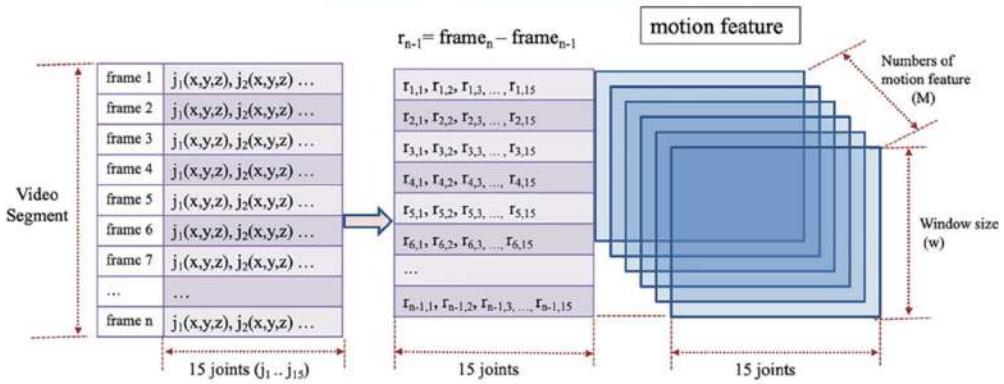


Fig 5. Feature Extraction Method [21]

IV. DISCUSSION

A. Related Literature Review

Overall, various ANN methods have very accurate performance in studying complex patterns using large datasets. However, other classification techniques such as k-NN, SVM, and Random Forest also have a high degree of accuracy in studying these patterns, but in smaller, less complex datasets [42]. The authors compared the classification techniques studied in Fig. 4 to see the results more clearly.

CNN and ConvLSTM methods combined with 3D camera sensors have proven to be very effective in detecting and analyzing the body motion of subjects or objects. With high accuracy, it can be used to monitor activities carried out by the elderly at a distance in a quality and efficient manner without the need for many sensors installed on the body of the elderly or subject. The CNN method with the extraction of certain features is the best method to date that can be used to accurately detect various kinds of human movements or activities [21].

This study [21] uses the CNN method for extracting different features. The feature extraction technique used is the sliding window technique on time series data to construct motion features. Each video segment in the dataset consists of multiple frames. Next, the distance change will be calculated consecutively at each joint coordinate point or joint (fifteen joints) on the frame. Identification of features of motion is carried out through this process. The depiction of feature extraction techniques used in this study can be seen in Fig. 5.

In addition, the method used was also tried on other datasets, where the activities carried out by individuals include walking, sitting, standing, falling, and other motions. The experimental results showed a loss value of 0.1964 and an accuracy of 93.18%. So, it can be categorized that the model used can also classify different datasets with high accuracy [21].

B. Challenges

According to the existing research results that have been reviewed, Human Activity Recognition (HAR)-based Pattern Recognition has excellent results in identifying movements so that it can classify body motion with a high level of accuracy. 3D camera sensor or Depth Camera can be the latest application of the HAR method, where the results obtained are very accurate. However, this method can be improved again in a more varied dataset to provide validation, considering that this new method is more efficient regarding equipment quantity. That is a challenge that must be developed in the future so that the activities of the elderly can be known and monitored by caregivers and families of the elderly who are monitored. By combining computer vision with medical technology, this system can significantly improve the quality of health for the elderly. At the same time, this system can also reduce accident rates and maintenance costs. Nevertheless, on the other hand, the application of cameras also has limitations regarding installation in private areas, such as bathrooms and toilets.

C. Development Approach

In this comparative literature review, error detection/classification exists when using several models in other datasets. It can be reduced by converting depth maps into 3D Point Clouds by grouping and segmenting distance information of observed objects to improve analyzing patterns of daily activities and behavior using statistical models such as the Hidden Markov Model (HMM). Therefore, how to improve accuracy during observation and interaction between objects will be the subject of further research in the future. In addition, combining with WBAN will support the detection and classification accuracy of fall events by measuring the impulse response to the signal generated by the IMU sensor. Combining WBAN with computer vision systems will extend the monitoring range to private areas.

The HAR-based pattern recognition using a classification algorithm, especially for elderly fall detection in this review, performs well and can accurately recognize the fall, but it still has several limitations. In this literature review, the authors have not discussed other classification algorithms on multilabel classification methods and hierarchical decision making-based classification algorithms. Furthermore, only wearable device-based and vision-based methods are discussed in this study.

V. CONCLUSION

Human Activity Recognition (HAR) is one example of AI based on computer vision in the scope of wireless sensor networks. It is a new method that must be developed in the future in order to be useful in many fields, especially in terms of supervision or monitoring. Computer vision is a relatively inexpensive technological breakthrough with high classification accuracy. There have been many implementations of motion recognition methods using AI, and one example can be applied to monitor activities carried out by the elderly directly from a distance. Starting from daily activities, such as walking, sitting, standing, sleeping, and other involuntary activities (for example, falling), and can analyze the behavior of the elderly, such as sluggishness, confusion, and other signs of cognitive impairment. The latest development of existing methods is needed to optimize the quality of results and the quantity of tools used. So, this opens up opportunities for future research to find new ways or methods to carry out remote monitoring more efficiently and effectively.

REFERENCES

[1] S. Giovannini *et al.*, "Falls among Older Adults: Screening, Identification, Rehabilitation, and Management," *Appl. Sci.*, vol. 12, no. 15, 2022, doi: 10.3390/app12157934.

[2] WHO, "Falls," 2021. https://www.who.int/news-room/fact-sheets/detail/falls#:~:text=Falls are the second leading,greatest number of fatal falls (accessed Aug. 09, 2023).

[3] S. Usmani, A. Saboor, M. Haris, M. A. Khan, and H. Park, "Latest research trends in fall detection and prevention using machine learning: A systematic review," *Sensors*, vol. 21, no. 15, pp. 1–23, 2021, doi: 10.3390/s21155134.

[4] R. Vaishya and A. Vaish, "Falls in Older Adults are Serious," *Indian J. Orthop.*, vol. 54, no. 1, pp. 69–74, 2020, doi: 10.1007/s43465-019-00037-x.

[5] J. B. Fernandes, S. B. Fernandes, A. S. Almeida, D. A. Vareta, and C. A. Miller, "Older adults' perceived barriers to participation in a falls prevention strategy," *J. Pers. Med.*, vol. 11, no. 6, 2021, doi: 10.3390/jpm11060450.

[6] S. Hayashi, Y. Misu, T. Sakamoto, and T. Yamamoto, "Cross-Sectional Analysis of Fall-Related Factors with a Focus on Fall Prevention Self-Efficacy and Self-Cognition of Physical Performance among Community-Dwelling Older Adults," *Geriatr.*, vol. 8, no. 1, 2023, doi: 10.3390/geriatrics8010013.

[7] J. J. Severance, S. Rivera, J. Cho, J. Hartos, A. Khan, and J. Knebl, "A Collaborative Implementation Strategy to Increase Falls Prevention Training Using the Age-Friendly Health Systems Approach," *Int. J. Environ. Res. Public Health*, vol. 19, no. 10, 2022, doi: 10.3390/ijerph19105903.

[8] R. Woltsche, L. Mullan, K. Wynter, and B. Rasmussen, "Preventing Patient Falls Overnight Using Video Monitoring: A Clinical Evaluation," *Int. J. Environ. Res. Public Health*, vol. 19, no. 21, pp. 1–12, 2022, doi: 10.3390/ijerph192113735.

[9] S. Upadhyay *et al.*, "Challenges and Limitation Analysis of an IoT-Dependent System for Deployment in Smart Healthcare Using Communication Standards Features," *Sensors*, vol. 23, no. 11, 2023, doi: 10.3390/s23115155.

[10] M. Yaghoubi, K. Ahmed, and Y. Miao, "Wireless Body Area Network (WBAN): A Survey on Architecture, Technologies, Energy Consumption, and Security

Challenges," *J. Sens. Actuator Networks*, vol. 11, no. 4, 2022, doi: 10.3390/jsan11040067.

[11] I. Ha, "Technologies and research trends in wireless body area networks for healthcare: A systematic literature review," *Int. J. Distrib. Sens. Networks*, vol. 2015, 2015, doi: 10.1155/2015/573538.

[12] L. Zhong *et al.*, "Technological Requirements and Challenges in Wireless Body Area Networks for Health Monitoring: A Comprehensive Survey," *Sensors*, vol. 22, no. 9, 2022, doi: 10.3390/s22093539.

[13] J. Muangprathub, A. Sriwichian, A. Wanichsombat, S. Kajornkasirat, P. Nillaor, and V. Boonjing, "A novel elderly tracking system using machine learning to classify signals from mobile and wearable sensors," *Int. J. Environ. Res. Public Health*, vol. 18, no. 23, 2021, doi: 10.3390/ijerph182312652.

[14] Y. A. Andrade-Ambriz, S. Ledesma, M.-A. Ibarra-Manzano, M. I. Oros-Flores, and D.-L. Almanza-Ojeda, "Human activity recognition using temporal convolutional neural network architecture," *Expert Syst. Appl.*, vol. 191, p. 116287, 2022, doi: https://doi.org/10.1016/j.eswa.2021.116287.

[15] O. Sarbishei, "A Platform and Methodology Enabling Real-Time Motion Pattern Recognition on Low-Power Smart Devices," *IEEE 5th World Forum Internet Things, WF-IoT 2019 - Conf. Proc.*, pp. 269–272, 2019, doi: 10.1109/WF-IoT.2019.8767219.

[16] N. E. Tabbakha, W. H. Tan, and C. P. Ooi, "Elderly action recognition system with location and motion data," *2019 7th Int. Conf. Inf. Commun. Technol. ICoICT 2019*, pp. 1–5, 2019, doi: 10.1109/ICoICT.2019.8835224.

[17] E. Casilari-Pérez and F. García-Lagos, "A comprehensive study on the use of artificial neural networks in wearable fall detection systems," *Expert Syst. Appl.*, vol. 138, p. 112811, 2019, doi: https://doi.org/10.1016/j.eswa.2019.07.028.

[18] O. Ribeiro, L. Gomes, and Z. Vale, "IoT-Based Human Fall Detection System," *Electron.*, vol. 11, no. 4, 2022, doi: 10.3390/electronics11040592.

[19] T. T. Zin *et al.*, "Real-time action recognition system for elderly people using stereo depth camera," *Sensors*, vol. 21, no. 17, 2021, doi: 10.3390/s21175895.

[20] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "Skeleton-based human activity recognition using ConvLSTM and guided feature learning," *Soft Comput.*, vol. 26, no. 2, pp. 877–890, 2022, doi: 10.1007/s00500-021-06238-7.

[21] E. S. Rahayu, E. M. Yuniarno, I. K. E. Purnama, and M. H. Purnomo, "Human activity classification using deep learning based on 3D motion feature," *Mach. Learn. with Appl.*, vol. 12, no. January, p. 100461, 2023, doi: 10.1016/j.mlwa.2023.100461.

[22] M. Udin Harun Al Rasyid, S. Sukaridhoto, A. Sudarsono, A. N. Kaffah, and M. Udin Harun Al Rasyid, "Design and Implementation of Hypothermia Symptoms Early Detection with Smart Jacket Based on Wireless Body Area Network," *IEEE Access*, vol. 8, pp. 155260–155274, 2020, doi: 10.1109/ACCESS.2020.3018793.

[23] S. Gupta *et al.*, "Modeling of On-Chip Biosensor for the in Vivo Diagnosis of Hypertension in Wireless Body Area Networks," *IEEE Access*, vol. 9, pp. 95072–95082, 2021, doi: 10.1109/ACCESS.2021.3094227.

[24] I. Rodríguez-Rodríguez, J.-V. Rodríguez, and M. Campo-Valera, "Applications of the Internet of Medical Things to Type 1 Diabetes Mellitus," *Electronics*, vol. 12, no. 3, 2023, doi: 10.3390/electronics12030756.

[25] Sonal, S. R. N. Reddy, and D. Kumar, "Early congenital heart defect diagnosis in neonates using novel WBAN based three-tier network architecture," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3661–3672, 2022, doi: 10.1016/j.jksuci.2020.07.001.

[26] Y. Qu, G. Zheng, H. Ma, X. Wang, B. Ji, and H. Wu, "A survey of routing protocols in WBAN for healthcare applications," *Sensors (Switzerland)*, vol. 19, no. 7, 2019, doi: 10.3390/s19071638.

[27] H. Taleb, A. Nasser, G. Andrieux, N. Charara, and E. Motta Cruz, "Wireless technologies, medical applications and future challenges in WBAN: a survey," *Wirel. Networks*, vol. 27, no. 8, pp. 5271–5295, 2021, doi: 10.1007/s11276-021-02780-2.

[28] C. A. Tavera, J. H. Ortiz, O. I. Khalaf, D. F. Saavedra, and T. H. H. Aldhyani, "Wearable wireless body area networks for medical applications," *Comput. Math. Methods Med.*, vol. 2021, 2021, doi: 10.1155/2021/5574376.

[29] Y. Albagory, "An efficient WBAN aggregator switched-beam technique for isolated and quarantined patients," *AEU - Int. J. Electron. Commun.*, vol. 123, p. 153322, 2020, doi: https://doi.org/10.1016/j.aeue.2020.153322.

[30] M. S. Hajar, M. O. Al-Kadri, and H. K. Kalutarage, "A survey on wireless body area networks: architecture, security challenges and research opportunities," *Comput. Secur.*, vol. 104, 2021, doi: 10.1016/j.cose.2021.102211.

[31] T. Ivașcu and V. Negru, "Activity-Aware Vital Sign Monitoring Based on a Multi-Agent Architecture," *Sensors*, vol. 21, no. 12, 2021, doi: 10.3390/s21124181.

[32] S. Javaid, S. Zeadally, H. Fahim, and B. He, "Medical Sensors and Their Integration in Wireless Body Area Networks for Pervasive Healthcare Delivery: A Review," *IEEE Sens. J.*, vol. 22, no. 5, pp. 3860–3877, 2022, doi: 10.1109/JSEN.2022.3141064.

[33] M. M. Kamruzzaman and O. Alruwaili, "Energy efficient sustainable Wireless Body Area Network design using network optimization with Smart Grid and Renewable Energy Systems," *Energy Reports*, vol. 8, pp. 3780–3788, 2022, doi: 10.1016/j.egyr.2022.03.006.

[34] A. Hayat, M. D. Fernando, B. P. Bhuyan, and R. Tomar, "Human Activity Recognition for Elderly People Using Machine and Deep Learning Approaches," *Inf.*, vol. 13, no. 6, pp. 1–13, 2022, doi: 10.3390/info13060275.

[35] M. G. Morshed, T. Sultana, A. Alam, and Y.-K. Lee, "Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities," *Sensors*, vol. 23, no. 4, 2023, doi: 10.3390/s23042182.

[36] P. P. Ariza-Colpas *et al.*, "Human Activity Recognition Data Analysis: History, Evolutions, and New Trends," *Sensors*, vol. 22, no. 9, 2022, doi: 10.3390/s22093401.

[37] M. H. Arshad, M. Bilal, and A. Gani, "Human Activity Recognition: Review, Taxonomy and Open Challenges," *Sensors*, vol. 22, no. 17, 2022, doi: 10.3390/s22176463.

[38] M. E. Karar, H. I. Shehata, and O. Reyad, "A Survey of IoT-Based Fall Detection for Aiding Elderly Care: Sensors, Methods, Challenges and Future Trends," *Appl. Sci.*, vol. 12, no. 7, 2022, doi: 10.3390/app12073276.

[39] A. Hua *et al.*, "Accelerometer-based predictive models of fall risk in older women: a pilot study," *npj Digit. Med.*, vol. 1, no. 1, p. 25, 2018, doi: 10.1038/s41746-018-0033-5.

[40] M. Saleh and R. L. B. Jeannes, "Elderly Fall Detection Using Wearable Sensors: A Low Cost Highly Accurate Algorithm," *IEEE Sens. J.*, vol. 19, no. 8, pp. 3156–3164, 2019, doi: 10.1109/JSEN.2019.2891128.

[41] H. A. Alharbi, K. K. Alharbi, and C. A. U. Hassan, "Enhancing Elderly Fall Detection through IoT-Enabled Smart Flooring and AI for Independent Living Sustainability," *Sustainability*, vol. 15, no. 22, 2023, doi: 10.3390/su152215695.

[42] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, and J. Liu, "Human Action Recognition From Various Data Modalities: A Review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3200–3225, 2023, doi: 10.1109/TPAMI.2022.3183112.

ISSN 2085-4552

Application of Convolutional Neural Network (CNN) Using TensorFlow as a Learning Medium for Spice Classification

Muhammad Naufal Adi Saputro[1], Febri Liantoni [2], Dwi Maryono [3]
[1,2,3] Information Engineering and Computer Education, Sebelas Maret University
E-mail : novalxena27@gmail.com[1] , febri.liantoni@gmail.com[2], dwimarus@yahoo.com[3]

Accepted 3 September 2023
Approved 6 June 2024

***Abstract*— The purpose of this research are: (1) To determine the accuracy of the CNN method in the development of a website for classifying spices, (2) To assess the feasibility of the spice classification websiteas a learning medium, (3) To ascertain user responses to the spice classification website as a learning medium. The method employed in this research is research and development. This study utilizes the ADDIE development method, which comprises 5 stages: (1) Analysis, (2) Design, (3) Development, (4) Implementation, and (5) Evaluation. The research yielded a significantly high accuracy rate. This is demonstrated by the results showing an accuracy of 96%, precision of 97%, and recall of 96%. Moreover, the research found the developed website to be feasible. This is supported by the evaluation using the Learning Object Review Instrument (LORI), resultingin a score of 88% from media experts and a score of 90% from subject matter experts. Additionally, user response was positive. This is evidenced by testing thelearning media on 10th-grade culinary students from SMK N 4 Surakarta, which yielded a score of 76% using the System Usability Scale (SUS), indicating a favorable usability assessment. In conclusion, the spiceclassification website, as a learning medium, can be employed as a suitable educational tool.**

***Index Terms*— Convolutional Neural Network (CNN); Development; Learning; Spices; TensorFlow; Website**.

I. INTRODUCTION

As we know, Indonesia is a country rich in spices. These spices can be used as a source of healthy food ingredients. However, over time, spices have been overshadowed by fast food or so-called junk food. Consequently, the current generation is less familiar with spices. Therefore, in this article, we will explore the effects or benefits of various types of spices on physiological functions withinthe body.

Indonesia is a country abundant in spices. There are various types of spices in Indonesia, and their usage has long been practiced within the community. Spices are widely used in the pharmaceutical and food industries, among others [1]. Contemporary lifestyles have led the Generation Z to be less acquainted with Indonesia's natural wealth, namely spices [2].

The concern of Generation Z towards spices is diminishing due to their inclination towards instant and fast items such as fast food or junk food. Consequently, the utilization of spices as cooking ingredients or for medicinal purposes has become less common. As a result, Generation Z is no longer familiar with the benefits and properties inherent in spices, and even worse, manyyoungsters are unfamiliar with the various types of spices around them [3]. Recognizing the different types of spices poses a challenge for the millennial generation. Some spices mayappear similar at first glance without knowing their characteristics. Based on a survey conducted in this study, involving 100 respondents attempting to identify five types of Indonesian spices, only 31% of respondents accurately identified more than three types of spices [4].

Generation Z is more interested in smartphones than in learning traditional knowledge. However, smartphones can alsoserve as highly effective learning tools. Educators and learners, when exposed to digital technology systems, are motivatedwhen they perceive benefits from such technological systems [5]. Utilizing smartphones has several advantages. One ofthem is their internet connectivity. Creating teaching materials through digital technology can be more engaging and motivating, as content can be presented not only through text but also through images, audio, video, and animations, influencing improved learning behavior [5]. This facilitates access to and implementation of various knowledge fields widely available. Moreover, smartphones enable learning anytime and anywhere, irrespective of time and place. These affordable devices are accessible to the general public.

The literature review encompasses several studies related to the application particularly TensorFlow, in various domains. "Implementation of TensorFlow-based deep learning in the learning application of around things in English" by Heru Budianto et al., they introduces a learning application aimed at

teaching English to children by recognizing objects in their environment using TensorFlow. However, this study primarily focuses on language learning rather than object classification [17]. In contrast, our research targets spice classification, providing a unique application of TensorFlow in a specific domain.

In another study titled “Machine learning in medicine using JavaScript: building web apps using TensorFlow.js for interpreting biomedical datasets” Jorge G. Pires discusses the utilization of TensorFlow.js for interpreting biomedical datasets, achieving high accuracy in tasks like diabetes detection and surgery complications prediction [18]. While this study demonstrates the effectiveness of TensorFlow.js in medical applications, it does not directly relate to our research on spice classification. Our study addresses a different domain, focusing on image classification for spice recognition rather than medical data analysis.

Then, Noor Mohd Ariff Brahin et al. present "LearnWithIman" in “Development of vocabulary learning application by using machine learning technique,” a vocabulary learning application for children using TensorFlow object detection API. Similar to Budianto et al., this study emphasizes language learning through object recognition but targets a different audience and language [19]. Our research, however, concentrates on spice classification, offering a distinct application of TensorFlow in educational technology.

Other than that, Sona Saitou et al. (2018) apply TensorFlow to recognize characteristic structures in fragment molecular orbital (FMO) calculations in “Application of TensorFlow to recognition of visualized results of fragment molecular orbital (FMO) calculations.” Although this study shares the use of TensorFlow for pattern recognition, it deals with molecular structures rather than object classification [20]. Our research diverges from this by focusing on image classification for spice identification, showcasing the versatility of TensorFlow in various fields.

Lastly, Haim A Abenhaim (2023) develops an object detection application for a forward collision early warning system using TensorFlow Lite in “Object Detection Application for a Forward Collision Early Warning System Using TensorFlow Lite on Android.” While this study employs TensorFlow for object detection, it is oriented towards automotive safety rather than spice classification [21]. Our research explores a different application domain, demonstrating the adaptability of TensorFlow in diverse contexts.

In summary, while existing literature demonstrates the versatility of TensorFlow in various domains such as language learning, medical diagnostics, and object detection, there remains a gap in the application of this technology specifically for spice classification. Our research addresses this gap by employing TensorFlow for image classification in the context of spice recognition, offering a novel contribution to the field of applications.

The introduction of spice types can utilize deep learning techniques. Deep learning involves processing data using artificial neural networks. This algorithm takes data as input and processes it through hidden layers. Subsequently, the algorithm performs nonlinear transformations on input data to generate output values. A widely used deep learning technique is Convolutional Neural Network (CNN), capable of recognizing spice types. Hence, in this study, the researcher will employ a website-based Convolutional Neural Network (CNN) method. Several studies on image processing using CNN have yielded high accuracy rates. For instance, a study conducted by Wulandari, Yasin, Widiharih, and Statistika on Classification Of Digital Images Of Spices Using Convolutional Neural Network (Cnn). achieved an accuracy rate of 88.89% [6].

Based on the aforementioned background, many people are still unfamiliar with the various spices in Indonesia, necessitating technology to facilitate their recognition. The utilization ofthe CNN algorithm can be used for classification. Therefore, this research aims to perform website- based spice classification using the CNN algorithm.

II. RESEARCH METHOD

The research method I am using in this study is the Research and Development (R&D) method. R&D, or Research and Development, is a research approach used to develop and test products that will be appliedin the field of education [7]. Utilizing the ADDIE approach, which stands for Analysis,Design, Develop, Implement, and Evaluate. The ADDIE model is employed to establish a foundation for performance in the learning process, by implementing the concept of developing learning product designs [8]. Thisstudy employs a questionnaire adapted from the Learning Object Review Instrument (LORI). The material being assessed covers several aspects, namely content quality, learning goal alignment, feedback and adaptation, and motivation.

The application, tools and library I will use is as follows:

Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a type of neural network that is specialized for processing data with a grid structure, such as two-dimensional images. The term "convolution" refers to a linear algebraic operation involving matrix multiplication between the filter and the image being processed. This process takes place in convolutional layers, which are one of several types of layers that can exist in a network. Convolution layers are the main and critical component of CNN. In addition to the convolution layer, another type of layer that is often used is the Pooling Layer, which is used to take the maximum or average value of the pixels in the image parts. The following is an example of a CNN architecture.

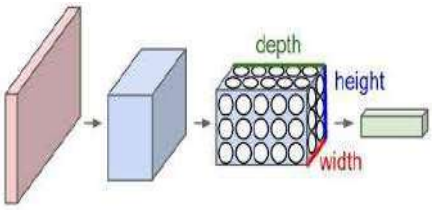


Fig 1. Neural Network Structure

The figure above shows that each input layer that is entered has a different volume and is represented by depth, height and width. Each quantity obtained depends on the previous layer's filtration results and the number of filters used. The network models have proven effective in dealing with image classification problems [11].

The function of CNN is to process data in the form of multiple arrays. There are three layers or layers in CNN, namely the Convolutional Layer, the Pooling Layer, and the Fully Connected Layer. An illustration of the CNN architecture is shown in Figure 2.

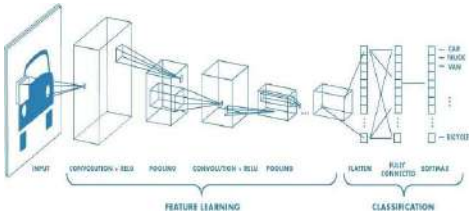


Fig 2. CNN Architecture

Input Layers

In this layer, CNN will store the pixel value of the input image. In general, each image has a different size. For example, an image with a size of 224x224 and 3 color channels Red, Green, Blue (RGB) will be used as input for CNN in the form of an array with a size of 224x224x3.

Convolution Layers

The Convolution Layer is the first component in CNN. This layer performs convolution between the input image with predetermined filters without changing the structure of the original image. The function of the convolution layer is to extract features from the image to be used in model training [12]. An example image of the convolutional layer is shown in Figure 3.

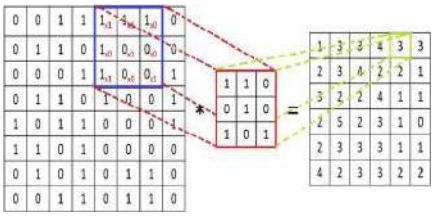


Fig 3. Convolution Layer

The Convolution Layer performs a process that produces a new image that contains extracted features from the input image. This process uses a filter (kernel) in the form of a 2-dimensional array with a size of 5x5, 3x3, or 1x1. Each image passes through the filter, resulting in a feature map. The result of the feature map from this layer is then used in the next layer, namely the Activation Function.

Pooling Layer

Pooling layer is the process of reducing the number of parameters and number of calculations in the network, as well as preventing overfitting of the image. This layer is divided into average pooling and max pooling. Average pooling is taking the average value of the selected area, while max pooling is taking the largest value of the selected area [12].

The pooling layer is a screen that utilizes the feature map function as an input and then processes it with various statistical operations that have been implemented by the system being managed. This pooling is a layer that is used sequentially in a CNN architecture in a progressive manner. The purpose of using this pooling layer is to take the max pooling or average value of parts of the image. The use of pooling layers is to reduce the size of the image so that it can be easily replaced with another layer, namely the convolution layer [13].

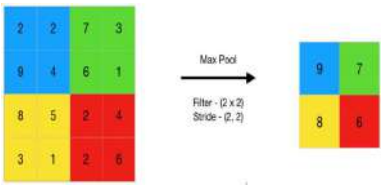


Fig 4. Pooling layer

Rectified Linear Units (RELU)

Rectified Linear Unit (RELU) is an activation function that has the advantage of being able to process large data quickly, which is used between the convolutional layer and the pooling layer. RELU maintains the results of the convolution image in a positive definite domain, so that every negative value that comes from the convolution process will go through the RELU process, and make the negative value equal to 0 [12].

$$f = \max(0, x)$$

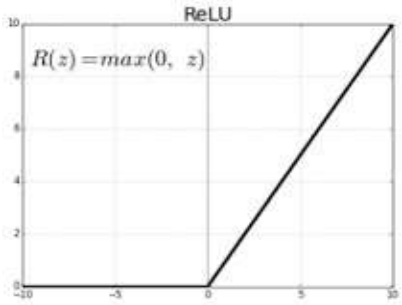


Fig 5. Curve of ReLu function

Softmax

Softmax is an activation function that is often used in neural networks that have many output categories (multi-class). The softmax function changes calculated values into probability values, this makes the calculated values comparable. By using the softmax function, it can be seen which class has the greatest possible value. Then, the biggest possibility will be the selected class and the next input will be classified into that class [12].

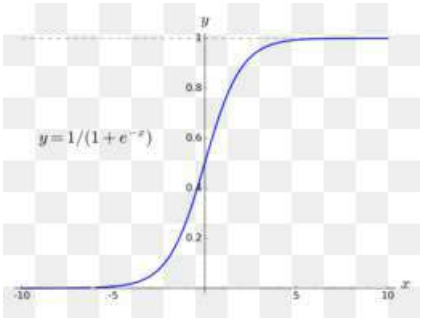


Fig 6. Curve of ReLu function

Fully Connected Layer

Fully Connected Layer is part of the neural network architecture where all data will be converted into one dimension. This process is known as flatten, which changes the dimensions of the data. This layer consists of nodes that are interconnected, and have weights and activation functions. The output of this layer is a prediction based on input data [12].

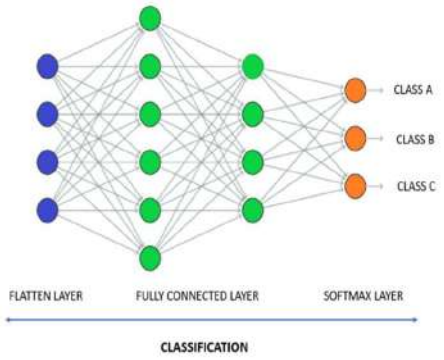


Fig 7. Fully Connected Layer

Confusion Matrix

Confusion matrix is a tool used to evaluate a classification model with the aim of estimating the truth or error of objects. This is in the form of a matrix of predictions that will be compared with the original input class or in other words, contains information about the actual and predicted values in the classification [16]. The Confusion Matrix is a way to assess the performance of a classification model, namely through the accuracy of the model. Some terms that are important in determining the level of accuracy are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These terms are usually combined in a matrix known as a confusion matrix, as shown below [14].

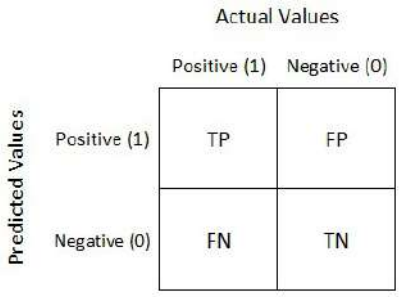


Fig. 8 Confusion Matrix

The accuracy value in classification is the percentage of accuracy of data records that are classified correctly after testing the classification results. Calculation of accuracy with the confusion matrix is as follows.

Accuracy = (TP+TN) / (TP+FP+FN+TN)

Tensorflow

TensorFlow, a freely available open-source software library, serves multiple purposes, with its primary emphasis being on neural network training and

inference. This library operates on a dataflow model and utilizes programming techniques [15].

The sampling technique used in this research is Random Sampling Technique, in which the sample determination technique ensures that each analytical unit has an equalchance and is considered to represent a population.

The data I will use for this study isobtained from two sources. The first source is data collection conducted by AwalTry (2020),which can be accessed from the website https://www.kaggle.com/datasets/awaltry/re mpah. The second source involves manualdata collection using a smartphone. From theabove data, it is divided into 500 training data,100 validation data, and 25 test data.

In the application of the convolutional neural network method for classifying spices, Confusion Matrix is employed. Confusion Matrix is a technique used to calculate accuracy in the context of data mining [9]. The Confusion Matrix is a table that represents the classification of the correct and incorrect test data.

An example of a Confusion Matrix for binary classification is shown as follow.

TABLE I. CONFUSION MATRIX

		Prediction	
		Positive	Negative
Class	Positive	TP	FP
	Negative	FN	TN

Explanation:

TP (True Positive) = Positive data correctlypredicted.

TN (True Negative) = Negative data correctly predicted.

FP (False Positive) = Negative data incorrectly predicted as positive.

FN (False Negative) = Positive data incorrectly predicted as negative.

Confusion matrix formulas to calculate accuracy, precision, and recall are as follows:

- Accuracy is a measure of how accurate a model is in predicting correct outcomes. It is calculated by dividing the total correct predictionsmade by the model by the total number of predictions made.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

- Precision is a measure of how accurately a model predicts positive outcomes. It is calculated by dividing the number of true positives (correct positive predictions) by the total number of positive predictions made by the model.

Precision = TP / (TP + FP)

- Recall is a measure of how accurately a model detects all positive occurrences. It is calculated by dividing the number of true positives (correct positive predictions) by the total number of actual positive occurrences.

Recall = TP / (TP + FN)

Since this study uses the SUS (System Usability Scale) method, the data analysis technique employed will use the SUS score calculation formula. For each respondent, it can beformulated as follows:

- For each odd-numbered question (1, 3,5, 7, 9), subtract 1 from the score (X-1).
- For each even-numbered question (2, 4,6, 8, 10), subtract its value from 5 (5-X).

Add the values of the even and odd-numbered questions. Then multiply the sum by 2.5. By following the procedures explained earlier, you can calculate the SUS score for eachrespondent and then calculate the averagescore to obtain the overall SUS score.

III. RESULT

An evaluation was conducted on the accuracy of the created model using aconfusion matrix. After the data processing is completed, the Confusion Matrix is employedfor evaluation, revealing the values of TP, TN, FP, and FN for each class. This yields accuracy, precision, recall, and f1-score values.

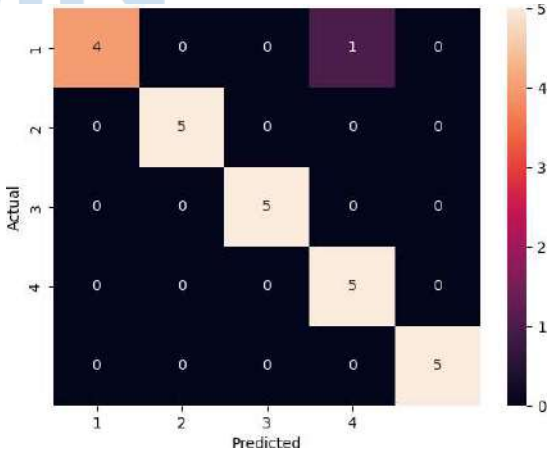
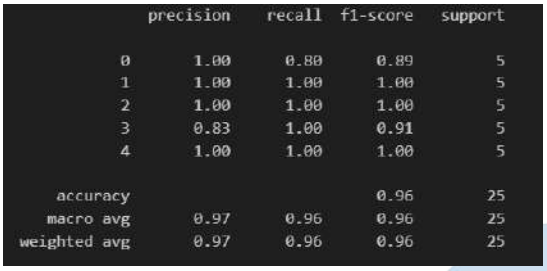


Fig 9. Confusion Matrix's Result

TABLE II. CONFUSION MATRIX'S RESULT

Class	*True Positive (TP)*	*False Positive (FP)*	*False Negative (FN)*
Jahe	4	0	1
Kencur	5	0	0
Kunyit	5	0	0
Lengkuas	5	1	0
Temulawak	5	0	0



	precision	recall	f1-score	support
0	1.00	0.80	0.89	5
1	1.00	1.00	1.00	5
2	1.00	1.00	1.00	5
3	0.83	1.00	0.91	5
4	1.00	1.00	1.00	5
accuracy			0.96	25
macro avg	0.97	0.96	0.96	25
weighted avg	0.97	0.96	0.96	25

Fig 10. Accuracy,Precision,Recall,F1-Score

1. Accuracy

Based on the above figure, it can be observedthat the evaluation results of the model using theconfusion matrix approach have an accuracy value of 0.96, indicating that the model is sufficiently suitable for use.

2. Precision

TABLE III. PRECISION RESULT EACH CLASS

Class	Precision
Jahe	1,00
Kencur	1,00
Kunyit	1,00
Lengkuas	0,83
Temulawak	1,00

The precision score is computed by comparing the number of correct positive predictions to the total number of positive predictions. The evaluation outcomes presented in the table above reveal the precision scores for individual classes as well as the overall precision score.

$$\text{Precision} = \frac{\text{P(Jahe)} + \text{P(Kencur)} + \text{P(Kunyit)} + \text{P(Lengkuas)} + \text{P(Temulawak)}}{\text{Sum of Class}}$$

$$\text{Presisi} = \frac{1{,}0 + 1{,}0 + 1{,}0 + 0{,}83 + 1{,}00}{5}$$

$$\text{Presisi} = 0{,}97$$

3. Recall

TABLE IV. Recall Result Each Class

Class	Recall
Jahe	0,80
Kencur	1,00
Kunyit	1,00
Lengkuas	1,00
Temulawak	1,00

The Recall value, also known as sensitivity, represents the proportion of accurate positive predictions in relation to the overall true positives within the dataset. Examining the data table presented earlier, we can see that the sensitivity values for different classes are as follows: 0.8 for the Ginger class, 1.00 for the Lesser Galangal class, 1.00 for the Turmeric class, 1.00 for the Galangal class, and 1.00 for the Javanese Ginger class.

$$Recall = \frac{\text{R(Jahe)} + \text{R(Kencur)} + \text{R(Kunyit)} + \text{R(Lengkuas)} + \text{R(Temulawak)}}{\text{Sum of Class}}$$

$$Recall = \frac{0{,}8 + 1{,}0 + 1{,}0 + 1{,}0 + 1{,}0}{5}$$

$$Recall = 0{,}96$$

4. F1-Score

The F1-Score represents an average of precision and recall, as indicated in the table above. In the case of the Ginger class, where Precision is 1.00 and Recall is 0.8, the F1 Score is calculated to be 0.89. Similarly, for the Lesser Galangal class, where Precision is 1.00 and Recall is 1.00, the resulting F1 Score is 1.00. The Turmeric class, with a Precision of 1.00 and Recall of 1.00, also yields an F1 Score of 1.00. As for the Galangal class, which has a Precision of 0.83 and Recall of 1.00, the computed F1 Score is 0.91. Lastly, for the Javanese Ginger class, having a Precision of 1.00 and Recall of 1.00 results in an F1 Score of 1.00.

IV. DISCUSSION

After testing using the confusion matrix, an average precision value of 97% was obtained. The average Recall value is 96%. And an accuracy of 96% was achieved. The results above indicate that the outcomes obtained are quite favorable in terms of accuracy, precision, and recall. This is demonstrated by comparing them with the research conducted by Evan Tanuwijaya andAngelica Roseanne in "Modification of

VGG16 Architecture for Classification of Indonesian Spices Digital Images." In that study, an accuracy rate of 85%, recall value of 80%, and precision value of 84% were obtained using the VGG16 base model [10].

TABLE V. Comparison Result

	Researcher Results	Previous Research Results
Accuracy	96%	85%
Precision	97%	84%
Recall	96%	80%

The researcher's results in the table are higher because the researcher used a larger dataset, specifically 500 training data,compared to the previous researcher whoonly used 100 training data.

Development Stage

During the development stage, the process began with creating a CNN model for classifying spices on the developed website. Firstly, the necessary libraries were imported, followed by preparing the dataset, including dividing it into classes and pre-processing the data. Subsequently, the CNN architecture was designed, with researchers opting for a custom-built model without adding a base model to it. Then, the data underwent training using the previously created model. Researchers utilized Google Colab for training the data due to its superior specifications for such tasks. Once the training was completed, the model was saved in .h5 format for implementation into the spice classification learning website.

Implementation Stage

During the implementation stage, the model was integrated into the website. The first step involved creating the website's interface, with researchers using a Bootstrap template. Next, researchers customized several aspects of the interface and added desired features. After designing the website's interface, researchers implemented the model using the Flask framework in Python and connected the back end with the front end. Lastly, the researchers uploaded the completed website to hosting to make it accessible to everyone.

Model Evaluation

After testing using a confusion matrix, the model yielded an average precision score of 97%, an average recall score of 96%, and an accuracy of 96%. These results indicate that the obtained outcomes are quite satisfactory in terms of accuracy, precision, and recall.

Evaluation of Spice Classification Learning Website

The media expert validation was conducted by an individual proficient in the field of learning media. The assessment instrument comprised four aspects: Presentation Design, Interaction Usability, Accessibility, and Reusability. The expert's evaluation resulted in a total score of 88 and an average score of 4.4, indicating an excellent rating.

The content expert validation was conducted at SMK N 4 Surakarta by experts in kitchen spices. The validation instrument consisted of 12 assessment points divided into four aspects: Content Quality, Learning Goal Alignment, Feedback and Adaptation, and Motivation. Overall, the validation results showed that the website received a total score of 54 and an average score of 4.5, indicating an excellent rating.

Thirty-two students from SMK N 4 Surakarta's culinary department participated in the usability test. After the test, the participants were asked to fill out a questionnaire assessing their response to the learning media. The questionnaire consisted of 10 assessment points. The results indicated that the students gave a total score of 76.4 and an average score of 30.56, indicating a good rating.

V. CONCLUSION

Based on the research findings, it can be concluded that the developed spice classification learning website is deemed suitable for use. Additionally, there are several strengths and weaknesses identified for further website improvement in future research.

Strengths:
1. The website can be accessed on various devices and screen resolutions.
2. Attractive website interface.
3. Comprehensive content.
4. High classification accuracy.

Weaknesses:
1. The website is still static.
2. Limited number of classes.
3. Lack of error handling mechanisms.

The use of the CNN method for classifying spices through a website-based learning media results in a relatively high accuracy rate of 96%. This figure is considered sufficient for implementation within the spice classification learning media website. With such a high accuracy result, the classification process within the spice classification website can perform well. This spice classification website can serve as an effective means to enhance understanding and mastery of concepts taught in the subject of spice introduction. With the presence of this technology, teaching and learning become more innovative, enjoyable, and provide significant benefits for learners.

REFERENCES

[1] Hambali, E., & Permanik, R. (2006). *Membuat Bumbu Instan Kering*. Penebar Swadaya Grup.

[2] Sinatra, C., Damajanti, M. N., & Milka, R. M. (2016). Perancangan Buku Pengenalan Rempah-rempah bagi Masyarakat Modern.*Jurnal DKV Adiwarna*, *2*(9), 9.

[3] Hikmatulloh, E., Lasmanawati, E., &Setiawati, T. (2017). Manfaat Pengetahuan Bumbu Dan Rempah Pada Pengolahan Makanan Indonesia Siswa Smkn 9 Bandung. *Media Pendidikan, Gizi, Dan Kuliner*, *6*(1).

[4] Putra, A. E., Naufal, M. F., & Prasetyo, V. R. (2023). Klasifikasi Jenis Rempah Menggunakan Convolutional Neural Network dan Transfer Learning. 9(1), 12–18.

[5] Muhasim, M. (2017). Pengaruh Tehnologi Digital terhadap Motivasi Belajar Peserta Didik. *Palapa*, *5*(2), 53–77. https://doi.org/10.36088/palapa.v5i2.46.

[6] Wulandari, I., Yasin, H., Widiharih, T., Statistika, D., & Diponegoro, U. (2020). Klasifikasi citra digital bumbu dan rempah dengan algoritma convolutional neural network (cnn) 1,2,3. 9, 273–282.

[7] Maydiantoro, A. (2020). Model Penelitian Pengembangan. *Chemistry Education Review (CER)*, *3*(2), 185.

[8] Hidayat, F., & Nizar, M. (2021). Model Addie (Analysis, Design, Development, Implementation and Evaluation) Dalam Pembelajaran Pendidikan Agama Islam. *Jurnal Inovasi Pendidikan Agama Islam (JIPAI)*, *1*(1), 28–38.https://doi.org/10.15575/jipai.v1i1.1104.

[9] Pratiwi, B. P., Handayani, A. S., & Sarjana, S. (2021). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix. *Jurnal Informatika Upgris*, *6*(2), 66–75. https://doi.org/10.26877/jiu.v6i2.6552.

[10] Tanuwijaya, E., & Roseanne, A. (2021). Modifikasi Arsitektur VGG16 untuk Klasifikasi Citra Digital Rempah- Rempah Indonesia Classification of Indonesian Spices Digital Image using Modified VGG 16 Architecture. *Jurnal Manajemen, TeknikInformatika, Dan Rekayasa Komputer*, *21*(1), 191–198. https://doi.org/10.30812/matrik.v21i1.xxx.

[11] Yunanto, R., Purfini, A. P., & Prabuwisesa, A. (2021). *Survei Literatur : Deteksi Berita Palsu Menggunakan Pendekatan Deep Learning*. *xx*, 118–130. https://doi.org/10.34010/jamika.v11i2.493

[12] Budi, R. S., Patmasari, R., & Saidah, S. (2021). KLASIFIKASI CUACA MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN) WEATHER CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK (CNN) METHOD. 8(5), 5047–5052.

[13] Mustafi, K., Prima, A., Dimas, N., & Arya, M. (2022). *Klasifikasi sampah menggunakan Convolutional Neural Network*. *3*(2), 72–81. https://doi.org/10.56705/ijodas.v3i2.33

[14] Wulandari, I., Yasin, H., Widiharih, T., Statistika, D., & Diponegoro, U. (2020). Klasifikasi citra digital bumbu dan rempah dengan algoritma convolutional neural network (cnn) 1,2,3. 9, 273–282.

[15] Primatama, Y., Rhamadani, A. E., Ramtomo, F. D., Cahya, D., & Buani, P. (2018). *Menggunakan Pemindai Wajah Berbasis Android*. 59–65.

[16] Nawangsih, I., Melani, I., Fauziah, S., & Artikel, A. I. (2021). Pelita Teknologi Prediksi Pengangkatan Karyawan Dengan Metode Algoritma C5.0 (Studi Kasus Pt. Mataram Cakra Buana Agung. *Jurnal Pelita Teknologi*, *16*(2), 24–33.

[17] H. Budianto, T. Khalimi, R. Ismaya, E. Kurniadi, and E. Dharmawan, "Implementation of tensor flow-based deep learning in the learning application of around things in English," J. Phys. Conf. Ser., vol. 1933, no. 1, p. 012007, 2021.

[18] J. G. Pires, "Machine learning in medicine using JavaScript: building web apps using TensorFlow.js for interpreting biomedical datasets," bioRxiv, 2023.

[19] N. Mohd Ariff Brahin, H. Mohd Nasir, A. Zakwan Jidin, M. Faizal Zulkifli, and T. Sutikno, "Development of vocabulary learning application by using machine learning technique," Bull. Electr. Eng. Inform., vol. 9, no. 1, pp. 362–369, 2020.

[20] S. Saitou et al., "Application of TensorFlow to recognition of visualized results of fragment molecular orbital (FMO) calculations," Chem-Bio Inf. J., vol. 18, no. 0, pp. 58–69, 2018.

[21] B. Satya, Hendry, and D. H. F. Manongga, "Object detection application for a forward collision early warning system using TensorFlow lite on android," in Third Congress on Intelligent Systems, Singapore: Springer Nature Singapore, 2023, pp. 821–834.

ISSN 2085-4552

Comparing Karate Framework with Others for Automated Regression Testing: A Case Study of PT Fliptech Lentera Inspirasi Pertiwi

Afina Putri Dayanti[1], Tony Tony[2]
[1, 2] Department of Information Systems, Faculty of Information Technology, Tarumanagara University
Jakarta, Indonesia
[1]afina.825200049@stu.untar.ac.id, [2]tony@fti.untar.ac.id

Accepted 10 November 2023
Approved 19 April 2024

***Abstract*—In the rapidly evolving digital era, applications, and software systems increasingly rely on Application Programming Interfaces (APIs) to enable interaction, integration, and functionality extension. However, manual testing of APIs is often inefficient and challenging to reuse when changes occur. To address this, automation testing has become a more effective choice, where test scripts can verify and execute tests repeatedly, easily adapting to API changes. Essentially, automation testing plays a vital role in software maintenance, particularly in regression testing, which tests modified or upgraded software versions to ensure that their core functions remain unchanged and unaffected. One approach to automation testing is employing the Software Testing Life Cycle (STLC), which follows a systematic series of stages conducted by the testing team to ensure software product quality. This paper utilizes PT Fliptech Lentera Inspirasi Pertiwi's public API to conduct testing on 25 scenarios from two modules. The objective is to utilize the Karate Framework to conduct these automated regression tests, resulting in an impressively short testing duration, averaging only 42.645 seconds, or approximately 1.706 seconds per scenario. A comparison with the Behave framework, using the same scenarios but with differences in steps, reveals that Behave achieves a duration of 18.762 seconds, or 0.750 seconds per scenario, making it 127.295% faster than Karate. However, in terms of the number of steps, Behave covers only 188, while Karate includes 543. This means that Behave requires 0.100 seconds per step, while Karate necessitates 0.079 seconds per occurrence. Karate provides more detailed results by 188.830% per step or 26.582% in terms of step duration. The primary goal is to enhance testing efficiency, expedite issue identification and resolution, provide a clearer testing process, and potentially improve overall software quality.**

***Index Terms*—API; automation testing; Karate framework; regression testing; STLC.**

I. INTRODUCTION

In the era of advancing information technology, software continually undergoes development and refinement to align with ever-changing needs and the rapid progression of technology [1]. Each change applied to the software, whether it involves improvements, feature additions, or other modifications, has the potential to influence the overall performance and stability of the entire system. Consequently, the significance of regression testing has increasingly become an essential foundation. This type of testing ensures that any alterations do not disrupt the functionality that was previously operating smoothly. The primary objective of regression testing is to verify whether these changes have caused disruptions in pre-existing functions, with the purpose of mitigating the risk of potential system failures resulting from these modifications [2].

PT Fliptech Lentera Inspirasi Pertiwi, aka Flip, is an Indonesian fintech company established in 2015 [3]. As a prominent player in the software development industry, the company encounters challenges similar to those faced by its industry peers while managing its multifaceted operations. The developer team engaged in various projects and features is tasked with ensuring the stability of the system. However, manual regression testing consumes significant time and resources. In this context, the implementation of an automation testing tool through an Application Programming Interface (API) emerges as a practical solution. APIs find widespread application in the creation of distributed software systems featuring interconnected components. Despite their absence of a visible interface [4], APIs play a pivotal role in enabling machine-to-machine communication and serving as a means to foster interaction, integration, expansion, and data exchange among distinct software functions or entities [5].

This paper contains the following contributions. Firstly, it provides insights into the challenges faced by Flip and the importance of system stability in software development. Secondly, it sheds light on the resource-intensive nature of manual regression testing and advocates for the use of automation testing, which involves executing tests through specialized tools or software [6] via APIs, as a solution to these challenges.

Thirdly, it promotes the use of the Karate Framework, an automation testing tool that utilizes Gherkin syntax from Cucumber BDD (behavior-driven development). While Karate is based on Java, it does not always require advanced programming skills for basic software testing. Instead, it encourages a deeper understanding of Cucumber and specific framework development [7]. Further, this paper conducts a case study using Flip's public API, the Flip for Business Platform, to create a specialized tool for regression testing. Fourthly, it anticipates reduced time and resource requirements for issue detection following system changes. Although automation can accelerate the testing process, it is essential to ensure that the benefits outweigh the initial setup and maintenance expenses. Lastly, it emphasizes the potential enhancement in testing efficiency, accuracy, and coverage through an approach of automation testing for regression.

The rest of this paper is organized as follows. Section II discusses related works to the research while Section III presents the research's results and discussion. The details of our solution and its performance are described in Section IV. Finally, Section V concludes the paper and provides future research directions.

II. RELATED WORKS

To the best of our knowledge, there has been no prior work addressing a problem similar to ours, except for the study conducted by Gidvarowart et al. [8]. In their work, the authors explored automated API testing using the Karate Framework and presented a case study of an online assessment web application demonstrating reduced testing time per iteration in contrast to manual testing and comparisons with other frameworks. Our approach extends to a comparison with the Behave framework, closely associated with the Python programming language and commonly identified using the search terms "bdd" and "behave" [9]. We opted for this framework because Behave incorporates a concept similar to Karate, namely BDD, and our aim is to compare them to identify more efficient regression testing automation.

When juxtaposing our work with other related studies, several distinct patterns emerge. Putri [10] applied regression automation testing using Katalon Studio for the "Teman Diabetes" mobile app, citing the perceived limitations of manual regression testing. In contrast, this work centers on regression testing for Flip for Business's API, presenting a different context. Directed attention, Yutia [11] highlighted automated functional API testing using the Robot framework for KALcare.com, emphasizing the efficiency gains brought about by automation over manual methods. In this proposed approach, the STLC methodology is applied to API regression testing, with the Karate framework harnessed for test execution. Additionally, automated load and performance testing for DiTenun's API were extensively explored by Barus et al. [12], while this work specifically highlights regression automation testing executed after system changes. Lastly, automation testing challenges in the context of Hospital Management Systems were discussed by Saputra and Stefanie [13], with this work predominantly focusing on API testing, where the Karate framework serves as the primary automation testing tool, contributing to the growing body of research on automation testing within various contexts.

III. METHODOLOGY

Testing is a process with its own stages, even though it's an integral part of the Software Development Life Cycle (SDLC) [14], see Fig. 1. Software Testing Life Cycle (STLC) refers to a systematic series of stages run by the testing team to test the software product. In essence, STLC constitutes the testing phase embedded within the SDLC, running in parallel but with its distinct cycle [15].

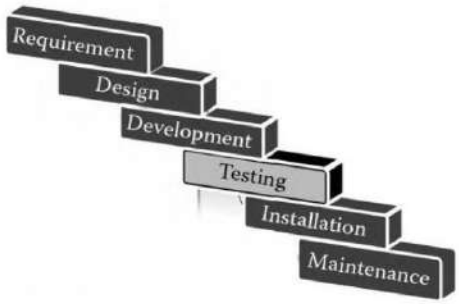


Fig 1. Testing Phase in Software Development Life Cycle [16]

In the design process, STLC approach serves as the primary methodology. Although the implementation of testing may vary depending on each SDLC's specific approach, the key steps in each software STLC remain consistent. This STLC approach provides a structured framework for governing all software testing stages, much like SDLC, and consists of stages (see Fig. 2) [16]. Each phase is designed to enhance the quality of the product [17] and is characterized by well-defined entry and exit criteria, activities, and associated deliverables [18].

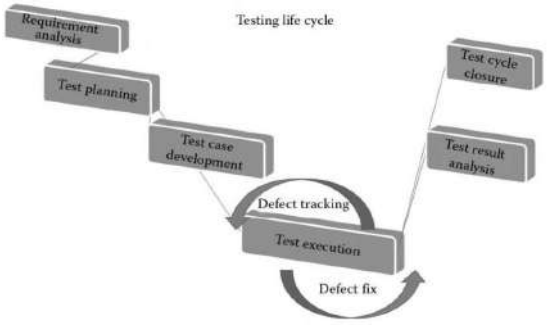


Fig 2. Stages of the Software Testing Life Cycle [16]

1) Requirements Analysis: This phase involves identifying the target, goals, scope, and testing approach to be taken. The plan will provide detailed explanations of how regression automation testing will be conducted, including resource allocation and testing schedules.
2) Test Planning: Analyzing testing requirements in detail. This includes an analysis of functional, non-functional features, and relevant testing scenarios for selected API features.
3) Test Case Development: Designing testing scenarios and automation testing scripts based on the requirements analysis results. The testing plan includes test steps, test data, and the test environment.
4) Test Execution (defect tracking and fixing): Involves creating and executing automation testing scripts according to the plan. Automation testing tools are developed using API concepts to test selected features. After obtaining test results, if defects are found, the next step is to return them to this phase for further analysis. Every bug and error in the API will be thoroughly analyzed. Afterward, corrective actions and updates will be implemented to address these defects.
5) Test Result Analysis: A post-conditional process involving data collection from end-users. After testing execution, the team evaluates the results of regression testing.
6) Test Cycle Closure: Discussion and evaluation of testing artifacts to identify strategies to be applied in the future, using the experience gained from the completed regression testing cycle. The goal is to reduce process constraints in subsequent testing cycles and share best practices for similar projects in the future.

IV. Result and Discussion

The Institute of Electrical and Electronics Engineers (IEEE) defines a process as the actions required to carry out a task or as a written description of those actions, as in documented testing procedures. From this perspective, it can be concluded that "the testing process" is the "actions necessary to carry out testing" or the "approach used in conducting testing" [19]. In the design of this process, we refer to the STLC methodology to create a structured overview of the API automation process, as shown in Fig. 3, where several key stages are detailed in the form of a flowchart. Commencing with "Start", the process navigates through stages such as analyzing requirements, selecting APIs, and conducting manual API inspections. Subsequently, the automation structure is aligned with the folder hierarchy, and testing scripts are created or adjusted. The results of automation testing are analyzed; if the outcome is "no", manual testing is revisited, while "yes" results lead to presenting the outcomes. System artifact evaluation ensures a thorough review, and the process loops back for continuous testing. The cycle concludes with the "End" phase, marking the conclusion of the API automation testing workflow.

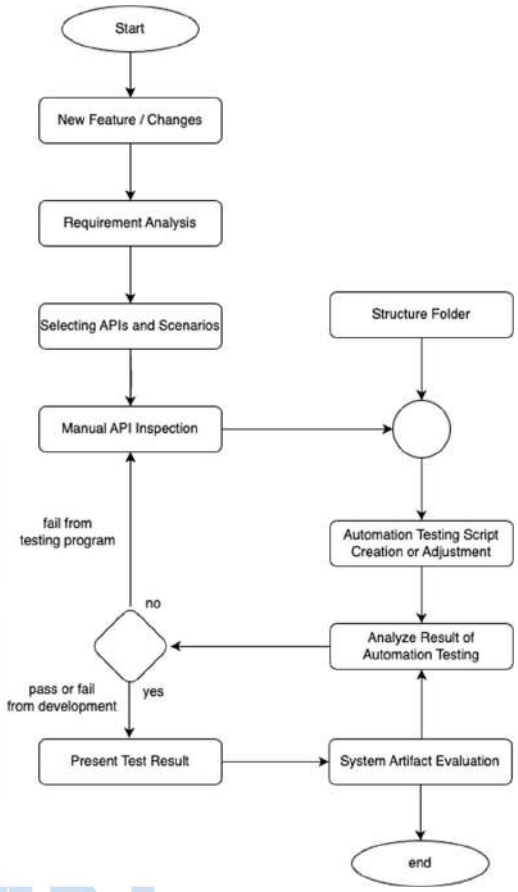


Fig 3. Flowchart of Automation Testing Process

A. Folder Structure

In the context of designing tests using Karate, it is different from Java code development that follows conventions like com.mycompany.foo.bar and results in nested sub-folders. The Karate documentation actually suggests having a folder structure with only one or two levels, where the folder names clearly identify the resource, entity, or API under test. However, in this design, we choose to adopt an automation structure customized to the company's needs (see Fig. 4).

Initially, all files are placed in the **src-test** folder. The **src-test-java** folder is used to store all automation files, including the Karate runner for execution and HTML report generation. The service folder is assigned for storing scenarios (.feature), the config folder for global configuration, the spec folder for payload requests, and the utils folder for utility data. Meanwhile, the **src-test-resource** folder is designated for application configuration files.

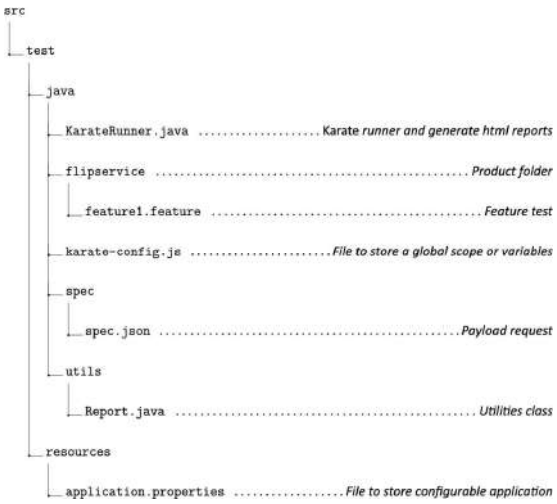


Fig 4. Folder Structure

B. Requirements Analysis

The process commences with the identification and analysis of the requirements and specifications of Flip for Business APIs. The required data is acquired through an interview with one of the Test Engineers from the Business and Solution Team responsible for Flip for Business. This data-gathering process involves collecting information about the relevant API usage, analyzing the API's flow, and identifying its integration with the database.

C. Selection APIs and scenario

The selection of APIs and scenarios involves a process that encompasses defining use cases, endpoints, actions, and test scenarios. Within Flip for Business, there are numerous use cases to choose from. However, in this study, only two modules will be selected, i.e., Money Transfer and Special Money Transfer [20] because these modules are among the most frequently used and represent the core of Flip's business operations, thus significantly contributing to Flip for Business' revenue. In each of the selected modules, there are four endpoints. The Money Transfer module has a total of 15 scenarios, while the Special Money Transfer module has a total of 10 scenarios. Therefore, for this research, the combination of these two modules results in a total of 25 scenarios. For more detailed information, the definition of the scope is based on the chosen use cases, as exemplified in Table I and Table II.

TABLE I. SCENARIO OF MONEY TRANSFER

No	Action	Endpoint	Scenario
1	POST	https://bigflip.id/api/v3/disbursement	Should success create disbursement with one beneficiary email
			Should return error create disbursement params required
			Should return error create disbursement only number
			Should return error create disbursement amount minimum
			Should return error create disbursement amount maximum
			Should return error create disbursement invalid bank code
			Should return error create disbursement max email
			Should return error create disbursement invalid email
2	GET	https://bigflip.id/api/v3/disbursement?pagination=pagination&page=page&sort=sort&attribute=value	Should success get all disbursement
			Should success get all disbursement with pagination
			Should success get all disbursement filter by status
			Should success get all disbursement filter by bank
			Should success get all disbursement filter by created from
3	GET	https://bigflip.id/api/v3/get-disbursement?idempotency-key=idempotencykey	Should success get disbursement by idempotency key
4	GET	https://bigflip.id/api/v3/get-disbursement?id=id	Should success get disbursement by id

TABLE II. SCENARIO OF SPECIAL MONEY TRANSFER

No	Action	Endpoint	Scenario
1	POST	https://bigflip.id/api/v3/special-disbursement	Should success create special disbursement for do mestic transfer
			Should success create special disbursement for foreign inbound transfer
			Should return error create special disbursement params required
			Should return error create special disbursement only number
			Should return error create special disbursement amount minimum
			Should return error create special disbursement amount maximum
			Should return error create disbursement invalid bank code
2	GET	https://bigflip.id/api/v2/disbursement/city-list	Should success get city list
3	GET	https://bigflip.id/api/v2/disbursement/country-list	Should success get country list
4	GET	https://bigflip.id/api/v2/disbursement/city-country-list	Should success get city and country list

D. Manual API Inspection

Before delving into the automation testing phase, it is highly advantageous to initiate the process with a manual API inspection. This initial step entails using software tools like Postman, which enable testers to manually interact with the API. By doing so, testers gain valuable insights into how the API functions and communicates. This manual exploration serves as a foundation for the subsequent phases and ensures a comprehensive understanding of the API's behavior. This phase is illustrated in Fig. 5, showcasing how manual testing aids in comprehending the intricacies of API interactions.

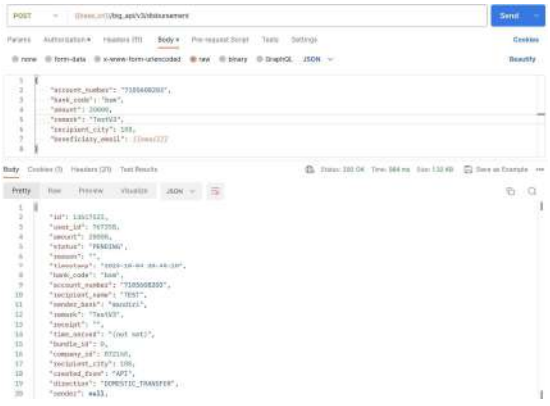


Fig 5. Testing Manually with Postman

E. Automation Script Creation or Adjustment

Before proceeding with the creation or modification of automation scripts, it's crucial to fulfill the prerequisites for system implementation, including software, hardware, personnel, installation procedures, and user guides. This phase is a pivotal point in the system's life cycle as it initiates the solution's deployment in the production environment. Proper software installation on prepared hardware, especially with the presence of personnel like test engineers, is vital. The installation process should ensure optimal system component functionality, meet performance standards, and provide user manuals for accurate system utilization.

Once the prerequisites for system implementation have been satisfied and gaining a comprehensive understanding of the API's requests and responses through manual methods, the next phase involves the transformation of this knowledge into automated tests. This step is essential for achieving efficiency and repeatability in the testing process. Automated tests are designed to mimic the interactions that were previously tested manually. The creation or adjustment of automation scripts allows for the seamless execution of these tests, as demonstrated in Fig. 6. Essentially, these scripts function as a comprehensive set of directives meticulously guiding the testing framework, effectively streamlining and optimizing the entire testing process for enhanced accuracy and efficiency.

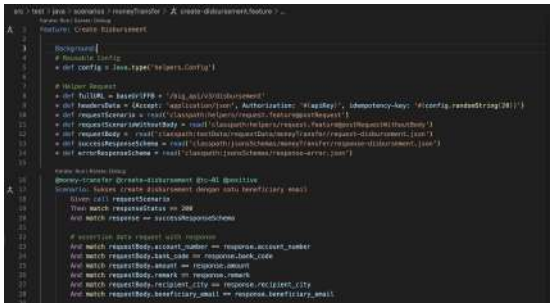


Fig 6. Transitioning from Manual Testing to Automation Testing

F. Analyze result of automation testing

In this phase, the focus shifts towards analyzing the outcomes of the automation testing process. After the automated tests have been executed as illustrated in Fig. 7, the results obtained need to be comprehensively examined and assessed. This entails scrutinizing the data, logs, and metrics generated during the testing process. One of the primary goals is to identify any anomalies, errors, or issues that might have surfaced during the automation testing. It's crucial to thoroughly evaluate the collected data to gain insights into the performance and behavior of the tested API. This phase serves as a critical checkpoint for quality assurance and ensures that the automated tests have been carried out effectively.

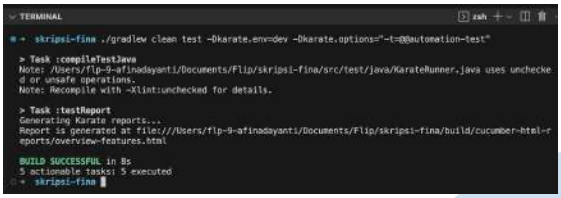


Fig 7. Automation Testing Execution

G. Present test result

The subsequent phase involves generating an HTML-formatted report to present the test results in a structured and informative manner. This comprehensive report accounts for various aspects of the testing process, including executed test cases, their outcomes, identified issues or defects, and performance metrics. This structured report plays a crucial role in facilitating in-depth test analysis, offering stakeholders a clear overview of the API's behavior and areas that need attention. Referenced as Fig. 8, it aids in problem identification and necessary improvements.

Fig 8. Automation Testing Execution

In the context of the initial selection of 25 scenarios, covering 8 endpoints across 2 previously chosen modules, the process of automation script creation and refinement resulted in a total of 543 occurrences. These scenarios underwent 5 consecutive trial runs, as seen in the variability of results in Table III, which are influenced by CPU and memory resources. The calculated average duration of 42.645 seconds implies that each scenario takes approximately 1.706 seconds to execute, equivalent to 0.079 seconds per occurrence.

TABLE IV. AUTOMATION TESTING ACROSS 10 TEST RUNS

Execute	Duration Result
1	43.580 s
2	42.693 s
3	45.216 s
4	40.311 s
5	43.254 s
6	42.111 s
7	42.918 s
8	41.289 s
9	43.246 s
10	41.832 s
Average	**42.645 s**

$$\text{Average Duration} = \frac{\text{Total Duration}}{\text{Trial Runs}} = \frac{426.45}{10} \approx 42.645\text{s}$$

$$\text{Average Scenario} = \frac{\text{Average Duration}}{\text{Total Scenario}} = \frac{42.645}{25} \approx 1.706\text{s}$$

$$\text{Average Occ.} = \frac{\text{Average Duration}}{\text{Total Occurrences}} = \frac{42.645}{543} \approx 0.079\text{s}$$

In addition to reporting test results in the .html format, there is also the possibility of integration, such as using the Slack platform. Through this integration, it becomes feasible to automatically send notifications when testing failures occur. For example, the integrated output within Slack, as depicted in Fig. 9, will provide immediate notifications to the development team when there are issues that require immediate attention during testing. Integrations like this ensure that information regarding testing problems

can be rapidly received by developers, allowing for more effective and timely responses to potential issues that may arise during the testing process.

Following the presentation of test results, a comprehensive evaluation of the results or report will be conducted. If any defects or issues are detected during this assessment, the testing process might regress to the manual testing phase to further investigate and resolve these problems. However, if no defects are found, the implementation will progress to the evaluation of system artifacts, ensuring the overall robustness and quality of the system.

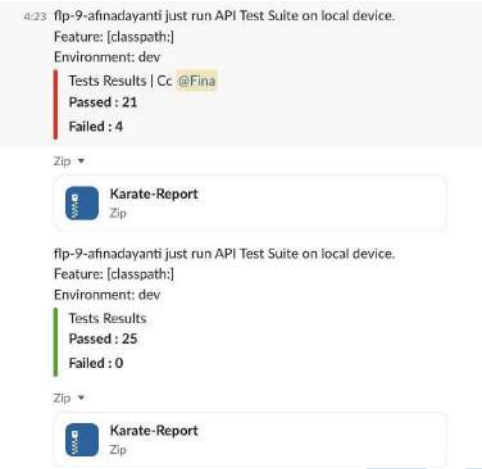


Fig 9. Automation Testing Report Integration with Slack

H. *System artifact evaluation*

The discussion and evaluation of testing artifacts aim to identify strategies that will be applied in the future, leveraging the experience gained from the on-going regression testing cycle. The goal is to minimize process constraints in the next testing cycle and share best practices for similar projects in the future.

I. *Comparison with Other Framework*

For the purpose of comparing the duration of test results, the author employed the behave framework. As detailed in Table IV, it became apparent that after conducting 10 test runs using the same 25 scenarios, encompassing 8 endpoints across 2 previously selected modules, with a slight difference in the number of steps (precisely 188 steps, as shown in Fig. 10), Behave achieved an average duration of 18.762 seconds. This suggests that each scenario takes roughly 0.750 seconds to execute, which translates to approximately 0.100 seconds per step.

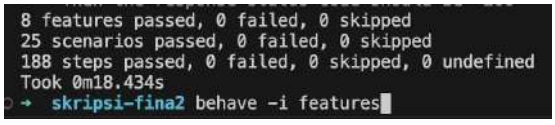


Fig 10. Execution of Automation Testing Using Behave Framework

TABLE IV. COMPARING THE DURATION RESULTS OF AUTOMATION TESTING FRAMEWORKS

Execute	Duration Result Karate	Duration Result Behave
1	43.580 s	19.215 s
2	42.693 s	18.551 s
3	45.216 s	18.293 s
4	40.311 s	18.176 s
5	43.254 s	19.385 s
6	42.111 s	18.898 s
7	42.918 s	18.972 s
8	41.289 s	17.978 s
9	43.246 s	20.180 s
10	41.832 s	17.969 s
Average	**42.645 s**	**18.762 s**

$$\text{Average Duration} = \frac{\text{Total Duration}}{\text{Trial Runs}} = \frac{187.617}{10} \approx 18.762\text{s}$$

$$\text{Average Scenario} = \frac{\text{Average Duration}}{\text{Total Scenario}} = \frac{18.762}{25} \approx 0.750\text{s}$$

$$\text{Average Step} = \frac{\text{Average Duration}}{\text{Total Occurrences}} = \frac{18.762}{188} \approx 0.100\text{s}$$

In evaluating the merits of these two options, we must consider testing objectives and priorities. Utilizing formulas to calculate the percentage increase or decrease can measure the extent of changes in a value [21]. In this context, it is necessary to identify the differences between the values of Karate and Behave first. The resulting difference is then divided by the value of Karate or Behave, depending on whether we are looking for a percentage increase or decrease. The result is then multiplied by 100 to convert it into a percentage.

$$\text{Percent decrease (\%)} = \frac{\text{Original Value} - \text{New Value}}{\text{Original Value}} \times 100\%$$

$$\text{Percent increase (\%)} = \frac{\text{New Value - Original Value}}{\text{Original Value}} \times 100\%$$

Behave boasts a faster execution time, completing tests in 18.762 seconds, compared to Karate's 42.645 seconds. This results in Behave reducing execution time by 127.295%, making it more efficient in terms of time and resource utilization compared to Karate. It's important to emphasize that this difference in duration cannot be attributed to a single factor. Instead, it's influenced by various variables, including test complexity, the testing environment, parallel execution, hardware and resource disparities, optimization, tool-specific factors, and more, all of which collectively contribute to these variations.

Percentage difference between the average duration of Karate and Behave:

$$\text{Percent decrease (\%)} = \frac{\text{Behave - Karate}}{} \times 100\% = \frac{18.762 - 42.645}{18.762} \times 100\% \approx -127{,}294\%$$

Moreover, shifting the focus to the number of steps, Behave employs 188 steps, whereas Karate uses 543 steps. This implies that utilizes's per-step duration is 0.100 seconds, while Karate's per-occurrence duration is 0.079 seconds. Consequently, Karate holds a 188.830% advantage in providing more detailed and comprehensive insights into the behavior of the tested software, particularly when considering step duration, where Karate outperforms Behave by 26.582%. Therefore, when deciding between these options, it's essential to consider the trade-off between execution speed and the depth of analysis while considering specific testing requirements and objectives.

Percentage difference between Karate and Behave steps:

$$\text{Percent decrease (\%)} = \frac{\text{Behave} - \text{Karate}}{\text{Behave}} \times 100\% = \frac{188 - 543}{188} \times 100\% \approx -188.830\%$$

Percentage difference between the duration of Karate and Behave steps:

$$\text{Percent increase (\%)} = \frac{\text{Behave} - \text{Karate}}{\text{Karate}} \times 100\% = \frac{0.100 - 0.079}{0.079} \times 100\% \approx 26.582\%$$

CONCLUSION

This paper utilizes the Karate Framework as an automation tool for testing to investigate the use of the public Flip for Business API during the development process. The execution of 25 scenarios selected from two modules resulted in an impressively short testing duration of only 42.645 seconds, which translates to approximately 1.706 seconds per scenario. This reduction in testing time is a significant improvement over manual methods, leading to substantial time savings of several minutes per test. Additionally, when compared with the Behave framework using the same scenarios but with differences in steps, Behave achieved 18.762 seconds, or 0.750 seconds per scenario, making it 127.295% faster than Karate. However, when considering the number of steps, Behave only covers 188 steps, while Karate includes 543 steps. This means that Behave requires 0.100 seconds per step, while Karate requires 0.079 seconds per occurrence. Karate provides more detailed results by 188.830% per step or 26.582% in terms of step duration. Therefore, the choice between Behave and Karate depends on your primary testing objectives. Since the main goal of this paper is efficiency to obtain results as quickly as possible and in-depth analysis, Karate will be the preferable choice. This acceleration in the testing process contributes to faster development cycles, ensuring consistent API quality with each modification. Additionally, the quality assurance report verifies the online assessment system's quality and deployment readiness. Thorough preparation is essential for seamless parallel testing, avoiding conflicts and overlaps in test cases. An immediate area of future work involves integrating this procedure into CI/CD (Continuous Deployment or Continuous Delivery) solutions, to speed up release cycles and address potential issues during code integration.

ACKNOWLEDGEMENT

We acknowledge the support from Muhamad Rizal Indrabayu as the Engineering Manager, Henry Suryawirawan as the Vice President of Engineering, and Dwina Apriliasari as Corporate Communications at PT Fliptech Lentera Inspirasi Pertiwi.

REFERENCES

[1] J. A. Gerding, B. W. Brooks, E. Landeen, and more, "Identifying needs for advancing the profession and workforce in environmental health," American Journal of Public Health. 2020 Mar, 110(3):288-94.

[2] G. Blokdyk, Regression Testing: A Complete Guide - 2019 Edition. Emereo Pty Limited, 2019.

[3] Flip. (2023) Tentang flip. Accessed on August 10, 2023. [Online]. Available: https://flip.id/tentang-flip

[4] M. Biehl, API Architecture. CreateSpace Independent Publishing Platform, May 2015.

[5] AWS. (2023) What is an api? Accessed on 7 August 2023. [Online]. Available: https://aws.amazon.com/id/what-is/api/

[6] M. Baumgartner, T. Steirer, M.-F. Wendland, S. Gwihs, R. Seidl, and more, Test Automation Fundamentals: A Study Guide for the Certified Test Automation Engineer Exam – Advanced Level Specialist – ISTQB® Compliant. dpunkt.verlag, August 30 2022.

[7] P. A. Chaubal, Mastering Behavior-Driven Development Using Cucumber. BPB Publications, August 2021.

[8] S. Gidvarowart, A. Suchato, D. Wanvarie, N. Pratanwanich, and N. Tuaycharoen, "Automated api testing with karate framework: A case study of an online assessment

web application," in 2023 20th International Joint Conference on Computer Science and Software Engineering (JC- SSE). Phitsanulok, Thailand: IEEE, June 28 - July 01 2023.

[9] T. Storer and R. Bob, "Behave nicely! automatic generation of code for behaviour driven development test suites," in 2019 19th International Working Conference on Source Code Analysis and Manipulation (SCAM), 2019, pp. 228–237.

[10] Y. F. Putri, "Automation regression testing pada aplikasi teman diabetes dengan menggunakan metode black box testing," Ph.D. dissertation, Universitas Atma Jaya Yogyakarta, 2020.

[11] S. N. Yutia, "Automated functional testing pada api menggunakan keyword driven framework," Journal of Informatics and Communication Technology (JICT), vol. 3, no. 1, pp. 65–78, 2021.

[12] A. C. Barus, J. Harungguan, and E. Manulu, "Pengujian api website untuk perbaikan performansi aplikasi ditenun," Journal of Applied Technology and Informatics Indonesia, vol. 1, no. 2, pp. 14–21, 2021.

[13] B. D. Saputra and A. Stefanie, "Automation testing api, android, dan website menggunakan serenity bdd pada software sistem manajemen rumah sakit," Jurnal Ilmiah Wahana Pendidikan, vol. 9, no. 10, pp. 114–126, 2023.

[14] S. Desai and A. Srivastava, Software Testing. Phi Learning, January 30 2016.

[15] G. Singh, "A study on software testing life cycle in software engineering," Int. J. Manag. IT, vol, vol. 9.

[16] A. S. Mahfuz, Software Quality Assurance: Integrating Testing, Security, and Audit. CRC Press, April 27 2016, ebook.

[17] A. Anand and A. Uddin, "Importance of software testing in the process of software development," International Journal for Scientific Research and Development, vol. 12, no. 6, 2019.

[18] A. Nordeen, Learn Software Testing in 24 Hours: Definitive Guide to Learn Software Testing for Beginners. Guru99, October 31 2020.

[19] R. Drabick, Best Practices for the Formal Software Testing Process: A Menu of Testing Tasks. Pearson Education, July 15 2013.

[20] Flip. (2023) Flip for business. Accessed on August 10, 2023. [Online]. Available: https://flip.id/business

[21] D. Sharma. (2023) How to calculate percentage increase: Formula & examples. Updated on August 1, 2023. Accessed on November 20, 2023. [Online]. Available: https://www.indeed.com/career-advice/career-development/percent-increase-formula

ISSN 2085-4552

Educational Game Design For Carbon Emission Using Game Development Life Cycle Method

Dewi Tresnawati[1], Sri Rahayu[2], Randi Maulana[3]
[1,2,3] Ilmu Komputer, Program Studi Teknik Informatika, Institut Teknologi Garut, Garut, Indonesia
[1] dewi.tresnawati@itg.ac.id, [2] srirahayu@itg.ac.id, [3] 1906052@itg.ac.id

Accepted 28 November 2023
Approved 27 March 2024

***Abstract*— Carbon emissions are gases that arise from human actions, such as burning fossil fuels and industrial waste. Climate change is currently a problem that is increasingly attracting the attention of the wider community around the world, including Indonesia. The educational game about carbon emissions applies the Game Development Life Cycle (GDLC) approach which consists of six stages, including Initiation, Pre-Production, Production, Testing, Beta, and Launch. The educational game on carbon emissions is expected to help raise youth awareness about the importance of reducing carbon emissions and provide information about efforts to reduce carbon emissions for the younger generation and the general public.**

***Index Terms*— carbon emissions; educational game; GDLC method.**

I. INTRODUCTION

Climate change is currently an issue that is increasingly attracting the attention of the wider community worldwide, including Indonesia. The introduction of climate change and public concern over issues arising from climate change have resulted in new environmental regulations in recent decades [1]. Efforts to conserve energy and reduce carbon emissions, as a very important action in dealing with climate change and promoting sustainable economic development, have attracted great attention from countries around the world [2], [3], [4]. Since the adoption of the Kyoto Protocol in 2005, governments around the world have made great efforts to reduce carbon emissions by various methods. However, total greenhouse gas emissions worldwide have remained stable and have not decreased [5]. The significant increase in global temperature has led to climate change that impacts the environment and human life. Based on data from the Global Carbon Atlas, Indonesia ranks 9th with the highest amount of carbon emissions in Southeast Asia in the last two years. The total emissions produced reached 619 Metric Tons of CO2 [6]. This shows how important it is to reduce carbon emissions in Indonesia to minimize the impact of climate change which is increasingly detrimental to humans and the environment. Carbon emissions themselves are gases produced from human activities, such as the burning of fossil fuels and industrial waste. These carbon emissions are the main cause of climate change occurring around the world, such as increasing global temperatures, melting polar ice caps, and changing extreme weather patterns [7]. At the same time, human activities are predominantly related to energy production, industrial activities, and those related to forestry, land use, and land use change [8]. Climate change is becoming increasingly serious, and every country is paying more attention to carbon emissions. These activities have an impact on the environment in which humans are located. The environment is a place of activity and interaction of living things that are interdependent on one another. Environmental sustainability plays an important role in the sustainability of the environment and ecosystem of living things. Environmental damage is caused by irresponsible human behavior so environmental damage has entered a very alarming stage [9]. The gradual increase in human awareness of the importance of environmental protection has encouraged active efforts to produce low-carbon products with higher intensity [10]. The role of the youth generation is needed considering that youth is a milestone of change, in the 2009 Constitution explains Youth is an Indonesian citizen who enters an important period of growth and development aged 16 (sixteen) to 30 (thirty) years [11]. Youth is an important factor because of their high fighting spirit, creative solutions, and innovative manifestations also have the potential to make positive changes to the environment and reduce carbon emissions in the future [12]. They are the group that will inherit this earth from the previous generation and potentially become the decision-makers of the future. Therefore, raising young people's awareness about the importance of reducing carbon emissions can help create sustainable behavior change in the long run. In addition, the younger generation also has easier access to technology and information that can be used to reduce carbon emissions, educational games are considered as one of the interesting ways to reduce carbon emissions. Based on these problems, an educational game about carbon emissions will be built.

II. METHODS

The methodology applied in this research is the Game Development Life Cycle (GDLC), GDLC is a game development process that applies an iterative approach consisting of 6 development phases, starting from the initialization/concept creation, pre-production, production, testing, and release phases [13], which are contained in Figure 1 below.

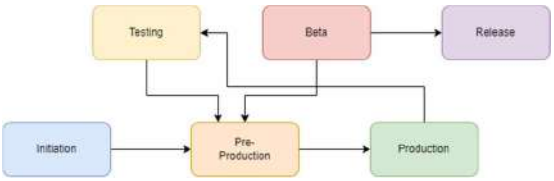


Fig. 1. GDLC Phase [13]

There are 6 phases used in the GDLC development method, namely:

1) Initiation, In the initial stage, problem identification, identification of needs in making games, identifying users, and identifying objects to have benefits in accordance with the purpose of making games.

2) Pre-Production, this stage is one of the main and important stages in the production cycle [14], planning will be carried out in the form of concepts and scenarios as well as features for the game to be created. Concepts are general ideas or ideas that underlie a project or product, including in terms of making games. Game concepts can include themes, gameplay mechanics, game characters and environments, and educational goals of the game.

3) Production, At this stage the author works on the core part of making the game. The work starts from collecting the required assets to coding and game development.

4) Testing, This stage is the stage where the game that has been made is tested to ensure its quality. At this stage, functional testing and bug fixes are carried out.

5) Beta, The Beta stage is the stage where the game is further tested by a small group of users. The goal is to find out the user's response to the game and make improvements if needed.

6) Release, The final stage of all game development processes is when the game is released to the public. This stage includes marketing the game, launching the game, and maintaining the game after its release.

Where the six stages are implemented into a series of activities presented in Figure 2.

Based on the picture of the framework, it is divided into three main parts in conducting this research.

a) The first stage, in this section, a series of activities are carried out such as Initial Identification and Analysis Initial identification in order to get results in the form of problem formulation, and research gaps to GDLC research methodology.

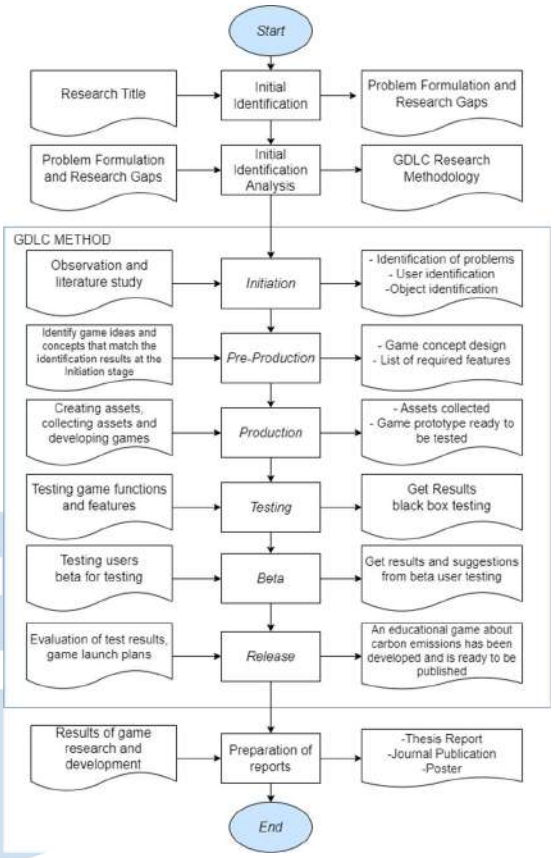


Fig. 2. Framework

b) The second stage, in this section, the software development process is carried out based on the stages in the GDLC methodology.

c) The third stage, in this last stage, is the preparation of reports and journals on the games that have been made.

III. RESULT AND DISCUSSION

A. Research results

The result of the research that has been done is the design and construction of a carbon emission educational game where the creation is by applying the GDLC method.

1) Initiation

Is an initial stage that involves creating a rough concept of the game, starting from determining the main characteristics of the game to be created [15].

a) Problem Identification

At this stage, problem identification is carried out, which is expected that youth can understand and internalize the importance of playing an active role in

reducing carbon emissions in everyday life, by presenting information about efforts to reduce carbon emissions, with educational results obtained by cleaning up garbage, saving energy, reducing disposable plastic waste, and planting trees.

b) User Identification

At this stage, user identification is carried out which is designed for the younger generation with an age range of 16 to 30 years. Gameplay that actively involves players, as well as content that is relevant to real life, all aim to provide a learning experience on the issue of carbon emissions.

c) Object Identification

At this stage, object identification is carried out where the objects include garbage, trees, electronics, and disposable plastic waste. These objects will be used in the context of the game to teach players about different aspects of carbon emissions with solutions.

2) Pre-Production

This stage is the process of planning, creating, and depicting sketches and other elements that form a unified whole in the game, involving the development of game prototypes along with the formation of the conception and basic design of the game [16]. This includes game concept and design, storyline, flowchart, map design, and storyboard.

a) Game Concept and Design

The concept and design of this carbon emission educational game invite players to understand the importance of reducing carbon emissions in everyday life. Players will play the role of city residents who strive to protect the environment and reduce the impact of carbon emissions through actions such as cleaning up trash, saving energy, planting trees, reducing disposable plastic waste, and adopting other environmentally friendly habits. The game features real-life situations that allow players to make decisions that impact carbon emissions.

b) Storyline

In RPG games, the storyline plays an important role because it is the main foundation of the game. It is told that Idnar is a young man who has lived in the big city for a long time, but he misses the village where he grew up and all the good memories there. He feels inspired by the environmental efforts he sees in the city and wants to contribute to protecting the environment and reducing carbon emissions in his own village.

c) Flowchart

This flowchart is organized based on the logic of thought that shows how the game will run, which is contained in Figure 3.

This game flow diagram explains that when the game starts, it displays the opening first then the player will enter the village intersection map where the map is the first time the player plays, after that, the player will explore the map until the player will meet with the NPC and will talk to the NPC about the mission, then if the player takes the mission then the mission will begin with different challenges faced depending on the mission taken after the player completes the mission then the mission will be completed. If the player does not take the mission, the player can explore again to find another mission. If all missions are completed, the player will be transferred to the transition ending, where a closing remark will appear and thank you for completing the game.

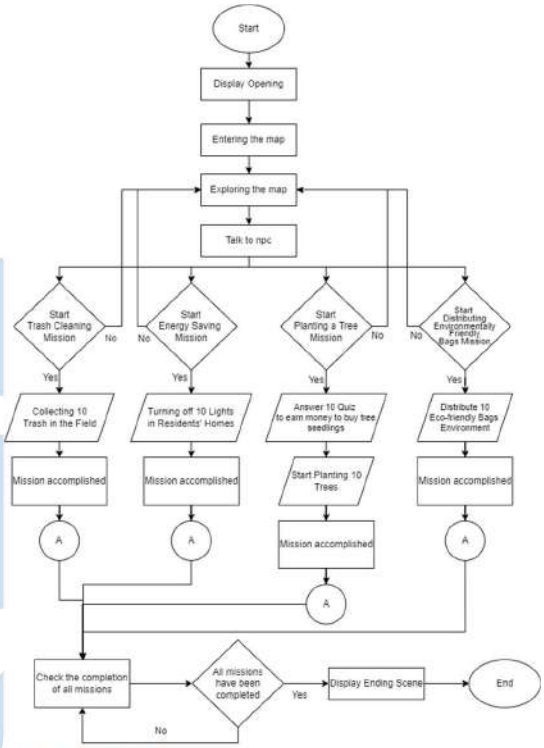


Fig. 3. Game flow diagram

d) Map Design

Is a description of the process of designing and creating the layout of the game environment in a game. The following is a description of the map that will be made in this research is contained in Figure 4:

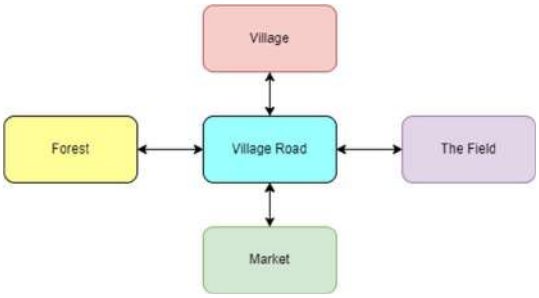


Fig. 4. Map design

In Figure 4 the map design describes how the process of designing and creating the layout of the game

environment, starting from the village road, field, village, forest, and market is described below:

Description of Mapping Design:

1) Village Road:

- The main road becomes the center of access to various areas in the village hall.

- Players will start the game on this main road.

2) Field:

- To the right of the main road is a large field.

- The garbage-clearing quest is located in this field. The player has to clean up the scattered garbage and complete the task of cleaning up the environment.

3) Village:

- Above the main road, there is a village with residents' activities.

- Players can talk to villagers, and get information about the story and life in the village.

- Some villagers may give side missions or tasks that are connected to the main story or to the daily activities of the villagers.

4) Forest:

- To the left of the main road, there is a shady forest that presents beautiful natural scenery.

- Tree planting missions and quizzes are found in this forest. Players must plant trees in designated areas to help protect the environment and answer the quiz.

5) *Market:*

- At the bottom of the main road, there is a market with various small shops. Players can shop at the market to purchase mission necessities, such as farming equipment and eco-bags.

e) Storyboard

The storyboard in this design aims to describe in detail the appearance and flow of the game to be created [17]. Users interact with the system through the interface. To ensure that users can easily use and interact with the system even if this is their first experience, the interface is designed to be user-friendly or easy. The following is an overview of the storyboard design of the carbon emission educational game contained in Table 1.

3) Production

The production stage in this game development includes several important steps that focus on producing collected assets and game prototypes that are ready to be tested. This stage involves the process of implementing the game concept that has been designed previously [18].

TABLE I. SUMMARY OF CARBON EMISSION EDUCATIONAL GAME STORYBOARD

No	*Name*	*Description*
1.	Beginning Storyboard	When the game is run, the user will be faced with the main menu display or the game's home page.
2.	Cutscene Storyboard	The appearance of the initial game narration text forms the flow of the main character in the game.
3.	Road Map Storyboard	Map design after the intro is complete, the main character will be placed on this map.
4.	Market Map Storyboard	Map design for the market street where the main character will go to a convenience store to buy an item.
5.	Village Conversation Storyboard	This view will emerge when the main character starts interacting with supporting characters where narration text appears.
6.	Field Map Storyboard	Map design for the field where the main character will start a mission.
7.	Market Shop Storyboard	Map design for a market shop where players when buying an item will appear first narration.
8.	Forest Storyboard	Map design for the forest where the main character will start a mission.
9.	Final Storyboard	The appearance of the end-of-game narration text that has completed the game.

a) Event Creation Production

Events in RPG Maker play a role in driving the story of the game. These events are created through the event editor which offers a variety of options, such as text display, show choices, switches, and variable control.

b) Character Assets

The main character to be played is named Idnar Analuam, called Idnar. The main character will later interact with supporting characters (Npc), namely Mr. Village Head, Mr. Coby, Mrs. Sarah, Mr. Heri, Mr. Budi, Mr. Tejo, and Mr. Bonbon.

Fig. 5. Character assets

c) Game Prototype

A game prototype is an early version of a game created to test and illustrate key features and basic gameplay. Initial Display Screen: This is the implementation of the storyboard into the game where it displays according to the storyboard, namely the game title, and buttons such as New Game, Continue, Options, and Game End. As shown in Figure 5.

4) Testing

The production stage in this game development includes several important steps that focus on producing collected whether the game function runs

optimally [19]. The results of black box testing are contained as follows: In the dynamic world of the game app, every scenario seamlessly unfolds with resounding success, from the melodic main menu opening to embarking on triumphant journeys with the "New Game" button, effortlessly continuing ongoing adventures with the "Continue" button, exploring personalized settings through the "Options" button, and gracefully concluding sessions with the "Game End" button, all while immortalizing accomplishments with the "Save Game" button and seamlessly revisiting checkpoints using the "Load" button, as players navigate through intricate details, unlock new dimensions in the "Item" view, engage in conversations and quests, traverse diverse environments, and successfully complete missions, contributing to a virtuous impact on the virtual world's environment and economy, while quizzes add an intellectual twist, choices contribute to the evolving storyline, noble missions showcase commitment to a sustainable virtual environment, and the climax rewards players with an ending scene and narration, ultimately finding themselves back at the initial menu, basking in the glow of a triumphant adventure, and ready for future digital escapades.

The results of testing with a total test cases using black box testing showed that the game was successfully run according to the expectations that had been set. All test cases were successfully passed without any significant problems. With these satisfactory results, it can be concluded that the game is ready to enter the beta testing stage.

Fig. 6. Early game prototype

5) Beta

The stage where testing is carried out by a third party or external [20]. At this stage there are test results, testing is carried out by applying the System Usability Scale (SUS) method, testing is carried out using a 10-question technique that has been specially adapted for testing this game. The testing involved as many as 25 young individuals, which included youth from a class at SMAN 11 Garut. The results of the beta testing are contained in Table 2.

Q1 assesses the perceived difficulty level of mission tasks, gauging players' challenges in completing game objectives. Q2 evaluates the clarity of instructions and objectives for each mission, focusing on players' comprehension of provided guidance. Q3 measures players' interest in replaying the game, determining its appeal for further engagement. Q4 examines the relevance of game features to carbon emission reduction, assessing alignment with the environmental theme. Q5 assesses players' understanding of the game's storyline, focusing on narrative comprehension. Q6 evaluates players' satisfaction with the game's visual aspects and design, considering graphics, interface, and overall presentation. Q7 measures the frequency of technical issues encountered during gameplay, assessing disruptions to game performance. Q8 evaluates the usefulness of game-provided information on carbon emission reduction, assessing its educational value. Q9 assesses the ease of interaction with the game's interface, focusing on user-friendliness. Q10 measures the likelihood of players recommending the game to friends, indicating overall satisfaction and endorsement potential.

The average score is obtained from the total SUS score divided by the total number of youth. With a total result of "79", according to the SUS assessment scale, it is included in the "B" category in the class scale with an adjective in the scope of "Good", this shows that the score is acceptable and also considered good and acceptable.

6) Release

At this stage, the game has been uploaded through the itch.io platform, a forum for indie game developers to share their work for free. With this step, the game can be accessed and played by users through the website.

B. Research results

The results of this study state that the creation of educational games using the quiz method and character interaction is alignment with previous research [21]. This study also applied the GDLC method in the development of educational games, which provides a systematic and structured approach to game design and development, improving the efficiency of the game creation process. In addition, this study used RPG Maker to facilitate the implementation of interactive educational games. This research aligns with previous research [22], which aims to increase the understanding and interest of the younger generation in a particular field of knowledge or expertise. In this study, young people were directed to better understand the important issue of carbon emissions and how they can contribute to reducing them.

The results and discussion of this study indicate that the research successfully designed and developed an educational game about carbon emissions through the application of the Game Development Life Cycle (GDLC) method. The main purpose of this game is to increase the understanding and awareness of the younger generation about the importance of reducing carbon emissions and their impact on the environment. The test results using the System Usability Scale (SUS) show that this educational game has a positive and

acceptable level of usability, with an average score of 79, indicating that this game is considered easy to use by players and has succeeded in increasing understanding of carbon emissions.

TABLE II. BETA TESTING

No	*Respondent*	*Question*										*Overall*	*Sus Score (Overall X 2.5)*
		Q1	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*	*Q7*	*Q8*	*Q9*	*Q10*		
1.	Respondent 1	4	5	4	3	4	3	3	4	3	3	36	90
2.	Respondent 2	3	4	3	4	3	4	3	4	3	2	33	78
3.	Respondent 3	2	3	4	4	4	4	4	5	4	2	40	95
4.	Respondent 4	5	4	2	2	2	2	2	3	2	2	26	73
5.	Respondent 5	3	3	4	3	3	3	3	3	3	2	32	90
6.	Respondent 6	4	4	3	4	4	4	3	4	3	3	36	75
7.	Respondent 7	2	3	5	4	4	3	3	1	3	3	33	65
8.	Respondent 8	5	4	4	4	4	4	4	2	4	3	44	90
9.	Respondent 9	4	5	3	3	3	2	2	3	2	2	29	98
10.	Respondent 10	3	4	5	3	4	3	3	3	4	4	41	75
11.	Respondent 11	4	3	2	4	3	3	3	3	3	2	30	98
12.	Respondent 12	3	4	4	3	2	2	2	2	2	2	26	70
13.	Respondent 13	5	3	3	4	3	4	4	4	3	3	37	93
14.	Respondent 14	3	5	5	4	5	4	3	3	4	3	43	75
15.	Respondent 15	2	4	4	2	3	3	3	3	3	3	30	78
16.	Respondent 16	4	5	4	4	4	4	2	4	4	4	41	65
17.	Respondent 17	3	4	2	3	3	3	3	3	2	2	28	65
18.	Respondent 18	3	3	5	5	4	4	3	4	3	3	39	78
19.	Respondent 19	4	4	4	4	3	2	2	3	2	2	30	73
20.	Respondent 20	3	3	3	3	3	3	3	3	4	3	37	63
21.	Respondent 21	2	4	1	2	3	3	3	3	3	2	31	90
22.	Respondent 22	4	3	4	2	2	2	2	3	2	2	28	78
23.	Respondent 23	4	4	3	3	4	3	3	1	3	3	37	95
24.	Respondent 24	3	3	3	3	3	3	3	3	3	2	32	73
25.	Respondent 25	4	3	2	3	2	2	2	3	2	2	29	90
Overall		86	94	86	83	82	77	71	77	74	64	794	1985
Average													79

IV. CONCLUSION

The conclusion based on the research conducted is to develop an educational game about carbon emissions through the application of the Game Development Life Cycle (GDLC) method. The use of the GDLC method can also be a reference in the development of other educational games for various fields and educational purposes. With the aim of increasing awareness and understanding of the younger generation about the importance of reducing carbon emissions and their impact on the environment. It is expected that the younger generation will be more interested and motivated to learn about environmental issues and take real action to reduce carbon emissions. Based on the SUS analysis, it can be concluded that this game demonstrates good usability, with relevant feature integration and positive user responses regarding interest, visual satisfaction, and understanding of the game's objectives, question 8 evaluates whether players find the information provided by the game about carbon emission reduction efforts informative and valuable. By receiving positive feedback on the usefulness of this information, the game can effectively fulfill its

objective of educating players about environmental issues and motivating them to take action to reduce carbon emissions. Suggestions that are expected for future researchers can be made such as adding a variety of content, and features and can be made in the form of a mobile version to make it more accessible.

REFERENCES

[1] Z. Borghei and P. Leung, "An Empirical Analysis of the Determinants of Greenhouse Gas Voluntary Disclosure in Australia," Accounting and Finance Research, vol. 2, Nov. 2013, doi: 10.5430/afr.v2n1p110.

[2] Z. Wang, F. Yin, Y. Zhang, and X. Zhang, "An empirical research on the influencing factors of regional CO2 emissions: Evidence from Beijing city, China," Appl Energy, vol. 100, no. C, pp. 277–284, 2012, doi: DOI: 10.1016/j.apenergy.2012.0,.

[3] J. Hou, Y. Hou, Q. Wang, and N. Yue, "Can industrial agglomeration improve energy efficiency? Empirical evidence based on China's energy-intensive industries," Environmental Science and Pollution Research, vol. 29, pp. 1–15, Jun. 2022, doi: 10.1007/s11356-022-21429-x.

[4] J. Hou, T. Teo, F. Zhou, M. K. Lim, and H. Chen, "Does industrial green transformation successfully facilitate a decrease in carbon intensity in China? An environmental regulation perspective," J Clean Prod, vol. 184, Mar. 2018, doi: 10.1016/j.jclepro.2018.02.311.

[5] P. Liu, "Pricing policies and coordination of low-carbon supply chain considering targeted advertisement and carbon emission reduction costs in the big data environment," J Clean Prod, vol. 210, pp. 343–357, 2019, doi: https://doi.org/10.1016/j.jclepro.2018.10.328.

[6] Atlas team, "Global Carbon Atlas." Accessed: Jun. 22, 2023. [Online]. Available: https://globalcarbonatlas.org/emissions/carbon-emissions/

[7] Z. Du and B. Lin, "Analysis of carbon emissions reduction of China's metallurgical industry," J Clean Prod, vol. 176, pp. 1177–1184, 2018, doi: https://doi.org/10.1016/j.jclepro.2017.11.178.

[8] O. Edenhofer et al., Climate Change 2014: Mitigation. Technical Summary. 2014.

[9] F. Nugraha, A. Permanasari, and I. D. Pursitasari, "Disparitas Literasi Lingkungan Siswa Sekolah Dasar di Kota Bogor," Jurnal IPA & Pembelajaran IPA, vol. 5, no. 1, pp. 15–35, Mar. 2021, doi: 10.24815/jipi.v5i1.17744.

[10] L. Sun, X. Cao, M. Alharthi, J. Zhang, F. Taghizadeh-Hesary, and M. Mohsin, "Carbon emission transfer strategies in supply chain with lag time of emission reduction technologies and low-carbon preference of consumers," J Clean Prod, vol. 264, Aug. 2020, doi: 10.1016/j.jclepro.2020.121664.

[11] Peraturan Pemerintah RI, Undang – Undang Republik Indonesia Nomor 40 Tahun 2009 Tentang Kepemudaan. 2009.

[12] P. O. Irianto and L. Y. Febrianti, "Pentingnya Penguasaan Literasi Bagi Generasi Muda Dalam Menghadapi Mea," Jurnal Unissula, pp. 640–645, 2017.

[13] R. Ramadan and Y. Widyani, "Game development life cycle guidelines," in 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2013, pp. 95–100. doi: 10.1109/ICACSIS.2013.6761558.

[14] R. Yanwastika Ariyana, E. Susanti, M. Rizqy Ath-Thaariq, and R. Apriadi, "INSOLOGI: Jurnal Sains dan Teknologi Penerapan Metode Game Devlopment Life Cycle (GDLC) pada Pengembangan Game Motif Batik Khas Yogyakarta," Media Cetak), vol. 1, no. 6, pp. 796–807, 2022, doi: 10.55123/insologi.v1i6.1129.

[15] R. A. Krisdiawan, M. Faza Rohmana, and A. Permana, "Pembuatan Game Runaway From Culik Dengan Algoritma Fuzzy Mamdani," Buffer Informatika, vol. 6, 2020, [Online]. Available: https://journal.uniku.ac.id/index.php/buffer

[16] A. Agung Saputra, F. Nonggala Putra, and R. Darma Rusdian Yusron, "Pembuatan Game Edukasi Pengenalan Kebudayaan Indonesia Menggunakan Metode Game Development Life Cycle (GDLC) Berbasis Android Design an Educational Game Introducing Indonesian Culture Using the Android-Based Game Development Life Cycle (GDLC) Method," 2022.

[17] J. T. Pendidikan, R. Winarni, E. Resnandari, and P. Astuti, "Pengaruh Penggunaan Media Pembelajaran Storyboard Terhadap Kreativitas Belajar Sisiwa Pada Mata Pelajaran Seni Budaya," vol. 4, 2019.

[18] A. A. Pratama, R. Roedavan, and A. P. Kurniawan, "Pembuatan Aplikasi Game Kimia Berbasis Android-Mekanik Sub Game Scramble Chemistry," 2023, pp. 1332–1338.

[19] A. Wahyudinata and H. Dirgantara, "Pengembangan Gim Edukasi 2D Pemilahan Sampah Daur Ulang Berbasis Android," MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 20, no. 1, Sep. 2020, doi: https://doi.org/10.30812/matrik.v20i1.860.

[20] R. A. Krisdiawan, "Implementasi Model Pengembangan Sistem Gdlc Dan Algoritma Linear Congruential Generator Pada Game Puzzle," Jurnal Nuansa Informatika, vol. 12, 2018.

ISSN 2085-4552

Designing a QR Code Attendance System Using BYOD (Bring Your Own Device)

Ahmad Raihan Djamarullah[1], Ilyas Nuryasin[2], Hardianto Wibowo[3]
[1, 2, 3] Department of Informatics, Universitas Muhammadiyah Malang, Malang, Indonesia
[1]raihandj@webmail.umm.ac.id, [2]Ilyas@umm.ac.id, [3]ardi@umm.ac.id

Accepted 5 February 2024
Approved 19 April 2024

***Abstract*—Attendance is an activity of collecting attendance data from each individual who attends events, work, and learning. The current application of attendance in certain companies, schools, or universities is still done manually using paper so it is considered less efficient and effective. Digitizing attendance activities can provide many benefits, such as making managing large amounts of attendance data easier. This is usually used in companies or schools. To reduce additional costs, this can be done by using a personal device as a medium for taking attendance, this can be called BYOD or Bring Your Own Device. The attendance that will be designed will use the user's smartphone or mobile device as a medium for taking attendance by scanning the QR code. The results of tests carried out using black box testing on mobile and web applications, shows that all the features contained in both applications are running according to their function. The use of QR Codes and also the implementation of BYOD can make it easier for users to take attendance. Apart from this, it is also easier for admins to manage user attendance data.**

***Index Terms*— Attendance System; QR Code; Bring Your Own Device; BYOD**.

I. INTRODUCTION

Currently, information technology plays an essential role in society. With the rapid development of technology, many areas of life have become more manageable[1]. With current technology, the information obtained is managed quickly and accurately [2]. Utilizing this technology helps everyone create various kinds of tools and programs that can help make every activity easier so that it becomes more productive [3], [4].

Various activities are currently becoming increasingly complex and dense, so a high level of mobility is required [3]. One example of an activity that can be made easier by using technology is attendance. Attendance is an activity of collecting attendance data from each individual who attends events, work, and learning[5], [6]. Attendance is very important because it is one of the factors used in the assessment aspect of an agency[6], [7]. The current application of attendance in certain companies, schools, or universities is still done manually using paper so it is considered less efficient and effective [8], [9]. Using manual attendance still has several disadvantages, such as the information that has been collected must be managed manually, which can take a lot of time and is also prone to errors in data management[2], [8], [10]. Apart from that, by using manual attendance the process of archiving attendance data can also take a long time and there is a risk that the archive data will be lost[10], [11]. Therefore, the aim of this research is to create a digital attendance system.

Digitizing attendance activities can provide many benefits, such as making managing large amounts of attendance data easier. This is usually used in companies or schools[12], [13]. Implementing a digital attendance system can use several additional tools that can scan cards, faces, and fingerprints [14]. However, implementing this attendance system can be categorized as expensive because it requires costs to purchase additional equipment [14]. To reduce additional costs, this can be done by using a personal device as a medium for taking attendance, this can be called BYOD or Bring Your Own Device [15], [16], [17].

Therefore, the use of technology such as smartphone is very necessary. This is because nowadays almost everyone has a smartphone. In Indonesia alone, in 2018, it is estimated that the number of active smartphone users will be more than 100 million people [18]. Smartphone has features that are very useful in helping users' daily activities [14], [19]. Smartphone are highly portable due to their compact size and low power requirements [15]. Utilizing a digital attendance system based on BYOD can be cost-effective and efficient since users don't need to sign for attendance, reducing the accumulation of attendance data files[20].

Based on the background explanation above, the theme raised is designing a QR Code attendance system with the implementation of BYOD (Bring Your Own Device). The attendance that will be designed will use the user's smartphone or mobile device as a medium for taking attendance by scanning the QR code.

II. RESEARCH METHOD

A. Research Flow

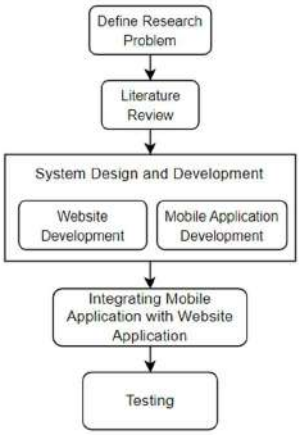


Fig. 1. Research Flow

The flow of the intended research is illustrated in Figure 1. The research will utilize the prototype model of software development to design an attendance system that will be accessible on both websites and mobile devices. The following stages will be involved in the research:

1. Website Development

The website-based attendance system will later be designed using the Laravel framework. The website will be used as a place for admins to manage user data, and attendance data from users and also as a medium for displaying attendance QR codes.

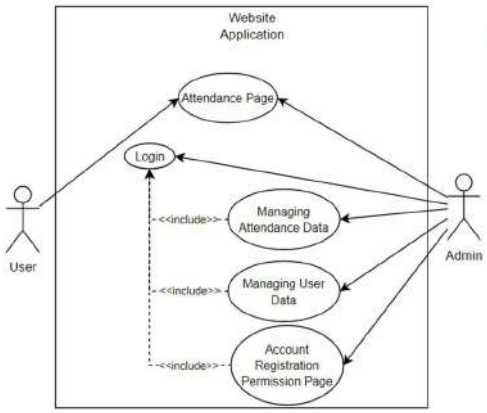


Fig. 2. Website Use Case Diagram

Figure 2 shows a picture of the Website Use Case Diagram. This website consists of several pages. The main page is the attendance page, which is accessible to all users. Additionally, there are two pages for managing attendance data and user data. These pages can only be accessed by admins and are designed for managing user data and attendance information. Finally, there is an account registration permission page, this page was created to prevent the creation of careless accounts.

2. Mobile Application Development

A mobile application will be designed using Flutter, which will have an API to send data to the website, where it will be managed by the admin. The purpose of the application is to allow users to take attendance by scanning the QR code displayed on the website.

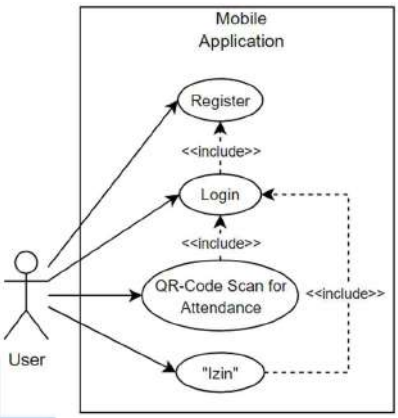


Fig. 3. Mobile Application Use Case Diagram

Figure 3 shows a picture of the Mobile Application Use Case Diagram. Users can interact with various pages, such as login, registration, and attendance. Users can register for an account, but their account will not be active immediately. They need permission from the admin to activate their account. The QR code scanning feature is available on the attendance page for users who wish to register their attendance. In addition, an "Izin" feature is available for users who are unable to attend.

3. Integrating Mobile with Website Application

During the website development, an API will be created and integrated with the mobile application. This API will include functions such as login, attendance, password change, and logout.

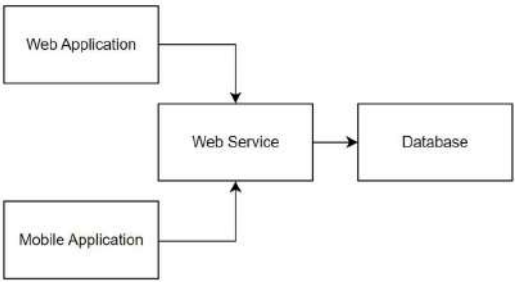


Fig. 4. System overview that has been integrated

Figure 4 shows a picture of the system after integration between the mobile application and website application. Users can mark their presence by scanning the QR code that appears on the website using the mobile application, which will transmit the data to the database via the API. The attendance information will

be displayed on the website, where the admin can access and manage it. The admin has the authority to oversee both attendance and user data for all system users.

4. Testing

After the prototype design and integration of two applications, the system will be tested using the Black Box Testing method to ensure its functionality and stability. The testing will be divided into two stages, the first on the website and the second on the mobile app.

B. Bring Your Own Device (BYOD)

Bring Your Own Device or BYOD is a concept that allows users to use their devices to connect, access data, or complete tasks from systems used in a company, institution, etc. [14], [15], [17]. BYOD is a growing trend since mobile device users started using personal devices to help with their work. According to Jeffrey in [21] Bring Your Own Device or BYOD is a strategy proposed by Malcolm Harkins, Intel's chief security and privacy officer, based on his observations which show that most employees bring personal mobile devices while working. This is why policy proposals take advantage of this trend to reduce costs and increase productivity[21].

Devices used in BYOD implementation can be laptops, tablets, and smartphones, but currently, smartphones are commonly used in BYOD implementation. Smartphones are considered easier to use because they are easy to carry due to their small size and can be connected to the internet network anywhere[16], [21].

BYOD can be applied not only in companies but also in the education sector, such as by helping students learn and develop skills by using these devices[17].

Implementing BYOD in the attendance system that will be designed can provide many benefits for users and companies. The benefits that can be obtained from implementation for companies include saving budget expenses and also not needing to prepare new devices for digital attendance tools. Apart from benefits for the Company, BYOD also provides benefits to users such as increased mobility and productivity because by using personal devices, users can work and access company data or materials wherever they are. Furthermore, it also provides convenience for users because it uses a device that is always used, making it easier to work[13], [15], [21].

III. RESULT AND DISCUSSION

In this section, we will explain the web application and mobile application for the QR code attendance system that will be developed as well as the test results using the black box testing method.

A. Web Application

The attendance system that will be run through the website will be designed using the Laravel framework. The designed website will be used to display QR Codes and manage attendance and account data for each user.

Fig. 5. Attendance Page

Figure 5 shows the main page or page used to display the QR Code which can be scanned by users when taking attendance. The QR Code on the attendance page changes every 5 seconds to prevent cheating. On this page, there is also a table to display a list of users who have missed arrivals and returns on that day.

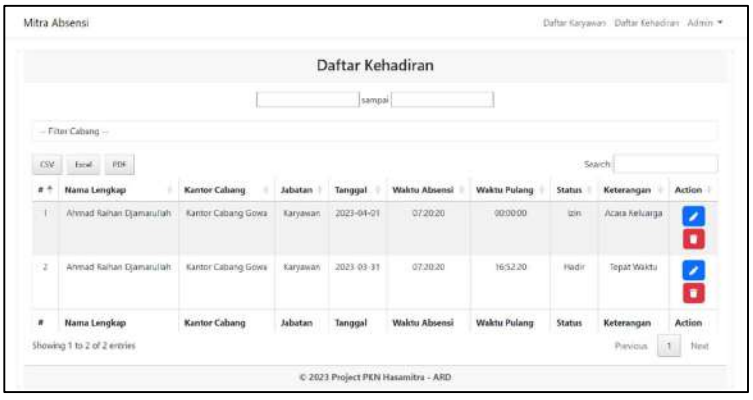


Fig. 6. Attendance Data Management Page

Figure 6 shows page that admins can use to manage attendance lists. The features presented include filters based on company branch, a search column to search for specific data, an attendance filter with a range of dates, months, and years to make the attendance recap process easier, and also a feature to export attendance data into CSV, Excel and PDF file formats for Facilitates the attendance recap process by admin.

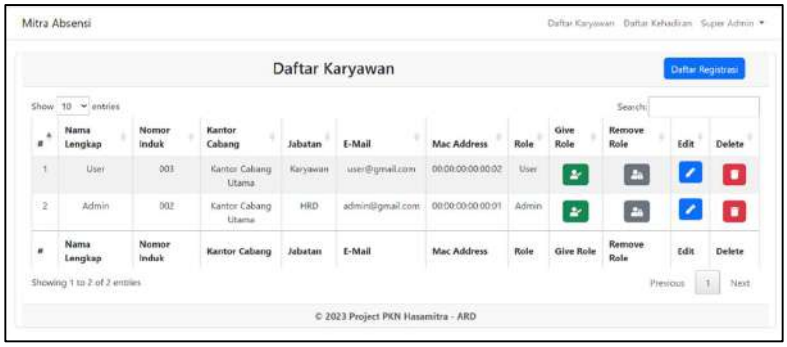


Fig. 7. User Data Management Page

Figure 7 shows page that can be used by admins to manage the user list and approve the user account registration process. Apart from that, admins can also assign and delete Admin roles when there is a change in position in the company.

B. *Mobile Application*

The attendance system that will be run via smartphone will be designed using Flutter and integrated into the website with an API that created from the website. The mobile attendance application will later be used by users who will take attendance into the system.

Fig. 8. Mobile Application Home Page

Figure 8 shows the main page of the mobile application. On this page there are features for absence, changing passwords, permissions, and also logging out. The attendance feature can be accessed by pressing the scan QR Code button and users can scan the QR Code on the website provided.

Then the permission feature can be accessed by selecting the permission menu which will then display a dialog alert to enter the type of permission, whether permission or sickness, and also information about absence. Next is the change password feature which users can use to change their old password to a new password. And the last one is the logout feature to delete running tokens and sessions.

C. *Black Box Testing*

At this stage, testing is carried out on the two applications using the black box testing method. This test is carried out to find out whether the features in the application function properly. Testing is divided into 2 parts, namely web application testing and mobile application testing.

In table 1 shows the test results of the web application using black box testing, testing is carried out on all features on the web. The results of this test show that all features are functioning according to their function.

Table 2 shows the test results of the mobile application using black box testing, testing is carried out on all the features contained in the mobile application. The results of this test show that all features are functioning according to their function.

TABLE I. WEB APPLICATION TESTING RESULTS

No	Testing	Validation	Input	Test Result	Status
1.	Login	Verify Username and Password	Username and Password are Correct	Login Success	Valid
			Username and Password are Incorrect	Login Failed	
2.	Updating User Data	Accessing the Web Application and Select the Update User Menu	New User Data	Update Success	Valid
3.	Deleting User Data	Accessing the Web Application and Select the Delete User Menu	Select User	Delete Success	Valid
4.	Giving Roles	Accessing the Web Application and Selecting the Give Admin Role Button	Select User	Role assignment was Successful	Valid
5.	Approve the Registration Request	Access the Web Application and Select the Registration List Menu	Select User	User Registration Successful	Valid
6.	Updating Attendance Data	Accessing the Web Application and Select the Update Attendance Menu	New Attendance Data	Update Success	Valid
7.	Deleting Attendance Data	Accessing the Web Application and Select the Attendance User Menu	Select Attendance Data	Delete Success	Valid
8.	Export Attendance Data	Accessing the Web Application and Selecting the Export Button	Select Attendance Data	Export Successful	Valid

TABLE II. MOBILE APPLICATION TESTING RESULTS

No	Testing	Validation	Input	Test Result	Status
1.	Login	Verify Username and Password	Username and Password are Correct	Login Success	Valid
			Username and Password are Incorrect	Login Failed	
2.	Register	Click on the 'Register' Button	User Data	Register Success	Valid
3.	Attendance with QR Code	Accessing the Mobile Application and Click on the QR Code Button	Scan QR Code in Web Application	Attendance was Successful	Valid
4.	Permission to be Absent	Accessing the Mobile Application and Click on the 'Izin' Button	Type of Permission to be Absent and Description of Absence	Permission to Absence Successful	Valid
5.	Change Password	Accessing the Mobile Application and Click on the 'Ganti Password' Button	Old Password and New Password	Changed Password Successfully	Valid

IV. CONCLUSION

From the research that has been carried out, the results obtained are that the design of the QR Code attendance system is well made. The use of QR Codes and also the implementation of BYOD can make it easier for users to take attendance. Apart from this, it is also easier for admins to manage user attendance data.

Applications are designed using web and mobile platforms. The web application can be used by admins to manage attendance data and also user data who can make attendance in the system. Then the mobile application can be used by users to take attendance or give permission not to attend.

The results of tests carried out using black box testing on mobile and web applications. It shows that all the features contained in both applications are running according to their function.

ACKNOWLEDGMENT

The Author would like to express his gratitude to the Universitas Muhammadiyah Malang and the

lecturers who helped and supported in preparing this research article.

REFERENCES

[1] T. Marlein Tamtelahitu, J. Sambono, and J. E. Unenor, "Designing a Smart Student Attendance System Using QR Code and Geolocation Techniques", "Perancangan Sistem Absensi Pintar Mahasiswa Menggunakan Teknik QR Code dan Geolocation," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 2021. , in press

[2] N. L. Khairina and M. Dedi Irawan, "Application of QR Code in Employee Attendance Application Using Bootstrap", "Penerapan QR Code Pada Aplikasi Absensi Karyawan Menggunakan Bootstrap," *JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING (CoSIE)*, vol. 01, no. 3, p. 2022, 2022. , in press

[3] M. Himyar, Mu. F. Mulya, and J. H. S. RIngo, "Android-based Employee Attendance Application with QR Code Implementation Accompanied by Personal Photo and Location as Case Study Validation: PT.Selindo Alpha", "Aplikasi Absensi Karyawan Berbasis Android Dengan Penerapan QR Code Disertai Foto Diri Dan Lokasi Sebagai Validasi Studi Kasus: PT.Selindo Alpha," *Jurnal Sistem Komputer dan Kecerdasan Buatan (SisKom-KB)*, 2021. , in press

[4] K. Sianturi and H. Wijoyo, "Megara Hotel Pekanbaru Employee Payroll and Attendance Web Based", "Penggajian Dan Absensi Karyawan Megara Hotel Pekanbaru Berbasis Web," *EKONAM: Jurnal Ekonomi Rancang Bangun System Informasi* , 2020. , in press

[5] N. Rubiati and S. Widya Harahap, "Student Attendance Application Using QR Code with PHP Programming Language at SMKIT Zunurain Aqila Zahra in Pelitung", "Aplikasi Absensi Siswa Menggunakan QR Code Dengan Bahasa Pemrograman PHP di SMKIT Zunurain Aqila Zahra di Pelitung," *Jurnal Informatika, Manajemen dan Komputer*, vol. 11, no. 1, 2019. , in press

[6] A. T. Faramita, S. Wiguna, and A. Fuadi, "Implementation of the Multi App V.1.0 Attendance Application Online in the Work Motivation of Islamic Religious Education Teachers at SMA Negeri 1 Wampu", "Implimentasi Aplikasi Absensi Multiapp V.1.0 Secara Online Dalam Motivasi Kerja Guru Pendidikan Agama Islam Di SMA Negeri 1 Wampu," *Khazanah : Journal of Islamic Studies* 2022. , in press

[7] W. Dinasari, A. Budiman, and D. Ayu Megawaty, "Mobile Based Teacher Attendance Management Information System (Case Study: SD Negeri 3 Tangkit Serdang)" , "Sistem Informasi Manajemen Absensi Guru Berbasis Mobile (Studi Kasus : SD Negeri 3 Tangkit Serdang)," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 1, no. 2, pp. 50–57, 2020. , in press

[8] D. Prasetyo, I. Fitri, A. Rubhasy, and U. Nasional, "Web-Based Online Attendance System With QR Code in Real Time Using Vigenere Cipher Algorithm", "Sistem Absensi Online Berbasis Web Dengan QR Code Secara Real Time Menggunakan Algoritma Vigenere Cipher," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 4, no. 1, 2021. , in press

[9] M. Nandi Susila, K. Salam Ruzki, A. Dwi Praba, and E. Wulansari Fridayanthie, ""QR Code Based E-Attendance With Extreme Programming", "E-Absensi Berbasis QR Code Dengan Extreme Programming," *Jurnal Sistem Informasi (JSI) STMIK Antar Bangsa*, 2022. , in press

[10] F. K. Adam, A. F. O. Pasaribu, and A. D. Wahyudi, "Ditlantas Employee Attendance Monitoring Application Using GPS Technology (Case Study: Ditlantas Polda Lampung)", "Aplikasi Monitoring Absensi Karyawan Ditlantas Dengan Penerapan Teknologi GPS (Studi Kasus: Ditlantas Polda Lampung)," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 1–9, Mar. 2023, doi: 10.33365/jatika.v4i1.723. , in press

[11] R. Roosdianto, A. O. Sari, and A. Satriansyah, "Design and Build an Online Employee Attendance Information System Application", "Rancang Bangun Aplikasi Sistem Informasi Absensi Karyawan Online," *INTI Nusa Mandiri*, vol. 15, no. 2, pp. 135–142, Feb. 2021, doi: 10.33480/inti.v15i2.1932. , in press

[12] D. S. Pratomo and C. Budihartanti, "Design and Development of an Employee Attendance Application Using the Mobile-Based QR Code Method at PT Bayarna Teknologi Nusantara", "Rancang Bangun Aplikasi Absensi Karyawan Menggunakan Metode QR Code Berbasis Mobile di PT Bayarna Teknologi Nusantara," *Journal of Information System, Applied, Management, Accounting and Research*, vol. 6, no. 4, pp. 804–814, 2022, doi: 10.52362/jisamar.v6i4.921. , in press

[13] I. Dzikria and A. Rizal, "Design and Build a Restaurant Self-Ordering System Based on Progressive Web Apps", "Rancang Bangun Sistem Pemesanan Mandiri Restoran Berbasis Progressive Web Apps," *Jurnal Sistim Informasi dan Teknologi* , vol. 5, no. 1, 2023, doi: 10.37034/jsisfotek.v5i1.252. , in press

[14] I. G. N. D. Paramartha and I. W. A. Suranata, "Analysis and Design of an Attendance System Using QR Code and BYOD Method", "Analisis dan Perancangan Sistem Absensi Dengan Menggunakan QR Code dan Metode BYOD," *Jurnal Teknologi Informasi dan Komputer*, vol. 6, no. 2, Jan. 2020, doi: 10.36002/jutik.v6i2.1023. , in press

[15] M. Idris, "Selection of Bring Your Own Device (BYOD) Implementation Solutions Based on Security Controls", "Pemilihan Solusi Penerapan Bring Your Own Device (BYOD) Berdasarkan Kontrol Keamanan," *Jurnal Ilmiah MATRIK*, vol. 21, no. 3, 2019. , in press

[16] Y. Mega Puspita and M. Hasanudin, "Mobile Device Management for the Use of Bring Your Own Device (BYOD) as Company Data Security during the Covid-19 Pandemic," *International Journal of Information System & Technology Akreditasi*, vol. 6, no. 158, pp. 528–536, 2022. , in press

[17] A. H. Arif, S. W. Tho, and S. K. Ayop, "Development of a BYOD (Bring Your Own Device) Integrated STEM Learning Module for Physics Education in Matriculation Colleges: A Needs Analysis", "Development of a BYOD (Bring Your Own Device) Integrated STEM Learning Module for Physics Education in Matriculation Colleges: A Needs Analysis," *Practitioner Research*, vol. 3, pp. 171–190, Jul. 2021, doi: 10.32890/pr2021.3.9. , in press

[18] I. Rahmayani, "Kementerian Komunikasi dan Informatika." Accessed: Jan. 29, 2024. [Online]. Available: https://www.kominfo.go.id/content/detail/6095/indonesia-raksasa-teknologi-digital%20asia/0/sorotan_media

[19] U. Rahmalisa, Y. Irawan, and R. Wahyuni, "Android-Based Teacher Attendance Application for Schools with QR Code Security (Case Study: SMP Negeri 4 Batang Gansal)", "Aplikasi Absensi Guru Pada Sekolah Berbasis Android Dengan Keamanan QR Code (Studi Kasus : SMP Negeri 4 Batang Gansal)," *Riau Journal of Computer Science,* 2020. , in press

[20] A. Febriandirza, "Designing an Online Attendance Application Using the Kotlin Programming Language,", "Perancangan Aplikasi Absensi Online Dengan Menggunakan Bahasa Pemrograman Kotlin," *Jurnal Pseudocode*, 2020. , in press

[21] P. Hadi Nugroho and R. Achmad Darajatun, "Design of a Village Development Monitoring Information System Based on Bring Your Own Device", "Perancangan Sistem Informasi Monitoring Pembangunan Desa Berbasis Bring Your Own Device," *METIK JURNAL*, vol. 5, no. 2, pp. 10–18, Dec. 2021, doi: 10.47002/metik.v5i2.286. , in press

ISSN 2085-4552

U-TAPIS Sal-Tik : Typing Error Detection Using Random Forest Algorithm

Bryan Glennardy[1], Marlinda Vasty Overbeek[2], Niknik Mediyawati[3], Samiaji Bintang Nusantara[4], Rudi Sutomo[5]

[1,2] Informatics Study Program, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]bryan.glennardi@student.umn.ac.id, [2]marlinda.vasty@umn.ac.id
[3,4] Digital Journalism Study Program, Universitas Multimedia Nusantara, Tangerang, Indonesia
[3]niknik@umn.ac.id, [4]samiaji.bintang@umn.ac.id
[5] Information System Study Program, Universitas Multimedia Nusantara, Tangerang, Indonesia
[5]rudi.sutomo@umn.ac.id

Accepted 27 March 2024
Approved 6 June 2024

***Abstract*—The result of this study indicate that the development of technology in the field of journalism has grown very rapidly. However, there are still frequent deviations in language usage on online news portal, particularly in terms of spelling and word usage. Spelling mistake in news articles can cause the information to be unclear and ambiguous. Based on these issues, a study was conducted to create a model to detects typos in Bahasa Indonesia. This model was conducted to create model to detect typos in Bahasa Indonesia. This model was created using the Random Forest algorithm. The Random Forest algorithm works by constructing several decision decisions tree and then combining the decisions from each tree, taking the majority vote from the predictions of each tree to produce stable and accurate prediction. The result of this study show that the model achieved an accuracy of 100%. However, it should be noted that this 100% result means that the model is able to detect words that are already present in the dataset. Based on the evaluation results, since the detected words were contained in the dataset, the accuracy reported is 100%. The model successfully detects typos in Tribunnews news articles.**

***Index Terms*—detection; news articles; random forest algorithm; type error**

I. INTRODUCTION

Nowadays, technology has developed very rapidly. Technology comes with the aim to help make it easier for humans to do their daily work. One field that has experience significant technological development is journalism. In the past, news media primarily relied on the print media to disseminate news, requiring people to read newspapers or other printed publications. However, with advancements in technology, news can now be spread online, including through websites. This allows people to access news easily and freely, anytime and anywhere. News itself a report about a recent event or the latest information regarding an event. In other words, news consists of interesting facts or important information conveyed to the public through various media channels [1].

In journalism, language is used to convey precise and accurate information to the public through mass media [2]. One of the functions of language is as a tool for communication so that the use of language, especially Indonesian, must use good and correct spelling based on existing rules. Spelling itself is a procedure for using words, sentences, and punctuation both in oral and written form [3].

Technological developments in the field of journalism have developed very rapidly, but there are still frequent deviations from the language on online news portals. Usually, this can be seen from the aspect of using spelling and words that are not in accordance with the established writing rules and this is not uncommon in online news portals [4]. This can occur due to the speed of the news dissemination itself, which usually causes errors when typing the news and also when it is in the editing process. Spelling errors that occur in the news can cause the information contained in the news to be unclear and ambiguous [5]. In research that has been conducted related to the analysis of language errors on online news portals with a case study of the Suara.com online news portal and research related to word writing and punctuation errors in online news, it is concluded that there are still many typing errors in the news [6] [7].

Based on the existing problems, research was conducted to detect type error on online news portals which in this study will use news from the Tribunnews news portal namely U-Tapis Sal-Tik. This is because there are requests from partners regarding the creation of several modules and one of them is a module for detecting type error. There is a reason why Tribunnews wants to make this module, namely because Tribunnews uploads 3000 to 5000 articles a day and each reporter is required to write 20 articles. With such a large amount of production, the possibility of errors in publishing is great. Language errors in news can reduce the credibility and public trust in the information presented by the media. Tribunnews itself is the number one online news portal in Indonesia

managed by PT Tribun Digital Online. Tribunnews.com has a network that has spread throughout Indonesia called Tribun Network [8].

U-tapis already conducted before in 2020 until 2023 [9-13]. The research that has been done is the detection of misspelled words using the Jaccard similarity algorithm where this research found an accuracy rate of 93.2% [14]. Then there is also research on detecting the use of conjunctions using the cosine similarity algorithm where this research found an accuracy rate of 92.2% [15].

This type error detection research will use the random forest algorithm. The reason why this research uses the random forest algorithm is because random forest is one of the algorithms in ensemble learning that is used to classify large data sets. In addition, there are several advantages of this algorithm, such as having good accuracy results, relatively strong against outliers and noise, simple and easy to parallelize [16]. In several studies that have been conducted related to text classification also show that the random forest algorithm has a fairly high accuracy rate such as sentiment analysis of the Ruangguru application which has an accuracy rate of 97.16% [17] and fake news detection which has an accuracy rate of 84% [18].

The detection model created can detect by learning from new data that has been trained so that the model can increase its knowledge so that it does not detect by matching words or string matching. In addition, the model does not only detect type error, but it can also provide suggestions for word correction from type error that have been previously detected by the model.

By doing this research, it is hoped that it can detect type error properly and correctly so that it can help news writers in checking type error in the news articles written. Then also with the success of this research, it is hoped that the information contained in the news articles written can be conveyed properly to readers.

II. METHODS

A. Type Error

Type error is an error that occurs during the process of typing text and can change the meaning of a word and even the meaning of a sentence [18]. The occurrence of type error can cause information not to be conveyed properly to the reader and can also cause misunderstanding of the information provided to the reader.

B. Online News Portal

Online News Portal is a site or web page that contains various types of news, such as politics, economics, social, culture to entertainment that is hard news and soft news [19].

C. Text Preprocessing

Text Preprocessing is a process that aims to select text data so that it will be more structured [20]. In text preprocessing, there are several steps that need to be done, including the following.

- Case Folding, is a process that aims to convert all characters into lowercase letters and also eliminate characters that do not include letters [21].
- Tokenizing, is a process for breaking sentences into words [22].
- Filtering, is an advanced stage of tokenizing where this stage is used to select important words from the tokenizing results that have been done previously by removing words that cannot be used or can also be called stopwords [23].
- Stemming, the process of changing a word into its base form [23].

D. Decision Tree

Decision tree is a method for classification using a representation of a tree structure where each node represents the attribute, then the branches represent the value of the attribute and the leaves represent the class [24].

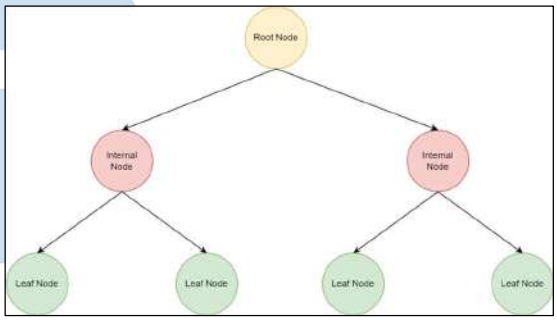


Fig. 1. Decision tree diagram [30]

Fig 1 is a diagram of a decision tree. As can be seen in Figure 1, the nodes in the decision tree are divided into three types, including the following [24].

- Root Node is the node located at the very top where this node has no input and allows it to have no output but it is also possible to have more than one output.
- Internal Node is a branching node. This node has one input and has an output. In the internal node, entropy calculation is performed. Entropy serves to measure the level of uncertainty or impurity of the attribute [25]. After determining the entropy, the calculation of the gain of each attribute is carried out. Gain is a value used to select attributes to generate new nodes. The following is the formula for calculating entropy and gain [25].

$$\boldsymbol{Entropy(S)} = \sum_{i=1}^{n} - \boldsymbol{pi} * \boldsymbol{log_2} * \boldsymbol{pi} \quad (1)$$

Where S is a set of cases, n is a number of S partitions and Pi is a proportion Si to S. And gain formula is

$$\boldsymbol{Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)} \quad (2)$$

- Leaf Node or also known as Terminal Node is the final node where this node only has one input and has an output in the form of a decision or final prediction

E. Ensemble Learning

Ensemble Learning is one of the machine learning paradigms where multiple models (base-models) are trained and combined to get better results[26]. In ensemble learning there are three models, including the following[26].

- Bagging is one method of ensemble learning which uses one type of base model by training in parallel and independently on each base model, then combining them to get the best results. Random Forest algorithm is one of the algorithms included in the bagging model.
- Boosting is one of the methods of ensemble learning which uses one type of base model where training is done sequentially and the results of each base model depend on the results of the previous base model. Adaptative boosting algorithm (AdaBoost) is one of the algorithms included in the boosting model.
- Stacking is one of the methods of ensemble learning in which training uses several base models which are then carried out in parallel and independently on each base model and then uses an algorithm derived from other learning to produce output from the combination of each base model. Blending algorithm is one of the algorithms included in the stacking model

F. Random Forest

Random forest a data mining algorithm method used to classify a dataset [27]. The way this random forest algorithm works is to build several decision trees then combine the decisions of each tree that has been built and take the most votes from the predictions of each tree so that later it will produce stable and accurate predictions [28].

G. Confusion Matrix

Confusion matrix is a table used to measure the performance of machine learning. In the confusion matrix table, there are four variables, including True Positive (TP) is data that is positive and predicted to be true positive by the system, False Positive (FP) is data that is negative but predicted to be positive by the system, False Negative (FN) is data that is positive but predicted to be negative by the system, and True Negative (TN) is data that is negative and predicted to be true negative by the system [29]. Table 1 is a form of confusion matrix table.

TABLE I. CONFUSION MATRIX

Actual Value	Predicted Value	
	1	0
1	TP	FN
0	FP	TN

Confusion matrix is used to calculate accuracy, precision, recall and F1 score. The following is the method used in the confusion matrix to calculate these four things [30].

- Accuracy is a description of how accurate the system that has been created in performing the classification correctly. The following is the formula for how to find the accuracy value.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+FP+FN+TN}} \quad (3)$$

- Precision is a description of the accuracy of the requested data with the prediction results given by the system. The following is the formula for how to find the precision value.

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \quad (4)$$

- Recall is the level of success of a system in finding back information. The following is a formula for how to find the recall value.

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \quad (5)$$

- F1 Score is the average value of precision and recall. Here is the formula for how to find the F1 Score value.

$$F1 = \frac{2*Recall*Precision}{Recall+Precisio} \quad (6)$$

III. RESULT AND DISCUSSIONS

A. Interface Display

The following is the interface of the type error detection model. Figure 2 is the home page of the website.

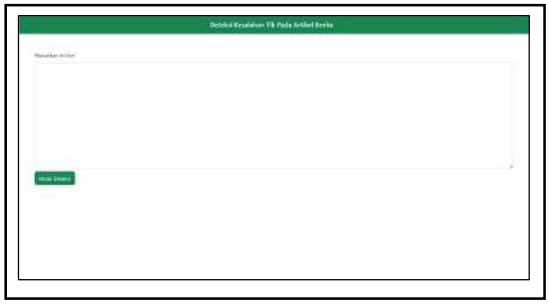


Fig. 2. Home Page Display

Fig 2 is the home page of the website. This page serves as a place for users to enter news articles to be

detected. Users can enter the news article they want to detect through the text area on the home page and then press the *'Mulai Deteksi*' button to start the detection process.

Fig. 3. Results Page Display

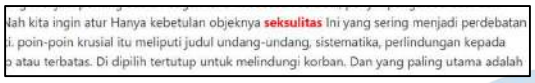


Fig. 4. Display of Detected and Highlighted Type Error Words on the Result Page

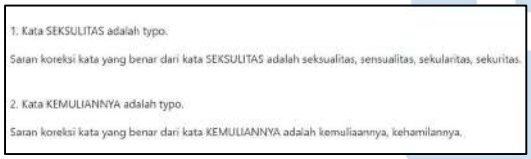


Fig. 5. Display a List of Type Error Words along with Suggestions for Correct Word Correction on the Result Page

Fig 3 is an outline of the results page when the entered news article has been detected. Then in Fig 4 is a display on the results page where the detected type error words will be marked and also in Fig 5 is a display of the list of detected type error words along with suggestions for correct word correction. list of detected mistyped words along with suggestions for correct word correction.

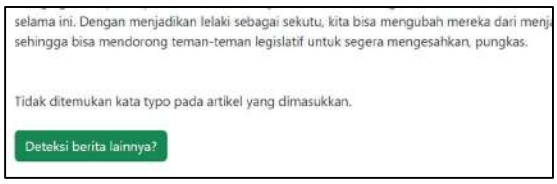


Fig. 6. Display of the Result Page with the Wrong Type Error Word Not Found

Fig 6 is the result page but with different conditions. In Figure 12 no mistyped words are detected so no words are marked and the list of mistyped words is displayed.

B. Testing

At this stage, testing is carried out on the type error detection model that has been created. There are two types of tests carried out, including testing based on the number of news and testing based on the cutoff value.

Testing Based on Number of News

In the first type of test, testing was carried out on the model that had been made with a total of 50 news articles that were entered with different amounts or gradually, namely 20 news to 40 news and to 50 news. Testing by entering the number of news articles in stages aims to see the model's ability to detect tick errors as the number of news increases whether there is a decrease in performance or not as the number of news increases and whether the model can handle larger amounts of data well.

Testing with 20 News

Testing based on the number of news starts with 20 news. Fig 7 is a display of the home page that entered 20 news to be tested.

Fig. 7. Home Page Display When 20 News Articles Are Entered

In the 20 news articles entered, there are 3809 words which is the total word result after passing text preprocessing with 29 type error. After detection, 26 type error were detected. Fig 8 is a display of the total type error detected on the results page of the test with 20 news articles.

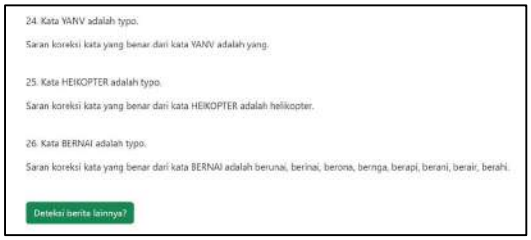


Fig. 8. Display of Test Results With 40 News Articles

Testing with 40 News

The next test was conducted with 40 news articles. Fig 9 is a display of the home page that includes 40 news articles to be tested.

Fig. 9. Home Page Display When 40 News Articles Are Entered

In the 40 news articles entered, there are 6821 words which is the total word result after passing text preprocessing with 61 type error. After detection, there are 48 detected type error. Fig 10 is a display of the total type error detected on the results page of the test with 40 news articles.

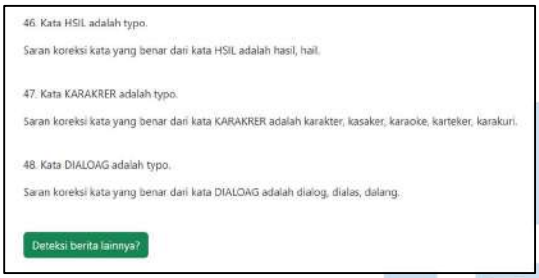


Fig. 10. Display of Test Results With 40 News Articles

Testing with 50 News

The next test was conducted with 50 news articles. Fig 11 is a display of the home page that includes 50 news articles to be tested.

Fig. 11. Home Page Display When 50 News Articles Are Entered

In the 50 news articles entered, there are 8535 words which is the total result of words after passing text preprocessing with 79 mistyped words. After detection, there are 60 detected type error. Fig 12 is a display of the total type error detected on the results page of the test with 50 news articles.

From the test results based on the number of news stories that have been carried out, it can be concluded that the model has successfully performed type error detection. However, there are still some type error words that are not detected. Fig 13 is a comparison graph of testing based on the number of news.

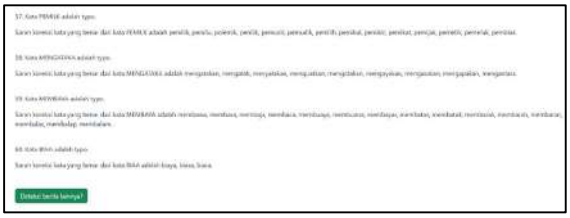


Fig. 12. Display of Test Results With 50 News Articles

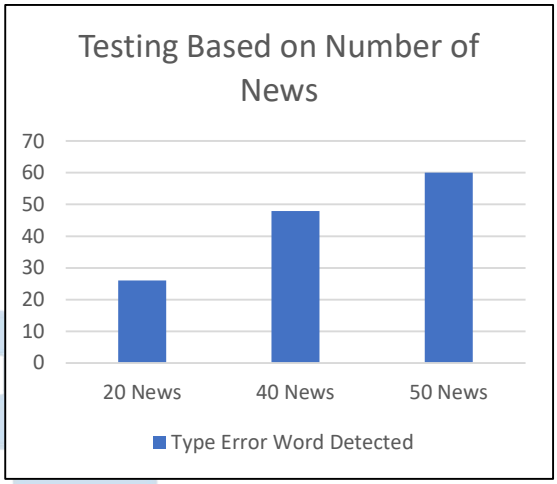


Fig. 13. Comparison Chart of The Test Based on The Number of News

Testing Based on Cutoff Value

In the second type of test, tests were conducted using different cutoff values. This aims to see the word correction generated from mistyped words with different cutoff values.

Testing with Cutoff Value 0.85

The test was conducted by giving a cutoff value of 0.85. Fig 14 is the display when a news article is entered to test with a cutoff of 0.85.

Fig. 14. Home Page View When Tested With a Cutoff Value of 0.85

After testing, in the tested news there is one type error word, namely '*kpeada*' which should be '*kepada*'. However, the result issued is that the word suggestion is not found. Fig 15 is a list of detected mistyped words on the results page with testing using a cutoff value of 0.85.

Fig. 15. Display Page Test Results With a Cutoff Value of 0.85

Testing with Cutoff Value 0.75

Further testing is done by giving a cutoff value of 0.75. The news used is the same as the news used in Fig 15. After testing using a cutoff value of 0.75, the results of word correction suggestions from the type error word '*kpeada*' were successfully obtained. Fig 16 is a list of type error words detected on the results page by testing using a cutoff value of 0.75.

Fig. 16. Display Page Test Results With a Cutoff Value of 0.75

Testing with Cutoff Value 0.65

Furthermore, testing is done by giving a cutoff value of 0.65. Fig17 is the display when the news article is entered for testing with a cutoff value of 0.65. testing with a cutoff value of 0.65.

Fig. 17. Home Page View When Tested With a Cutoff Value of 0.65

After testing, the tested news contained four mistyped words, namely '*seksulitas*', '*kemuliannya*', '*menberi*', and '*kemanusian*' which should be '*seksualitas*', ' *kemuliaannya*', ' *memberi'*, and ' *kemanusiaan*'. These four misspelled words have successfully obtained word correction suggestions. Fig 18 is a list of mistyped words detected on the results page by testing using a cutoff value of 0.65.

Then testing using the same news with a cutoff value of 0.75 can be used as a comparison of the resulting word correction. Fig 19 is a list of mistyped words detected on the results page by testing using a cutoff value of 0.75.

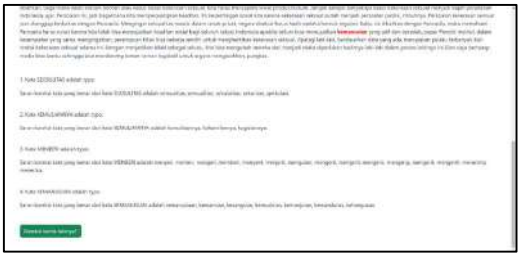


Fig. 18. Display Page Test Results With a Cutoff Value of 0.65

Fig. 19. Display Page The Second News Article Test Result With a Cutoff Value of 0.75

From the results of the tests conducted, the word correction suggestions from the type error words '*seksualitas*', ' *kemuliaannya*' displayed with a cutoff value of 0.75 resulted in a smaller number of correction suggestions than using a cutoff value of 0.65. This shows that the smaller the cutoff value, the more and further the similarity level of the word correction suggestions generated from the detected type error words

C. Evaluation

After conducting the testing stage, the next step is evaluation. In this evaluation stage, calculations are carried out using the confusion matrix. This aims to determine the accuracy, precision, recall, and F1 score of the model. In making the evaluation, the training data is 80% and the test data is 20% of the training data and the accuracy, precision, recall, and F1 score results are 100%. Fig 20 is a classification report from the results of the calculations that have been done.

Classification Report:	precision	recall	f1-score
correct	1.00	1.00	1.00
incorrect	1.00	1.00	1.00
accuracy			1.00
macro avg	1.00	1.00	1.00
weighted avg	1.00	1.00	1.00

Fig. 20. Classification Report Display of Calculations Performed

IV. CONCLUSION

Based on the research that has been done, it can be concluded that the typing error detection model using the random forest algorithm has been successfully built. From the results of the confusion matrix calculation carried out using 80% of the dataset as training data and 20% of the training data as test data at the evaluation stage, the results of accuracy, precision, recall, and F1 score are 100%. This shows that the model built has been able to detect type error, especially in the 50 Tribunnews news articles entered when testing. But keep in mind that this 100% result is that the model is able to detect words that are already contained in the dataset. Based on the evaluation results that have been carried out, because the detected word is contained in the dataset, the accuracy issued is 100%. However, if it is not contained in the dataset, such as one of the letters is mixed up or deleted, it will allow the word to be recognized by the model which is done by majority vote and will provide correction of the correct word from the word by looking at the level of similarity performed by the get close matches function from the difflib library. The reason for using 20% of the training data as data test data is because when calculating the evaluation, the model becomes very dependent on the dataset so that if in the evaluation calculation there is a word that is not in the dataset, the model will not be able to evaluate it. that is not in the dataset, it will cause the model to be unable to perform the evaluation calculation. perform evaluation calculations.

ACKNOWLEDGMENT

The authors would like to thank Universitas Multimedia Nusantara for supporting this research as well as Tribunews as a partner for this research.

REFERENCES

[1] D. N. Chandra, G. Indrawan and I. N. Sukajaya, "Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram," *Jurnal Ilmiah Teknologi dan Informasia ASIA (JITIKA),* vol. 10, no. 1, pp. 11-19, 2016.

[2] Waridah, "Ragam Bahasa Jurnalistik," *Jurnal Simbolika: Research and Learning in Communication Study,* vol. 4, no. 2, pp. 121-129, 2018.

[3] R. Tussolekha, "Kesalahan Penggunaan Ejaan Bahasa Indonesia," *AKSARA Jurnal Bahasa dan Sastra,* vol. 20, no. 1, pp. 35-43, 2019.

[4] N. Faizah and I. S. Ramadhani, "Analisis Kesalahan Berbahasa Pada Penulisan Berita OnlineLiputan6 Edisi 18 Juli 2022," *Jurnal Pendidikan dan Konseling,* vol. 5, no. 1, pp. 850-854, 2023.

[5] F. Achsani, "KESALAHAN BERBAHASA PADA PENULISAN," *SIROK BASTRA,* vol. 8, no. 2, pp. 246-255, 2020.

[6] A. Andriani, D. I. Sari, E. M. Choirunisa, N. P. Ariska and C. Ulya, "ANALISIS KESALAHAN BERBAHASA DALAM TATARAN MORFOLOGI PADA PORTAL BERITA ONLINE SUARA.COM," *Nivedana : Jurnal Komunikasi & Bahasa,* vol. 2, no. 2, pp. 128-139, 2021.

[7] R. R. Apriliana, A. Firdaus and F. Suparman, "KESALAHAN PENULISAN KATA DAN TANDA BACA PADA ONLINE NEWS," *BAHASTRA: Jurnal Pendidikan Bahasa dan Sastra Indonesia,* vol. 5, no. 1, pp. 13-19, 2020.

[8] Marsuki, Rasmila, R. Avindo and D. Safitri, "ANALISIS WEBSITE TRIBUNNEWS MENGGUNAKAN SUS (SYSTEM USABILITY SCALE)," in *Seminar Hasil Penelitian Vokasi (SEMHAVOK)*, Palembang, 2021.

[9] Mediyawati N, Bintang S. Platform Kecerdasan Buatan Sebagai Media Inovatif Untuk Meningkatkan Keterampilan Berkomunikasi: U-Tapis. Pros Semin Nas Pendidik Progr Pascasarj Univ PGRI Palembang 21 Augustus 2021. 2021;69–79

[10] Niknik M, Sutomo R, Nusantara, Samiaji Bintang Overbeek MV. U-Tapis Web: An Automated Indonesian News Text Error Detection System. J Syst Manag Sci. 2024;13(1):1–20

[11] Saputra AKB, Overbeek MV. Harnessing Long Short-Term Memory Algorithm for Enhanced Di-Di Word Error Detection and Correction. In: 1ST INTERNATIONAL CONFERENCE TRACK ON ARTIFICIAL INTELLIGENCE HORIZONS & SOCIETY. 2024

[12] Dwitya NR, Overbeek MV. Development of Detection and Correction of Errors in Spelling and Compound Words Using Long Short-Term Memory. In: 1ST INTERNATIONAL CONFERENCE TRACK ON ARTIFICIAL INTELLIGENCE HORIZONS & SOCIETY. 2024

[13] Siswanto VGA, Overbeek MV. Development of " Kata Terikat " Detection and Writing Errors Correction Using Rabin-Karp and Random Forest Algorithm. In: 1ST INTERNATIONAL CONFERENCE TRACK ON ARTIFICIAL INTELLIGENCE HORIZONS & SOCIETY. 2024

[14] N. Evan, "Deteksi kesalahan eja kata luluh pada berita dengan algoritma jaccard similarity (studi kasus : Tribunnews)," 2022.

[15] J. Olwen, "Deteksi penggunaan kata konjungsi pada portal berita dengan algoritma cosine similarity (studi kasus: Tribun news)," 2022.

[16] E. Fitri, Y. Yuliani, S. Rosyida and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *Jurnal Transformatika,* vol. 18, no. 1, pp. 71-80, 2020.

[17] N. G. Ramadhan, F. D. Adhinata, A. J. T. Segara and D. P. Rakhmadani, "Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic," *JURIKOM (Jurnal Riset Komputer),* vol. 9, no. 2, p. 251−256, 2022.

[18] A. I. Fahma, I. Cholissodin and R. S. Perdana, "Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan Levenshtein Distance," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2, no. 1, pp. 53-62, 2018.

[19] W. H. Kencana, I. V. O. Situmeang, Meisyanti, K. J. Rahmawati and H. Nugroho, "Penggunaan Media Sosial dalam Portal Berita Online," *Jurnal IKRAITH-HUMANIORA,* vol. 6, no. 2, pp. 136-145, 2022.

[20] A. N. Rohman, A. N. Rohman and S. Raharjo, "Deteksi Emosi Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing," *JURNAL EKSPLORA INFORMATIKA,* vol. 9, no. 1, pp. 70-76, 2019.

[21] N. A. Purwitasari and M. Soleh, "Implementasi Algoritma Artificial Neural Network Dalam Pembuatan Chatbot Menggunakan Pendekatan Natural Language Processing," *Jurnal IPTEK,* vol. 6, no. 1, pp. 14-21, 2022.

[22] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *Jurnal RESTI (Rekayasa*

Sistem dan Teknologi Informasi, vol. 1, no. 1, pp. 19-25, 2017.

[23] L. Ardiani, H. Sujaini and Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *JUSTIN (Jurnal Sistem dan Teknologi Informasi),* vol. 8, no. 2, pp. 183-190, 2020.

[24] A. Andriani, "SISTEM PENDUKUNG KEPUTUSAN BERBASIS DECISION TREE DALAM PEMBERIAN BEASISWA STUDI KASUS: AMIK "BSI YOGYAKARTA"," in *Seminar Nasional Teknologi Informasi dan Komunikasi 2013 (SENTIKA 2013)*, Yogyakarta, 2013.

[25] L. S. Muchlis, "PROSES DECISION TREE PADA DATAMINING DENGAN ALGORITMA ID3," *Jurnal Saintek,* vol. 2, no. 1, pp. 87-93, 2010.

[26] L. M. Cendani and A. Wibowo, "Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes," *Jurnal Masyarakat Informatika,* vol. 13, no. 1, pp. 33-43, 2022.

[27] F. Y. Pamuji and V. P. Ramadhan, "Komparasi Algoritma Random Forest Dan Decision Tree Untuk Memprediksi Keberhasilan Immunotheraphy," *Jurnal Teknologi dan Manajemen Informatika,* vol. 7, no. 1, pp. 46-50, 2021.

[28] R. Supriyadi, W. Gata, N. Maulidah and A. Fauzi, "Penerapan Algoritma Random ForestUntuk Menentukan Kualitas Anggur Merah," *JURNAL ILMIAH EKONOMI DAN BISNIS,* vol. 13, no. 2, pp. 67-75, 2020.

[29] R. Siringoringo, "KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR," *Jurnal ISD,* vol. 3, no. 1, pp. 44-49, 2018.

[30] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indonesian Journal of Data and Science,* vol. 1, no. 2, pp. 29-33, 2020.

ISSN 2085-4552

Recommendation System Coffee Shop using AHP and TOPSIS Methods

Christian Andreas Siagian[1], Eunike Endariahna Surbakti[2], Yaman Khaeruzzaman[3]
[1,3] Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]christian.andreas@student.umn.ac.id, [2]eunike.endariahna@umn.ac.id, [3]yaman.khaeruzzaman@umn.ac.id

Accepted 23 April 2024
Approved 26 June 2024

***Abstract*— Indonesian people generally like to spend time with friends, family and business colleagues while drinking coffee. This habit of consuming coffee can not only be done at home, but can also be done in other places such as traditional and modern coffee shops. This has also significantly influenced the growth of coffee shops, especially in Tangerang. So people are faced with so many choices and alternative coffee shops to visit. This research was conducted to create a system that can recommend coffee shops in Tangerang based on priority criteria input by the user. Therefore, this recommendation system uses the Multi Criteria Decision Making (MCDM) method, where the process of making decisions is based on several criteria. This research uses the method Analytical Hierarchy Process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). This research was tested using the Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire and received a very good rating with an overall score of 87.6%, so the conclusion was that the average respondent felt helped by this recommendation system.**

***Index Terms*— AHP; Coffee Shops; Recomendation System; TOPSIS; USE Questionnaire**.

I. INTRODUCTION

Nowadays coffee is no longer considered just a commodity, but has become a lifestyle. The International Coffee Organization (ICO) released data that the amount of coffee consumption in Indonesia increased by 4.04% in the 2021 period to 5 million 60 kg bags from previously 4.81 million 60 kg bags [1]. The increase in the amount of coffee consumption of Indonesian people is also in line with the increase in the number of coffee shops. Toffin released research stating that there was an increase in the number of coffee shops in big cities in Indonesia almost 3 times, from 1000 in 2016 to 2950 in 2019 [2]. Coffee shop growth also occurred significantly in Tangerang in general and South Tangerang in particular. The Tourism Office states that at least 600 coffee shops have been registered [3].

Previous research was carried out by Bambang Hermanto who designed a coffee shop recommendation system in the city of Yogyakarta using the collaborative filtering method [4]. According to Laksana, the collaborative filtering method is more suitable for situations where data is not classified based on specified criteria, because this method obtains recommendation results based on different user preferences and is not limited by the specified criteria [5]. In this way, collaborative filtering does not work with explicit criteria input by the user, so it requires another algorithm that can accept user input in the form of priority criteria. Therefore, this recommendation system uses the Multi Criteria Decision Making (MCDM) method, where the process of making decisions is based on several criteria. Methods that use MCDM include Analythical Hierarchy Process (AHP), Elimination Et Choix TRaduisant la reality (ELECTRE), and Simple Additive Weighting (SAW) [6]. So when comparing the three methods, Fernandes stated that the electre method was easier to implement but could not provide results with high accuracy like the AHP method [7]. Meanwhile, according to Saputra, the AHP method makes it easier to find weighting values compared to the SAW method [8].

Therefore, the AHP method was chosen to identify the weights for each criterion combined with the TOPSIS algorithm. The TOPSIS algorithm was chosen because the selected alternative not only has the closest distance to the positive ideal solution, but also the furthest to the negative ideal solution [9]. TOPSIS calculations are also not complicated, easy to understand, and can determine the value of each alternative with easy calculations [10]. So this research was carried out using the AHP-TOPSIS method, a combination of the two methods was also chosen because AHP has advantages in pairwise comparison matrices and consistency analysis, while TOPSIS is able to make decisions effectively and efficiently, because it is simple in concept, computationally efficient, and has the ability to measure performance. relative to each decision alternative [11]. In this research, AHP was used to weight each criterion, while TOPSIS was used to find the preference value for each coffee shop alternative. Based on interviews with experts, there are 4 criteria used, namely taste, price, service and atmosphere.

The website created must also be ensured to have quality standard so several questionnaire methods were compared, such as System Usability Scale (SUS), End User Computing Satisfaction (EUCS) and Usefulness,

Satisfaction, and Ease of Use (USE) Questionnaire.The SUS method is useful and easy to learn and use products [12]. EUCS method focuses more on user satisfaction, such as accuracy and format [13]. USE Questionnaire can measure various aspects of usability, including usability, user satisfaction, ease of use, and ease of learning, which can provide a more comprehensive understanding of the user experience of the information system [14]. The USE questionnaire also covers the ISO 9241 standard, namely usability is relevant to effective, efficient and user satisfaction measurements [15]. So the USE Questionnaire is used as a method for system evaluation.

II. LITERATURE REVIEW

A. Recommendation System for Coffe Shop

A recommendation system is a program that recommends the most suitable alternative by predicting a user's preference for an alternative based on information relating to the alternative, the user, and the interaction between the alternative and the user [16]. The way the recommendation system works is that user enters input which is then processed using a certain algorithm, and the results are returned to user as a recommendation of a particular alternative based on user preferences [17].

In general, Indonesian people who like to gather spend their time drinking coffee. Apart from being able to drink coffee at home, it can also be done in other places such as coffee shops, both traditional and modern [18]. Coffee shops are places that Indonesian people use to joke around, discuss together or just to soothe tired minds [19].

B. Analytical Hierarchy Process (AHP)

Analytical Hierarchy Process (AHP) is a method that works by weighting each criterion used. The criteria weight values are generated from calculations by comparing each criterion in pairs [20].

AHP has basic principles for solving a problem, namely[21]:

1) *Building a Hierarchy*
   Hierarchies are composed of criteria and alternatives which are fragments of a complex system.
2) *Make pairwise comparisons*
   Pairwise comparisons are made to assess criteria, the comparison scale can be seen in the table I.
3) *Synthesis*
   Several things are done at this stage, namely:
   a) Adds each value in a column in the matrix.
   b) Find the normalized value in the matrix by dividing each value in a column by the sum of all the values in that column.
   c) Adds up each value in each row, then divides by the total elements to produce an average value.

TABLE I. TABLE PAIRWISE COMPARISON SCALE

Scale	Description
1	Criterion X has the same effect as criterion Y
3	Criterion X is slightly more important than criterion Y
5	Criterion X is more important than criterion Y
7	Criterion X is clearly more important than criterion Y
9	Criterion X is absolutely more important than criterion Y
2, 4, 6, 8	For two adjacent values

4) *Measuring Consistency*
   The consistency value is measured by carrying out the following steps:
   a) Multiplies the value in the first column by the priority value of the first element, and continues until the last element.
   b) Sums each row and then divides by the priority value of that element.
   c) Add up the quotients in the previous point, then divide by the number of elements to get the value λ_{max}.
   d) Find the consistency index (CI) value based on the formula:

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

   Where n is the size of the matrix.
   e) Calculate the consistency ratio (CR) value based on the formula:

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

   With IR is Index Random Consistency
   f) Checking consistency in the hierarchy
   If the consistency ratio (CR) value obtained is less than 0.1 then the results of the calculation can be declared consistent and the weight values can be used [22].

C. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

The TOPSIS method is used to determine the available alternatives, where the selected alternative must have the shortest distance from the positive ideal solution and the farthest from the negative ideal solution [9]. The solution algorithm used in the TOPSIS method is [23]:

a) Create a normalized decision matrix using the equation below:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \quad (3)$$

With i=1,2,3,...,m and j=1,2,3,...,n.

b) Constructing a weighted normalized matrix The positive ideal solution (A+) and also the negative ideal solution (A−) are obtained based on the normalized weight value (yij) as in the formula below:

$$y_{ij} = w_i r_{ij} \quad (4)$$

With w=eigenvector; i=1,2,3,...,m and j=1,2,3,...,n.

c) Determining positive and negative ideal solutions The positive (A+) and negative (A−) ideal solution matrices are obtained based on the following equation:

$$A^+ = (y_1^+, y_2^+, y_3^+, .., y_n^+) \quad (5)$$

$$A^- = (y_1^-, y_2^-, y_3^-, .., y_n^-) \quad (6)$$

d) Find the distance from each decision alternative to the positive ideal solution and negative ideal solution. Calculation of the distance from alternative Ai to the positive ideal solution is carried out using the following formula:

$$D_i^+ = \sqrt{\sum_{j=1}^{n} (y_i^+ - y_{ij})^2} \quad (7)$$

With i = 1,2,3,..,m.
Calculation of the distance from alternative A_i to the negative ideal solution is carried out using the following formula:

$$D_i^- = \sqrt{\sum_{j=1}^{n} (y_{ij} - y_i^-)^2} \quad (8)$$

With i = 1,2,...,m.

e) Find the preference value for each alternative To determine the preference value for each alternative (V_i) can be seen in the following formula:

$$V_i = \frac{D_i^-}{D_i^- + D_i^+} \quad (9)$$

With i = 1,2,3,..,m.

III. RESEARCH METHODOLOGY

A. Requirement and Design

Coffee shop data needs from the pergikuliner website which contains names, pictures, list food and drink menus, locations and ratings. The assessment is in the form of a rating of taste, price, service and atmosphere. 50 coffee shop data were taken. This stage is the stage where all the results of the analysis and discussion of system specifications are applied into a system design.

1) *Flowchart Home User*: Flowchart Home User is on the home page/home that user sees when opening the Tangerang Coffee website.

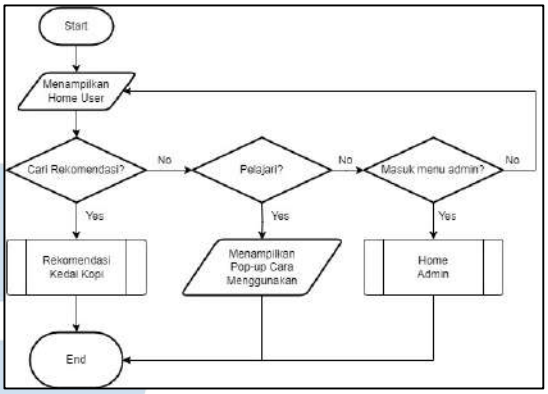


Fig 1. *Flowchart* Home User

2) *Flowchart Home Admin*: Flowchart Home Admin is on the home page which is seen when the admin successfully login using his account. On this page a list of coffee shops is displayed, search bar, search button, button delete search bar. 2.

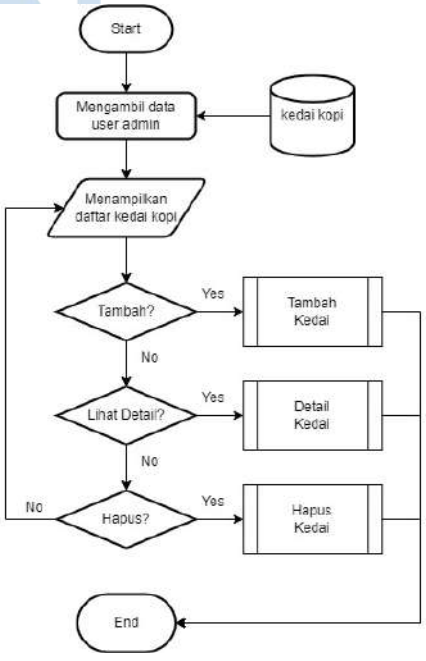


Fig 2. *Flowchart* Home Admin

IV. Results and Discussion

A. Interface Display

1) *Home User*: The image 3 is the result of implementing the display on the user's home page based on the mockup design that has been created. This page is the first page the user sees when opening the website. Displays an admin button to enter the admin menu, a start button to enter the criteria preferences menu and a learn button to display a pop-up on how to use the website.

Fig 3. Home User

2) *Criteria Preference*: The image 4 is the result of implementing the display on the criteria preference page based on the mockup design that has been created. Displays a comparison of each criterion for users to input values based on their personal preferences and a search button to process user input.

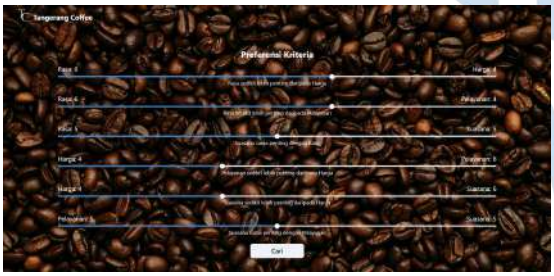


Fig 4. Criteria Preferences

B. AHP Method Calculation

1) *Creating a Criteria Pairwise Comparison Matrix:* The table II is a pairwise comparison of criteria entered by the user. Where each existing criterion is compared one by one. with calculation: Comparison of K1 with K2 = K1 − K2 + 1 = 3

TABLE II. Criteria Pairwise Comparison

Kode	Kriteria	Nilai	Nilai	Kriteria	Kode
K1	Rasa	6	4	Harga	K2
K1	Rasa	6	4	Pelayanan	K3
K1	Rasa	5	5	Suasana	K4
K2	Harga	4	6	Pelayanan	K3
K2	Harga	4	6	Suasana	K4
K3	Pelayanan	5	5	Suasana	K4

Because K1 > K2, then the ratio K1/K2 = 3/1 and the ratio K2/K1=1/3 or 0.333. The process is continued until all criteria are compared. If all criteria have been compared, then the pairwise comparison matrix of criteria has been successfully formed as in Table III.

TABLE III. Criteria Pairwise Comparison Matrix

	K1	K2	K3	K4
K1	1	3	3	1
K2	0,333	1	0,333	0,333
K3	0,333	3	1	1
K4	1	3	1	1
Total	2,667	10	5,333	3,333

2) *Normalization of Criteria Pairwise Comparison Matrix:* The next step is to normalize the pairwise comparison matrix of criteria. The normalization process is carried out by:
r_{11} = 1/2.667 = 0.375
r_{21} = 0.333/2.667 = 0.125
r_{31} = 0.333/2.667 = 0.125
r_{41} = 1/2.667 = 0.375
This process is continued until all rows are normalized, so that the results are as in table IV

3) *Determining the Weight of Each Criteria:* The next step is to determine the weight or eigenvector for each criterion in the following way:

$$\text{eigenVector}_1 = \frac{0.375 + 0.3 + 0.563 + 0.3}{4} = 0.384$$

$$\text{eigenVector}_2 = \frac{0.125 + 1 + 0.063 + 0.1}{4} = 0.097$$

$$\text{eigenVector}_3 = \frac{0.125 + 0.3 + 0.188 + 0.3}{4} = 0.228$$

$$\text{eigenVector}_4 = \frac{0.375 + 0.3 + 0.188 + 0.3}{4} = 0.291$$

So the eigenvector values can be seen in the table V

TABLE IV. Normalization of Criteria Pairwise Comparison Matrix

	K1	K2	K3	K4
K1	1	3	3	1
K2	0,375	0,3	0,563	0,3
K3	0,125	1	0,063	0,1
K4	0,375	0,3	0,188	0,3
Total	1	1	1	1

TABLE V. Determining the Weight of Each Criteria

Alternative	Total	Eigen Vector
K1	1,538	0,384
K2	0,388	0,097
K3	0,913	0,228
K4	1,163	0,291

4) *Measuring Consistency:* Measuring the consistency value begins by finding the λMax value by adding up all the multiplication results between each value in the eigen vector with the total value of each criterion based on the pairwise comparison

matrix of criteria. The λMax value that has been obtained is then reduced by the number of existing criteria and the results of this reduction are divided by the difference between the number of existing criteria and 1 to get the consistency index (CI) value. The final step in measuring consistency is to divide the CI value by the random index value for a matrix of size 4 to get the consistency ratio value. All these calculations are explained in the following calculations:

$$\lambda Max = (0.384 * 2.667) + (0.097 * 10) + (0.228 * 5.333) + (0.291 * 3.333) = 4.179$$

$$CI = \frac{4.179 - 4}{4 - 1} = \frac{0.179}{3} = 0.060$$

$$CR = \frac{0.060}{0.9} = 0.066$$

CR ≤ 0.1 (consistent)

If the CR value obtained is considered consistent, then the calculation process can be continued using the TOPSIS method.

C. TOPSIS Method Calculation

1) *Creating a Decision Matrix:* The first step in TOPSIS is to create a decision matrix obtained from alternative coffee shops and the weight value of each predetermined criterion. The decision matrix can be seen in Table VI. The table VI displays 4 alternatives as a representation of the 50 alternatives in the database, for calculations the method still uses 50 alternatives.

TABLE VI. DECISION MATRIX

Alternative	K1	K2	K3	K4
A1	4,2	3,7	3,9	4,1
A2	3,7	3,7	3,8	4,2
A3	4,3	3,9	4,4	4,6
A4	4,4	4	4,2	4,4

2) *Decision Matrix Normalization:* The decision matrix is then normalized by dividing each value by the square root of the sum of all values in that column squared. The decision matrix normalization process is described in the following calculations:

$$r_{11} = \frac{4,2}{\sqrt{4,2^2 + 3,7^2 + 4,3^2 + 4,4^2 + ..n^2}} = 0.056$$

$$r_{21} = \frac{3,7}{\sqrt{4,2^2 + 3,7^2 + 4,3^2 + 4,4^2 + ..n^2}} = 0.049$$

$$r_{31} = \frac{4,3}{\sqrt{4,2^2 + 3,7^2 + 4,3^2 + 4,4^2 + ..n^2}} = 0.057$$

$$r_{41} = \frac{4,4}{\sqrt{4,2^2 + 3,7^2 + 4,3^2 + 4,4^2 + ..n^2}} = 0.059$$

The calculation is continued for each criterion/column in the matrix. The results of the Decision Matrix Normalization can be seen in Table VII. Table VII only displays 4 alternatives out of 50 alternatives in the database, for calculations the method still uses 50 alternatives.

TABLE VII. DECISION MATRIX NORMALIZATION

Alternative	K1	K2	K3	K4
A1	0.145	0.136	0.134	0.138
A2	0.128	0.136	0.130	0.142
A3	0.149	0.143	0.151	0.155
A4	0.152	0.147	0.144	0.149

3) *Normalization of Weighted Decision Matrix:* The next process is Normalization of the Weighted Decision Matrix. This is done by multiplying each criterion value of all alternatives in the normalized decision matrix by the eigenvector of that criterion which has been obtained from the AHP process as follows:

$$y_{11} = 0.145 * 0.384 = 0.056$$
$$y_{21} = 0.128 * 0.384 = 0.049$$
$$y_{31} = 0.149 * 0.384 = 0.057$$
$$y_{41} = 0.152 * 0.384 = 0.059$$

The calculation is continued for each criterion / column in the matrix so that the normalization results of the weighted decision matrix are obtained as in Table VIII. Table VIII only displays 4 alternatives out of 50 alternatives in the database, for calculations the method still uses 50 alternatives.

TABLE VIII. NORMALIZATION OF WEIGHTED DECISION MATRIX

Alternative	K1	K2	K3	K4
A1	0.056	0.013	0.031	0.040
A2	0.049	0.013	0.030	0.041
A3	0.057	0.014	0.034	0.045
A4	0.059	0.014	0.033	0.045

4) *Searching for Positive and Negative Ideal Solutions:* After normalizing the weighted decision matrix, the next step is to look for positive and negative ideal solutions. The positive ideal solution is obtained from the largest value for each criterion from all alternatives, while the negative ideal solution is obtained from the smallest value for each criterion from all alternatives. The process is as follows:

$$A_1^+ = max(0.056,0.049,0.057,0.059, n5, n6, \dots, n50) = 0.061$$
$$A_1^- = min(0.056,0.049,0.057,0.059, n5, n6, \dots, n50) = 0.048$$

n50 in this process is intended as the 50th data because calculations are carried out on 50 alternative data. This process is continued for each criterion in the matrix. The results can be seen in Table IX.

TABLE IX. POSITIVE AND NEGATIVE IDEAL SOLUTIONS

Ideal Solution	K1	K2	K3	K4
A+	0.061	0.016	0.036	0.046
A-	0.048	0.011	0.027	0.035

5) *Finding the Distance between Positive and Negative Ideal Solutions:* Finding the distance to a positive ideal solution is done by adding up the difference between the criteria values for the alternative and the positive ideal solution squared, the results are then rooted. Meanwhile, the negative ideal solution is obtained by adding up the difference between the criteria values for the alternative and the negative ideal solution squared, then the results are then rooted. The calculation process can be seen as follows:

$$D_1^+ = \sqrt{(0.056-0.061)^2+(0.013-0.016)^2} + (0.031-0.036)^2+(0.040-0.046)^2$$
$$= \sqrt{0.000025+0.000009+0.000025+0.000036}$$
$$= \sqrt{0.000095}$$
$$= 0.010$$

$$D_1^- = \sqrt{(0.056-0.048)^2+(0.013-0.011)^2} + (0.031-0.027)^2+(0.040-0.035)^2$$
$$= \sqrt{0.000064+0.000004+0.000016+0.000025}$$
$$= \sqrt{0.000109}$$
$$= 0.010$$

The calculation is continued for each alternative, so that the distance results for positive and negative ideal solutions are obtained in Table X. Table X shows 4 alternatives out of 50 alternatives that were calculated.

TABLE X. DISTANCE OF POSITIVE AND NEGATIVE IDEAL SOLUTIONS

Alternative	D+	D-
A1	0.010	0.010
A2	0.015	0.007
A3	0.005	0.016
A4	0.005	0.015

6) *Searching for Preference Values:* The preference value is obtained from the distance to the negative ideal solution divided by the sum of the distances to the positive and negative ideal solutions. The preference values are then sorted from highest to lowest. The calculation process is as follows:

$$V_1 = \frac{0.010}{0.010+0.010} = 0.499$$
$$V_2 = \frac{0.007}{0.007+0.015} = 0.316$$
$$V_3 = \frac{0.016}{0.016+0.005} = 0.754$$
$$V_4 = \frac{0.015}{0.015+0.005} = 0.728$$

From this process, the preference value and ranking of each alternative is obtained in Table XI. The table XI shows 4 alternatives out of 50 alternatives in the database.

TABLE XI. PREFERENCE VALUES AND RATINGS

Alternatives	Preferences	Ratings
A1	0.499	29
A2	0.316	48
A3	0.754	1
A4	0.728	2

From Table XI, the data can be sorted from highest to lowest preferences as in Table XII. Based on Table XII it can be concluded that the G8 Coffee & Eatery coffee shop is the coffee shop that best suits user preferences, followed by Kopi Aah in second position and Black Campaign Coffee in third position. Table XII shows 4 alternatives out of 50 alternatives in the database.

TABLE XII. LIST OF COFFEE SHOPS BASED ON ORDER OF PREFERENCE VALUE

Alternatives	Coffee Shops	Preferences
A3	G8 Coffee & Eatery	0.754
A4	Aah Coffee	0.728
A6	Black Campaign Coffee	0.717
A11	Volks Coffee	0.705

The list of coffee shops based on manual calculation preference order in Table XII can be compared with Figure 5 which is the result of system calculations. It can be seen that both lists show accurate store order and preference values.

Fig. 5. Display of System Calculation Results

D. Evaluation System

The USE Questionnaire method which is divided into 4 parts, namely usefulness, ease of use, ease of learning, and satisfaction. The results of the questionnaire were filled in by 31 respondents, overall value are displayed:

TABLE XIII. CALCULATION PERCENTAGE CONVERSION RESULTS

Section	Percentage Calculation Results	Remarks
Usability	86.5%	Very Good
Convenience	87.3%	Very Good
Ease of Learning	88.7%	Very Good
Satisfaction	87.7%	Very Good
Overall	**87.6%**	**Very Good**

V. CONCLUSION

A recommendation system that uses the AHP method for weighting criteria and TOPSIS to find preference values to display recommendations for coffee shops in the Tangerang area according to user preferences and priorities was successfully built. questionnaires distributed for age start from 17-35 years old and have experienced come to coffe shop in Tangerang area. Testing and evaluation of the system was carried out by distributing questionnaires made based on the USE Questionnaire method to 31 respondents with a percentage of usefulness values of 86.5%, ease of use values of 87.3%, ease of learning values of 88.7%, and a satisfaction score of 87.7%, resulting in an overall score percentage of 87.6% with a very good predicate.

REFERENCES

[1] Ali Mahmudan, Berapa Konsumsi Kopi Indonesia pada 2020/2021?, https://dataindonesia.id/agribisniskehutanan/detail/berapa-konsumsi-kopi-indonesia-pada-20202021, 2022.

[2] Denny Adhietya Febrian, Riset Toffin: Tren Minum Kopi Dorong Prospek Bisnis Kedai Kopi Cerah, https://www.idntimes.com/business/economy/dennyadhietya/riset-toffin-tren-minum-kopi-dorong-prospek-bisniskedai-kopi-cerah?page=all, 2019.

[3] Rachman Deniansyah, Ratusan Kedai Kopi Menjamur, Tangsel Bersiap Jadi "Kota Kopi", https://tangerangnews.com/tangsel/read/37965/Ratusan-KedaiKopi-Menjamur-Tangsel-Bersiap-Jadi-Kota-Kopi, 2021.

[4] Bambang Hermanto, Sistem Rekomendasi Kedai Kopi dengan Metode Collaborative Filtering di Kota Yogyakarta Berbasis WEB, 2020: Universitas Islam Indonesia.

[5] Eka Angga Laksana, Collaborative Filtering dan Aplikasinya, Jurnal Ilmiah Teknologi Infomasi Terapan, vol. 1, no. 1, 2014.

[6] Rachman Jaya, Eka Fitria, Rizki Ardiansyah, dan lain-lain, Implementasi Multi Criteria Decision Making (MCDM) Pada Agroindustri: Suatu Telaah Literatur, Jurnal Teknologi Industri Pertanian, vol. 30, no. 2, 2020.

[7] Joao M Fernandes, Susana Prozil Rodrigues, Lino A Costa, Comparing AHP and ELECTRE i for prioritizing software requirements, 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 1–8, 2015: IEEE.

[8] M Saputra, O Salim Sitompul, P Sihombing, Comparison AHP and SAW to promotion of head major department SMK Muhammadiyah 04 Medan, Journal of Physics: Conference Series, vol. 1007, no. 1, pp. 012034, 2018: IOP Publishing.

[9] Desi Ratna Sari, Agus Perdana Windarto, Dedy Hartama, Solikhun Solikhun, Sistem pendukung keputusan untuk rekomendasi kelulusan sidang skripsi menggunakan metode AHPTOPSIS, Jurnal Teknologi dan Sistem Komputer, vol. 6, no. 1, pp. 1–6, 2018, publisher: Department of Computer Engineering, Engineering Faculty, Universitas Diponegoro.

[10] Oey Anita, Sistem Pendukung Keputusan Pemilihan Kuliner Lokal di Jakarta Menggunakan Metode AHP dan Topsis Berbasis Web, 2021.

[11] Abdul Ahmad Chamid, Alif Catur Murti, Kombinasi Metode AHP dan TOPSIS pada Sistem Pendukung Keputusan, 2017.

[12] Irma Salamah, Evaluasi usability website polsri dengan menggunakan system usability scale, Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI, vol. 8, no. 3, pp. 176–183, 2019.

[13] Annisa Uswatun Hasanah, Bayu Waspodo, Elsy Rahajeng, Analysis of MyPertamina Application User Satisfaction Using End User Computing Satisfaction Method, Journal of Software Engineering Ampera, vol. 4, no. 1, pp. 13–34, 2023.

[14] Yulmy Satria Mandala Putra, Rinabi Tanamal, Analisis usability menggunakan metode USE Questionnaire pada website Ciputra Enterprise System, 2020: Ikado.

[15] Ayu Ningtiyas, Siti Nurul Faizah, Metty Mustikasari, Irwan Bastian, Pengukuran Usability Sistem Menggunakan Use Questionnaire pada Applikasi Ovo, Jurnal Ilmiah KOMPUTASI, vol. 20, no. 1, pp. 101–108, 2021.

[16] Mohammad Fathurrahman, Dade Nurjanah, Rita Rismala, Sistem Rekomendasi pada Buku dengan Menggunakan Metode Trust-Aware Recommendation, eProceedings of Engineering, vol. 4, no. 3, 2017.

[17] Novianti Indah Putri, Yudi Herdiana, Zen Munawar, dan lainlain, Sistem Rekomendasi Hibrid Pemilihan Mobil Berdasarkan Profil Pengguna dan Profil Barang, TEMATIK, vol. 8, no. 1, pp. 56–68, 2021.

[18] Aldi An Nurfalah, Surti Zahra, Mohamad Bayi Tabrani, Pengaruh Kualitas Produk Dan Harga Terhadap Kepuasan Konsumen: Studi Kasus Kedai Kopi Mustafa85 Di Pandeglang Banten, Jurnal Bina Bangsa Ekonomika, vol. 13, no. 2, pp. 313–318, 2020.

[19] Diyah Widiyanti, Harti Harti, Pengaruh self-actualization dan gaya hidup hangout terhadap keputusan pembelian di kedai kopi kekinian pada generasi milenial Surabaya, Jurnal Manajemen Pemasaran, vol. 15, no. 1, pp. 50–60, 2021.

[20] Arista Qiyamullaily, Silvia Nandasari, Yusuf Amrozi, Perbandingan penggunaan metode SAW dan AHP untuk sistem pendukung keputusan penerimaan karyawan baru, Teknika: Engineering and Sains Journal, vol. 4, no. 1, pp. 7–12, 2020, publisher: Faculty of Engineering, Universitas Maarif Hasyim Latif.

[21] Umu Habibah, Miftahurrahma Rosyda, Sistem Pendukung Keputusan Penerima Bantuan Langsung Tunai Dana Desa di Pekandangan Menggunakan Metode AHP-TOPSIS, Jurnal Media Informatika Budidarma, vol. 6, no. 1, pp. 404–413, 2022.

[22] Istna Maratul Khusna, Novita Mariana, Sistem Pendukung Keputusan Pemilihan Bibit Padi Berkualitas Dengan Metode AHP Dan Topsis, Jurnal Sisfokom (Sistem Informasi Dan Komputer), vol. 10, no. 2, pp. 162–169, 2021.

[23] Putri Alit Widyastuti Santiary, Putu Indah Ciptayani, Ni Gusti Ayu Putu Harry Saptarini, I Ketut Swardika, Sistem Pendukung Keputusan Penentuan Lokasi Wisata dengan Metode Topsis, Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 5, no. 5, pp. 621–628, 2018.

[24] Ari Viyono, Dian Asmarajati, Saifu Rohman, PENGUJIAN USABILITY DALAM USER EXPERIENCE PADA APLIKASI BRIMOLA MENGGUNAKAN USE

QUESTIONNAIRE, Journal of Economic, Business and Engineering (JEBE), vol. 3, no. 2, pp. 321–330, 2022.

[25] Moch Rosid Noviansyah, Wildan Suharso, Didih Rizki Chandranegara, Muhammad Syofi Azmi, Moh Hermawan, Sistem Pendukung Keputusan Pemilihan Laptop Pada E-Commerce Menggunakan Metode Weighted Product, Prosiding SENTRA (Seminar Teknologi dan Rekayasa), no. 5, pp. 43–53, 2019.

[26] Andre Setiawan, Farica Perdana Putri, Implementasi Algoritma Apriori untuk Rekomendasi Kombinasi Produk Penjualan, Ultimatics: Jurnal Teknik Informatika, vol. 12, no. 1, pp. 66–71, 2020.

ISSN 2085-4552

Sentiment Analysis of IMDB Movie Reviews Using Recurrent Neural Network Algorithm

Aryasuta[1], Fenina Adline Twince Tobing[2]
[1,2] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]aryasuta.saputra@student.umn.ac.id, [2]fenina.tobing@umn.ac.id

Accepted 06 June 2024
Approved 02 July 2024

***Abstract*— IMDb is a well-known platform that provides user reviews and ratings of various movies. The number of reviews found on IMDb is quite large, reaching thousands of reviews. Although a movie can have a high overall rating, it is still possible to receive negative reviews from some viewers. Therefore, the purpose of this sentiment classification system is to provide a benchmark for the level of sentiment contained in the movie, and hope that filmmakers can use this information as a reference in the development of their next movie. In this research, reviews from IMDb users are classified into two types, namely positive reviews and negative reviews. The program was created using the Python language with the LSTM (Long Short-Term Memory) classification model of the RNN (Recurrent Neural Network) algorithm. The purpose of using this algorithm is to measure the level of prediction accuracy in the classification process. The results of three test ratios, namely 60:40, 70:30, and 80:20, show that in the scenario of 80% data training and 20% data testing has better performance with the results accuracy of 96%, precision of 97%, recall of 98%, f1-score of 97%.**

***Index Terms*— Sentiment Analysis; IMDb; Python; Recurrent Neural Network**

I. INTRODUCTION

The increasing use of digital platforms is a result of the world's growing population and the changing environment. Digital platforms or more famously known as social media, are very often used to exchange opinions and share experiences about a product or service. People express their emotions directly or indirectly through language, facial expressions, gestures or writing [1]. Finally, these expressions are expressed on one of the platforms for reviewing a movie called IMDb.

Internet Movie Database (IMDb) is a website that provides a collection of information about movies, tv shows, and the cast involved in the movie or show. The majority of IMDb site users are people who want to find some information about movies based on other audience reviews [2]. Viewers who provide reviews about the movie will also provide a rating related to the movie that has been seen. Based on research that ratings and reviews by the audience can have a significant effect on film production [3]. The many forms of reviews that are scattered are sometimes very difficult for humans to distinguish a person's emotions that are actually poured out from text, speech, or facial expressions [4]. Therefore, this research will create a sentiment analysis program in order to find out someone's sentiment on the topic.

Sentiment analysis is a process that uses human language processing and computer language to extract, identify, and classify diverse opinions expressed in text format. It is one of the most important and interesting areas of research, as it can determine the success of a product from reviews and ratings on the internet [5]. Sentiment analysis is basically a classification problem that covers two fields, namely Natural Language Processing (NLP) and Machine Learning (ML). This sentiment analysis system is carried out to see a person's view whether the person's opinion shows a positive or negative side expressed towards a movie, product, and other things [6]. The level of sentiment analysis is divided into three levels, namely document level analysis, sentence level analysis, and entity and aspect level analysis. In reviews of a movie, the level of analysis used is entity and aspect level analysis, where the person's opinion will refer to only two sentiments between positive and negative [7].

There is similar research on sentiment analysis of IMDb movie reviews using the Support Vector Machine (SVM) algorithm from 5000 reviews data getting 79% accuracy, 75% precision, and 87% recall [11].

There is also other research that has been done with the recurrent neural network (RNN) algorithm to analyze the sentiment of traveloka application users as much as 5,000 data and divided by 2,500 testing data and 2,500 training data, getting an accuracy value of 87.42%, and evaluating the performance of the algorithm obtained recall 87.17%, precision 87.53%, and f-measure 87.34% [12].

On the other hand, there are researchers who use the RNN algorithm for sentiment analysis on social media called twitter with the data used as many as 1,500,000 tweets which are divided for testing by 20% and then 80% are used for training and obtain an accuracy of 80.39%, recall 83.57%, precision 78.56% [13].

Based on the background above, from previous research, the RNN algorithm has been well tested for sentiment classification. However, in this study, the algorithm will be tested using the IMDb movie reviews dataset to evaluate whether the results will be comparable to previous research. The tested data will be grouped into two categories, namely positive and negative.

II. THEORETICAL FOUNDATION

A. Sentiment Analysis

Sentiment analysis is an area of machine learning research that focuses on extracting information from textual reviews. The field of sentiment analysis is closely related to natural language programming and text mining. This analysis is used to determine the attitude of a reviewer towards various topics or the review as a whole [14]. Commenting sites have become a popular place to share emotional impressions through short texts. Emotions include happiness, sadness, anxiety, fear, and more. Gathering opinions from movie reviews can be difficult because human language is quite complex, leading to situations where positive words have negative connotations and vice versa [15].

B. Text Mining

Text mining is a method of finding patterns in unstructured text and is done automatically by a computer to find useful information for certain purposes [16]. Through the text mining process, it is possible to see how a person's opinion or opinion on a topic will be classified into two or more classes [17]. The process of this text will require a document preprocessing method, where this process will separate the whole text only to be analyzed in order to facilitate the sentiment classification process. There are several text mining methods that can handle problems, including classification, clustering, information extraction, information retrieval [18].

C. Classification Method

Classification method is a process that aims to develop a model or function that is able to understand and distinguish between concepts or classes that exist in unlabeled data [19]. In data classification, there are two stages that must be carried out, consisting of training using the dataset (training data) and testing using the data to be tested (testing data). The training data used is data that already has a class label. The difference between classification and clustering is that classification requires a data training process and requires data that already has a class label, while clustering does not require a class label because the class label already exists [20].

D. Algoritma Recurrent Neural Network

RNN is one type of neural network family in the deep learning category because the data is processed automatically without defining features [21]. This algorithm is applicable in Natural Language Processing (NLP) in the form of speech recognition, music synthesis, and text. The calculation of hidden state (St) and output (Ot) in the RNN algorithm for the tth step can be formulated as follows [22].

$$S_t = f(U_{xt} + W_{st-1})$$

$$O_t = softmaxV_{st}$$

E. Confusion Matrix

Confusion matrix serves to display the identification results between correctly predicted data sets and incorrectly predicted data, then the results will be compared with the actual facts [23]. The matrix calculation is done based on the true class and predicted class, with the basis as shown below.

		True Class	
		Positif	Negatif
Predicted Class	Positif	TP	FP
	Negatif	FN	TN

Fig 1. Confusion Matrix Classification

- TP: True Positive (number of correct predictions of positive classes)
- TN: True Negative (number of correct predictions of negative classes)
- FN: False Negative (original positive class predicted negative)
- FP: False Positive (original negative class, predicted positive)

1) Accuracy

This matrix is used to indicate the extent to which the model can correctly predict the class. Although this method is widely used, there are drawbacks in the interpretation (notion) of the results, especially when used on unbalanced data. Unbalanced data can lead to incorrect interpretations and needs to be addressed with caution. The matrix value is obtained from the following equation [24]:

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

2) Precision

This indicator is used to measure the accuracy of the positive class prediction results with the true positive class. Although this indicator cannot clearly describe the number of correct predictions of the actual positive class, it gives an idea of how accurate the model is in identifying positive cases. The matrix value is obtained from the following equation [25]:

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

3) Recall

This indicator aims to illustrate the extent to which the predicted positive class is correct with the actual positive class. However, this indicator does not explicitly describe how well the actual positive class prediction results are correct. The matrix value is obtained from the following equation [26]:

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

4) F1-Score

This indicator aims to overcome the weakness in evaluating the performance of positive classes by taking into account precision and recall values in a balanced manner through the calculation of harmonic mean. The previous two indicators that focused only on positive classes meant that f-score could not provide a specific assessment of negative classes. However, all these drawbacks can be overcome by implementing a weighted version to consider all classes and their distributions. The metric value can be calculated using the following equation [27]:

$$f - score = 2 * \frac{precision * recall}{precision + recall} = 2 * \frac{2\,TP}{2\,TP + FP + FN}$$

III. METHODOLOGY AND SYSTEM DESIGN

The flow design of the recurrent neural network algorithm classification system for sentiment analysis of IMDb movie reviews will be described in the form of a flowchart, which consists of the main flowchart, preprocessing flowchart, and classification flowchart.

A. Main Flowchart

The main flowchart of the sentiment analysis system using recurrent neural network is shown in Figure 2. This diagram illustrates several stages performed in the system. The first stage is importing the library needed for this research. Next, import the dataset file in csv format. After that, the preprocessing stage is carried out to manage text data into structured sentences that can be processed by the program. The preprocessing process involves several steps, such as text cleaning, case adjustment, tokenization, removal of stopwords, and stemming. After the preprocessing stage is complete, sentiment labeling is performed on each review, which consists of two labels, namely positive and negative. The sentiment labeling process is done automatically using the Text Blob library. The last stage is classification using the Recurrent Neural Network algorithm with the LSTM (Long short-term memory) model. This model is used to predict the probability of each classification on reviews that have been labeled with sentiment before.

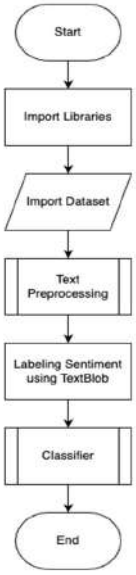


Fig 2. *Main Flowchart*

B. Preprocessing Flowchart

Figure 3 illustrates the preprocessing stage used to convert raw data collected from various sources into more structured information. The purpose of this process is to prepare the data so that it can be processed more efficiently and accurately in the next processing stage. The preprocessing process itself requires several steps, including the following:

1) Cleaning Text

This is a common process that is required during data processing, which is the removal of unnecessary symbols and punctuation marks in text data. This process includes filtering where non-letters or numbers are replaced with spaces. Then it removes excess spaces due to the replacement of special characters or numbers that have become spaces, and removes words that only consist of 1 or 2 characters, for example "the", "a", "is".

2) Case Folding

This is the stage of equalizing the size of letters from the letters "a" to "z", for example at the beginning of a normal word starting with a capital letter, it will be changed to lowercase letters for all words.

3) Tokenizing

Tokenization is the process of breaking sentences into individual word units. So the review sentences will be separated by the comma delimiter ",". An example of the tokenizing process, "Absolutely loved this film" becomes the tokens "Absolutely", "loved", "this", "film".

4) Stop words

This process is a step to remove common words that are considered irrelevant or known as stop words. Examples of these stop words include "the", "a", "an", "in", "of", and so on. This process is intended to improve the effectiveness and quality of text analysis

or modeling. The process of removing stop words is generally done after the tokenizing stage, where words have been separated into tokens which are then removed. This aims to make the words processed in the classification stage more relevant and meaningful [28].

5) Stemming

It is a step in text processing that aims to convert words into their base form or root word. For example, words like "play", "playing", and "played" will be converted into the same base form, namely "play", in the stemming process. This is done to reduce the variety of words that have the same root.

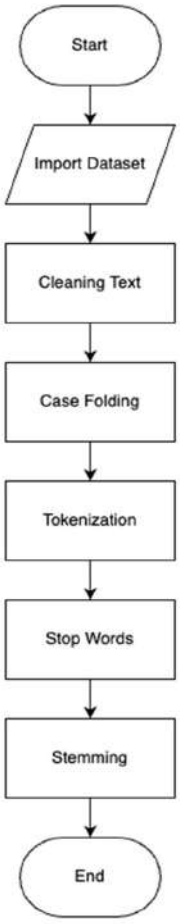


Fig 3. *Flowchart Preprocessing*

C. Classification Flowchart

Figure 4 shows the flowchart for classification using the Recurrent Neural Network algorithm. After the data set is processed at the preprocessing stage, the data is divided into two parts, namely training data and test data. There are three scenarios in separating training data and testing data, namely using a 60:40 ratio, a 70:30 ratio, and an 80:20 ratio. The train data is trained with the LSTM model of the RNN algorithm which will make predictions on the test data and match the X test and y test results to get the accuracy, precision, recall and f1-score values of the confusion matrix table model clustering.

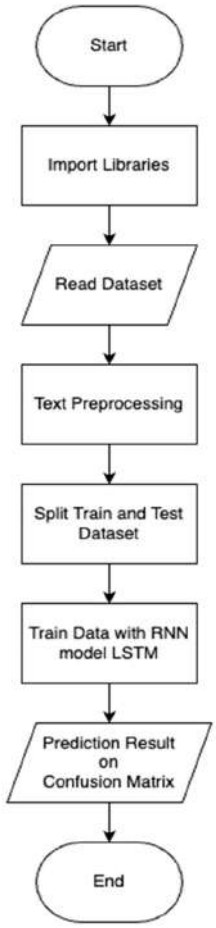


Fig 4. *Flowchart* Klasifikasi

IV. Result and Discussion

A. System Implementation

The implementation process in this research uses the Python programming language version 3.10.8. The data used are movie reviews from various genres and countries that come from an information platform called IMDb. The reviews used are 320,000 reviews data which are divided into two categories, namely positive and negative. The first process in this system will be data preprocessing, which begins with cleaning text.

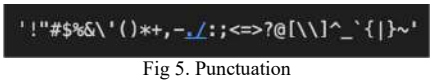


Fig 5. Punctuation

Figure 6 is the result of the cleaning text process created into a new column. The process removes some symbols or special characters as shown in Figure 5. The new column aims to show the results of the program work by displaying the comparison after and before the cleaning process.

	review	label	cleaning
0	"Yaara Sili Sili Virah Ki Raat Ka Jaina""Lekin...	8	yaara sili sili virah raat jalna lekin movie ...
1	Gulzar is at his best when he is telling such ...	9	gulzar his best when telling such intriiguing ...
2	I was completely mesmerized by Lekin and espec...	9	was completely mesmerized lekin and especiall...
3	Greatly enjoyed the development of the story l...	9	greatly enjoyed the development the story line...
4	The lines of time are very blurry. Past, prese...	10	the lines time are very blurry past present an...
5	This is a superb storyline and has excellent m...	10	this superb storyline and has excellent music ...
6	Criticizing is very easy. But when we see the ...	9	criticizing very easy but when see the thing w...
7	Man, where do I start? All three actresses ma...	10	man where start all three actresses make this ...
8	I'm trying to watch this movie and it would no...	2	trying watch this movie and would not play
9	Randsell Pearson's fact-based book provides th...	5	randsell pearson fact based book provides the ...

Fig 6. Cleaning Text

Figure 7 displays the results of the tokenizing process, which is a stage where the sentence will be separated into word units using a delimiter in the form of "," (comma). The sentence that will be used in the tokenizing process is taken from the cleaning column. After performing the process, the results are then entered into a new column named Tokenization.

cleaning	Tokenization
yaara sili sili virah raat jalna lekin movie ...	[, yaara, sili, sili, virah, raat, jalna, leki...
gulzar his best when telling such intriiguing ...	[gulzar, his, best, when, telling, such, intri...
was completely mesmerized lekin and especiall...	[, was, completely, mesmerized, lekin, and, es...
greatly enjoyed the development the story line...	[greatly, enjoyed, the, development, the, stor...
the lines time are very blurry past present an...	[the, lines, time, are, very, blurry, past, pr...
this superb storyline and has excellent music ...	[this, superb, storyline, and, has, excellent,...
criticizing very easy but when see the thing w...	[criticizing, very, easy, but, when, see, the,...
man where start all three actresses make this ...	[man, where, start, all, three, actresses, mak...
trying watch this movie and would not play	[, trying, watch, this, movie, and, would, not...
randsell pearson fact based book provides the ...	[randsell, pearson, fact, based, book, provide...

Fig 7. Tokenization

Figure 8 shows the result of the process of removing irrelevant words, commonly referred to as stopwords. This stopword process utilizes the NLTK (Natural Language Toolkit) library to assist in the removal of conjunctions such as "the", "was", "were", and other similar words. This stage aims to improve the performance of the analysis by removing common words that usually do not provide significant sentence changes. The words to be processed are taken from the Tokenization column. After the process, the results are then entered into a new column to verify the program's success in carrying out the process.

Tokenization	Stopwords
[, yaara, sili, sili, virah, raat, jalna, leki...	[, yaara, sili, sili, virah, raat, jalna, leki...
[gulzar, his, best, when, telling, such, intri...	[gulzar, best, telling, intriiguing, story, ea...
[, was, completely, mesmerized, lekin, and, es...	[, completely, mesmerized, lekin, especially, ...
[greatly, enjoyed, the, development, the, stor...	[greatly, enjoyed, development, story, line, m...
[the, lines, time, are, very, blurry, past, pr...	[lines, time, blurry, past, present, future, m...
[this, superb, storyline, and, has, excellent,...	[superb, storyline, excellent, music, set, bac...
[criticizing, very, easy, but, when, see, the,...	[criticizing, easy, see, thing, clarity, maybe...
[man, where, start, all, three, actresses, mak...	[man, start, three, actresses, make, movie, ge...
[, trying, watch, this, movie, and, would, not...	[, trying, watch, movie, would, play]
[randsell, pearson, fact, based, book, provide...	[randsell, pearson, fact, based, book, provide...

Fig 8. Stopwords

Figure 9 displays the results of the stemming process that has been entered into a new column. The stemming process is a process that will change from colloquial words to basic words, for example "directed: direct", "responsibility: response". The stemming process again uses the NLTK library because this library is one that is quite popular for use in natural language processing provided by Python. The words that will be processed in the stemming stage are taken from the results of the stopwords process.

Stopwords	Stemming
[, yaara, sili, sili, virah, raat, jalna, leki...	[, yaara, sili, sili, virah, raat, jalna, leki...
[gulzar, best, telling, intriiguing, story, ea...	[gulzar, best, tell, intriigu, stori, eas, per...
[, completely, mesmerized, lekin, especially, ...	[, complet, mesmer, lekin, especi, castl, dimp...
[greatly, enjoyed, development, story, line, m...	[greatli, enjoy, develop, stori, line, music, ...
[lines, time, blurry, past, present, future, m...	[line, time, blurri, past, present, futur, mer...
[superb, storyline, excellent, music, set, bac...	[superb, storylin, excel, music, set, backgrou...
[criticizing, easy, see, thing, clarity, maybe...	[critic, easi, see, thing, clariti, mayb, thin...
[man, start, three, actresses, make, movie, ge...	[man, start, three, actress, make, movi, gem, ...
[, trying, watch, movie, would, play]	[, tri, watch, movi, would, play]
[randsell, pearson, fact, based, book, provide...	[randsel, pearson, fact, base, book, provid, b...

Fig 9. Stemming

Figure 10 displays the results of the sentence return process after going through several processing stages. In the first stage, the numbers in the sentence are removed, then punctuation marks or special characters are removed in the second stage, and whitespace is removed in the third stage. Next, the words that have been tidied up are put into a variable named "split" to be made into a sentence like the beginning in the fourth stage. In the last stage, the text is returned after running a series of previous processing operations.

Stemming
yaara sili sili virah raat jalna lekin movi be...
gulzar best tell intriigu stori eas perfect di...
complet mesmer lekin especi castl dimpl haunt ...
greatli enjoy develop stori line music least a...
line time blurri past present futur merg one a...

Fig 10. Sentence Return Process

Figure 11 is the review labeling process. Label processing will use subjectivity and polarity calculations from the TextBlob library. The processing results will be classified into two parts, for positive reviews marked with the number one (1) and negative reviews marked with the number zero (0). The determination of positive and negative comes from the polarity value, where if the value gets a number above 0 it will be categorized as positive, and if the polarity value gets a value below 0 or equal to 0 it will be categorized as negative. The categorization results are entered into a new column called Analysis.

After running all these processes, the last stage in the program will be modeling. This process includes converting tokens to numeric or numbers which will then be able to proceed to the modeling process using the LSTM model of RNN. In this process, the data will be formed into 1000 tokens, after which the 'Stemming' column will create an internal tokenizer dictionary that

will map the words into numeric form. Furthermore, the LSTM(100) code will be used to determine the level of complexity and memory capacity. Sigmoid activation is useful for generating binary values for positive class probability results. This modeling will be run into several experimental scenarios.

Stemming	labelReview	subjectivity	polarity	Analysis
yaara sili sili virah raat jaina lekin movi be...	1	0.371296	0.051852	1
gulzar best tell intriligu stori eas perfect di...	1	0.557143	0.828571	1
complet mesmer lekin especi castl dimpl haunt ...	1	0.483009	0.324784	1
greatli enjoy develop stori line music least a...	1	0.486667	0.170000	1
line time blurri past present futur merg one a...	1	0.350000	0.050000	1
...	...	...	...	...
mike leigh work sever actor appear mani film j...	1	0.430556	0.063889	1
say previou fan movi said yet think mike leigh...	1	0.330000	0.255000	1
anoth one movi love much first time saw cri th...	1	0.294444	0.187778	1
mike leigh treat anoth masterpiec life sweet s...	1	0.550000	0.425000	1
one film subject sublim honestli portray peopl...	1	0.529762	0.161540	1

Fig 11. Labeling

B. Testing

In this section, the test results based on the three predefined scenarios will be discussed. After the test, the scenarios will be evaluated to see which scenario can provide the best classification performance.

1) Test Scenario

In this system, there are three scenarios that are divided based on the ratio comparison. The first scenario is 60:40, where 60% of the data is used for training and 40% of the data is used for testing. The second scenario is 70:30, where 70% of the data is used for training and 30% of the data is used for testing. While the third scenario is 80:20, with 80% of the data used for training and 20% of the data used for testing. These three scenarios are based on previous research in the division of training and testing data in sentiment analysis. While there are other variations in data sharing scenarios that may be used, these three scenarios are common choices and have been tested in the literature. All scenarios will use a dataset consisting of 320,747 total data, with 257,763 positively labeled data and 62,984 negatively labeled data. The training process is conducted using the epochs method, where the training data is shared and learned by the system. After learning, the system applies its learning results to the testing data to test the sentiment prediction capability.

2) Test Results

The system test results are visualized in the form of a confusion matrix table. The following is a display of the test results in the previous process, which will be evaluated to obtain accuracy, precision, recall and f1-score values.

Table 1 is the result of testing conducted on a ratio of 60% training data and 40% testing data formed into a confusion matrix.

Table 2 is the result of testing conducted on a ratio of 70% training data and 30% testing data formed into a confusion matrix.

TABLE I. CONFUSION MATRIX 60:40

Predicted Value	True Value	
	Positive	Negative
Positive	100734	2278
Negative	3203	22084

TABLE II. CONFUSION MATRIX 70:30

Predicted Value	True Value	
	Positive	Negative
Positive	73953	3235
Negative	1366	17617

Table 3 is the result of testing conducted on a ratio of 80% training data and 20% testing data formed into a confusion matrix.

TABLE III. Confusion Matrix 80:20

Predicted Value	True Value	
	Positive	Negative
Positive	50140	1413
Negative	1196	11401

3) Evaluation Results

The evaluation process will be carried out by manual calculation of each confusion matrix table, aiming to obtain and ensure the highest value of accuracy, precision, recall and f1-score. The following is a description of the manual calculation to clarify the results of the values of accuracy, precision, recall, and f1-score in the scenario of 60% training data and 40% testing data.

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

$$= \frac{100734 + 22084}{100734 + 3203 + 2278 + 22084}$$

$$= \frac{122818}{128299}$$

$$= 0.9572794$$

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

$$= \frac{100734}{100734 + 2278}$$

$$= \frac{100734}{103012}$$

$$= 0.97788607$$

$$Recall = \frac{TP}{TP + FP}$$
$$= \frac{100734}{100734 + 3203}$$
$$= \frac{100734}{103937}$$
$$= 0.9691832$$

$$F1 - Score = \frac{2\,TP}{2\,TP + FP + FN}$$

$$= \frac{201468}{201468 + 3203 + 2278}$$

$$= \frac{201468}{206949}$$

$$= 0.97351521$$

The following is a description of manual calculations to clarify the results of the values of accuracy, precision, recall, and f1-score in the scenario of 70% training data and 30% testing data.

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

$$= \frac{73953 + 17671}{73953 + 1366 + 3235 + 17671}$$

$$= \frac{91624}{96225}$$

$$= 0.95218498$$

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

$$= \frac{73953}{73953 + 3235}$$

$$= \frac{73953}{77188}$$

$$= 0.95808934$$

$$Recall = \frac{TP}{TP + FP}$$
$$= \frac{73953}{73953 + 1366}$$
$$= \frac{73953}{75319}$$
$$= 0.98186381$$

$$F1 - Score = \frac{2\,TP}{2\,TP + FP + FN}$$

$$= \frac{147906}{147906 + 1366 + 3235}$$

$$= \frac{147906}{152507}$$

$$= 0.96983089$$

The following is a description of manual calculations to clarify the results of the values of accuracy, precision, recall, and f1-score in the scenario of 80% training data and 20% testing data.

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

$$= \frac{50140 + 11401}{50140 + 1196 + 1413 + 11401}$$

$$= \frac{61541}{64150}$$

$$= 0.95932969$$

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

$$= \frac{50140}{100734 + 1413}$$

$$= \frac{50140}{51553}$$

$$= 0.97259131$$

$$Recall = \frac{TP}{TP + FP}$$
$$= \frac{50140}{50140 + 1196}$$
$$= \frac{50140}{51336}$$
$$= 0.97670251$$

$$F1 - Score = \frac{2\,TP}{2\,TP + FP + FN}$$

$$= \frac{100280}{100280 + 1196 + 1413}$$

$$= \frac{100280}{102889}$$

$$= 0.97464258$$

The following is a table created from the results of the previous manual calculations for each scenario. By using this table, it can easily provide an overview of the performance of each scenario based on the resulting percentage value. The percentage values include accuracy, precision, recall, and f1-score.

TABLE IV. Scenario Comparison of 60:40, 70:30, 80:20 Ratio

Matrix (%)	60:40		70:30		80:20	
	Positive	Negative	Positive	Negative	Positive	Negative
Accuracy	96%		95%		96%	
Precision	98%	87%	96%	93%	97%	91%
Recall	97%	91%	98%	85%	98%	89%
F1-Score	97%	89%	97%	88%	97%	90%

In Table 4, it can be seen that the 60:40 and 80:20 ratios have the same percentage value due to rounding results. However, there is a slight difference between the 60:40 ratio value in the manual accuracy calculation and the 80:20 ratio value in the manual accuracy calculation. The difference shows that in the scenario with the 80:20 ratio, there is a 0.002 higher increase in the accuracy value compared to the 60:40 ratio.

V. CONCLUSION

Based on the test results that have been carried out, it can be concluded that the implementation of the Recurrent Neural Network algorithm with the Long Short Term Memory model in the process of classifying movie reviews from the IMDb site has been successful. Tests were conducted using movie review data from the IMDb site with a total of 320,747 reviews. Tests were carried out using several scenarios that have different ratios, namely 60:40, 70:30, and 80:20. The test results show that the scenario with a ratio of 80% training and 20% testing provides higher performance compared to other scenarios, with an accuracy of 96%, precision of 97%, recall of 98%, and f1-score of 97%.

In addition, this study also shows that the division of training data and testing data has a significant influence on accuracy results. The more data used in the training process, the better the accuracy results. Of the total 320,747 data used, 80.36% had positive sentiments, while 19.63% had negative sentiments. Sentiment analysis on IMDb movie reviews shows that positive sentiment has a higher percentage than negative sentiment.

REFERENCES

[1] E. T. L. Joang Ipmawati, Kusrini, "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," Indonesian Journal on Networking and Security, vol. 6, no. 1, pp. 28–36, 2017.

[2] S. D. A. Y. P. I. A. S. Gita Cahyani, Wiwi Widayani, "Klasifikasi Data Review IMDb Berdasarkan Analisis Sentimen Menggunakan Algoritma Support Vector Machine," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 6, no. 3, 2022.

[3] M. A. F. Faisal Rahutomo, Pramana Yoga Saputra, "IMPLEMENTASI TWITTER SENTIMENT ANALYSIS UNTUK REVIEW FILM MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," JURNAL INFORMATIKA POLINEMA, vol. 4, no. 2, 2018.

[4] J. R. R. A. Kashfia Sailunaz, Manmeet Dhaliwal, "Emotion detection from text and speech: a survey," Social Network Analysis and Mining, vol. 8, no. 28, 2018.

[5] S. Z. M. Md. Rakibul Haque, Salma Akter Lima, "Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews," 2020.

[6] D. Oman Somantri, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," JEPIN (Jurnal Edukasi dan Penelitian Informatika), vol. 5, no. 2, pp. 191–196, 2019.

[7] N. J. Venkateswarlu Bonta, Nandhini Kumaresh, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis ," Asian Journal of Computer Science and Technology, vol. 8, no. S2, pp. 1–6, 2019.

[8] I. Y. B. Rosit Sanusi, Femi Dwi Astuti, "ANALISIS SENTIMEN PADA TWITTER TERHADAP PROGRAM KARTU PRA KERJA DENGAN RECURRENT NEURAL NETWORK," JIKO (Jurnal Informatika dan Komputer), vol. 5, no. 2, pp. 89–99, 2018.

[9] F. K. Muhamad Rizal Firmansyah, Ridwan Ilyas, "Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network," Industrial Research Workshop and National Seminar, pp. 488–495, 2020.

[10] Y. C. O. B. Mohamed Chiny, Marouane Chihab, "LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model," International Journal of Advanced Computer Science and Applications, vol. 12, no. 7, pp. 1097–1099, 2021.

[11] T. I. R. Nur Ghaniaviyanto Ramadhan, "Analysis Sentiment based on IMDB aspects from movie reviews using SVM)," Jurnal dan Penelitian Teknik Informatika, pp. 39–45, 2022.

[12] A. S. Lilis Kurniasari1, "Sentiment Analysis using Recurrent Neural Network," Journal of Physics: Conference Series, pp. 1–6, 2020.

[13] D. M. D. B. O. R. L. N. Sergiu Cosmin Nistor, Mircea Moca, "Building a Twitter Sentiment Analysis System with Recurrent Neural Networks," Journal of Physics: Conference Series, pp. 1–24, 2021.

[14] C. X. M. A. T. M. Zeeshan Shaukat, Abdul Ahad Zulfiqar, "Sentiment analysis on IMDB using lexicon and neural networks," 2020.

[15] S. M. Qaisar, "Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 5, pp. 109–114, 2019.

[16] A. H. Fira Fathonah, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Na¨ıve Bayes," Jurnal Sains dan Informatika, vol. 7, no. 2, pp. 155–164, 2021.

[17] M. A. Nengah Widya Utami, "TEXT MINING DALAM ANALISIS SENTIMEN PEMBELAJARAN DARING DI MASA PANDEMI COVID 19 MENGGUNAKAN

ALGORITMA K-NEAREST NEIGHBOR," JINTEKS (Jurnal Informatika Teknologi dan Sains), vol. 4, no. 2, pp. 140–148, 2022.

[18] A. F. r. O. y. P. a. De di Darwis, Eka Shintya Pratiwi, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," Jurnal Ilmiah Edutic, vol. 7, no. 1, pp. 1–11, 2020.

[19] A. R. Febie Elfaladonna, "ANALISA METODE CLASSIFICATION DECISSION TREE DAN ALGORITMA C.45 UNTUK MEMPREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN APLIKASI RAPID MINER," SINTECH (Science and Information Technology) Journal, vol. 2, pp. 10–17, 2019.

[20] S. M. Qaisar, "Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory," International Conference on Computer and Information Sciences, pp. 1–4, 2020.

[21] A. M. Gonzalo A. Ruz, Pablo A. Henr´ıquez, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers," Industrial Research Workshop and National Seminar, vol. 116, pp. 92–104, 2020.

[22] H. Utami, "Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network," Indonesian Journal of Applied Statistics, vol. 3, no. 1, pp. 31–38, 2022.

[23] F. H. R. Imamah, "Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regresion," Information Technology International Seminar (ITIS), vol. 7, no. 1, pp. 1–17, 2020.

[24] H. S. Gientry Rachma Ditami, Eva Faja Ripanti, "Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace," (JEPIN)Jurnal Edukasi dan Penelitian Informatika, vol. 8, no. 3, pp. 508–516, 2022.

[25] L. S. Merinda Lestandy, Abdurrahim Abdurrahim, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Na¨ıve Bayes," JURNAL RESTI(Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 2, pp. 802–808, 2019.

[26] M. S. Michael Suhendra, Windra Swastika, "ANALISIS SENTIMEN PADA ULASAN APLIKASI VIDEO CONFERENCE MENGGUNAKAN NA¨IVE BAYES," Jurnal Ilmiah Sains Teknologi, vol. 2, no. 1, 2021.

[27] F. H. Hilda Nuraliza1, Oktariani Nurul Pratiwi, "Analisis Sentimen IMBd Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance," Jurnal Mirai Manajemen, vol. 7, no. 1, pp. 1–17, 2022.

[28] M. N. F. M. Ulil Albab, Yohana Karuniawati P., "Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic," Jurnal TRANSFORMATIKA, vol. 20, no. 2, pp. 1–10, 2023.

[29] P. P. K. H. R. Prof. Praveen Gujjar J, "Sentiment Analysis:Textblob For Decision Making," International Journal of Scientific Research Engineering Trends, vol. 7, no. 2, pp. 1097–1099, 2021.

ISSN 2085-4552

Comparison of Fine-tuned CNN Architectures for COVID-19 Infection Diagnosis

Jonathan[1], Moeljono Widjaja[2], Alethea Suryadibrata[3]
Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]jonathan8@student.umn.ac.id, [2] moeljono.widjaja@umn.ac.id, [3]alethea.suryadibrata@lecturer.umn.ac.id

Accepted 2 June 2024
Approved 26 June 2024

***Abstract*— SARS-CoV-2 (COVID-19) virus spread quickly worldwide affects a variety of industries. The government took preventive steps to control the infection, such as diagnosing the human's lung by taking an X-Ray to see if the lungs were infected with COVID-19 or not. Using several pre-trained Convolutional Neural Network models as the basic model, this research deconstructs the comparison of fine-tuned architecture to identify which pre-trained model delivers the best outcomes in diagnosis by applying machine learning. Comparison is conducted using two scenarios that use batch sizes 64 and 32. Accuracy and f1 score are two evaluation metrics used to justify the model's good performance because the images in the real world, especially for positive classes, are scarce. According to the study, EfficientNetB0 outperforms other pre-trained models, namely ResNet50V2 and Xception, which achieved an accuracy of 0.895 and f1 score of 0.8871.**

***Index Terms*— Convolutional Neural Network; COVID-19; Machine Learning; X-Ray**

I. INTRODUCTION

Because of the rapid spread of the SARS-CoV-2 (COVID19) virus worldwide, many people have been contracted, isolated, and removed from the rest of society to prevent human-to-human transmission. Some of the economic consequences, such as economic loss due to flight cancellations (US$245 million) until a significant downfall in hotel occupancy rate (plunged 32.6% compared to June 2020 with June 2019), show how bad the virus affects the world [1]. It also affects the wellbeing of humans itself where the research shows that 41% of respondents feels that their happiness was deteriorated [2]. This caused a pandemic that brought unforeseen crises, including those for creative industry workers [3]. Not only did it affect creative industry workers, but also affected regency politics as explained by Daniel Susilo in his research of fighting the pandemic of COVID-19 in regency [4].Several measures, such as recognizing people exposed to the SARS-CoV-2 virus, were used to control the infection's rapid spread. Detection becomes critical to stop the spread and take preventative measures [5].

According to WHO [6], the primary symptom of the SARSCoV-2 virus, also known as COVID-19, is respiratory problems. Because respiratory issues indicate that the virus has infected the lungs, the X-Ray image will be able to reveal if the virus is present or not [75]. The use of X-Ray scanning machines in hospitals and laboratories can be used to discover this disease as early as possible. On the other hand, the diagnosis procedure is usually done manually by a doctor without technical improvements. Developing a model using a pre-trained model that provides results for detecting whether a person is infected with COVID-19 or not is very promising because it can reduce the time for the doctor to diagnose a patient, which is currently done manually.

Previous studies have shown that Convolutional Neural Network (CNN) is an appropriate method to classify a digital image. In [8], CNN is used to diagnose diabetic retinopathy from fundus image and obtained a model with ROC AUC score of 98%. In [9], CNN is used to classify between bacterial pneumonia and viral pneumonia on chest X-ray dataset.

This research aims to identify the most accurate pre-trained model among ResNet50V2 [10], Xception[11], and EfficientNetB0 [12] for diagnosing COVID-19 infections using the COVIDx CXR-2 chest X-ray dataset. These models consist of a range of architectures, from the old model like ResNet50V2 to the state-of-the-art model like EfficientNetB0. In contrast, many related studies compare older models such as VGG, Xception, ResNet, and Inception. This research also applies some basic fine-tuning by adjusting the batch size from 32 to 64. By determining the best pre-trained model, it can be utilized to train with a larger dataset, ensuring that the model performs well outside of the trained image.

II. RESEARCH METHOD

X-Ray images of human lungs are the subject of investigation in this study. The image will be used to determine whether the lung is infected with COVID-19 or not. Because of its availability and affordable cost, X-Ray is usually the first diagnostic test in patients with suspected or confirmed COVID-19, rather than the RT-PCR test [13]. Figure 1 displays the lungs image of individuals infected with COVID-19 and those who are not infected with COVID-19 for comparison.

(a)

(b)

Fig. 1. Example of a figure caption Human's lung, (a) Positive COVID-19, (b) Negative COVID-19. Source: COVIDx-CXR 2 dataset [10]

Figure 1a shows hazy or cloudy areas around the lungs, which may indicate a COVID-19 infection. In contrast, Figure 1b shows a clear lung X-ray, with no hazy or cloudy areas. This indicate the lung is negative COVID-19.

The dataset used as the study's object was derived from the COVIDx CXR-2 X-Ray lung dataset prepared by Alexander Wong and Linda Wang [14]. Because the dataset's negative and positive samples are unbalanced, they must be pre-processed before being used. Table I shows the properties of the dataset.

TABLE I. DATASET ATTRIBUTES

Data Type	Class Type	Before Preprocessing	After Preprocessing
2*training	Negative	13793	2158
	Positive	2158	2158
2*test	Negative	200	200
	Positive	200	200

To compare the diagnosis result, the researcher trained the model in three pre-trained Convolutional Neural Network (CNN) architectures, namely ResNet50V2, Xception, and EfficientNet B0. The steps taken in each of these architectures are shown in Figure 2.

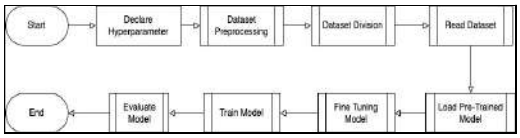


Fig. 2. The Flow of the Research

A. *Pre-Trained Model*

There are various reasons to utilize a pre-trained model, according to Krishna *et al.* [15]. First, training large models on large datasets requires significant computer power. By utilizing the pre-trained model, it can help reduce the computational burden. Second, pre-trained model can lead to faster outcomes when being fine-tuned. It is possible to reduce training time by using pre-trained models and training new models based on pre-trained weight. As a result, using the pre-trained model as a base model, then training the model with datasets and applying some additional layers is a smart move. The following are the pre-trained models that were used in this study:

1) ResNet50

ResNet consists of a "Residual Unit" stack. ResNet itself introduces a feature that can ignore one or more layers called "skip connections" and are the central part of the residual block. This method solves the problem where if the built network gets more profound, the accuracy will stagnate or not develop.

Gunraj *et al.* [14] use this architecture to diagnose COVID-19 infection. In his journal, the model is trained by three steps where at each stage, the machine will be trained with a different dataset. The training process with different datasets resulted in better accuracy and the ability to detect and ignore the noise in the image [16].

The ResNet architecture employed in this work is ResNet50V2, the second-generation one. Using a preactivation layer before being added to the residual block distinguishes this generation from the first one. In comparison to the first generation [17], the addition of this feature produces promising results.

2) Xception

Xception was developed by Francois Chollet, a Google employee, by modifying depth wise separable convolution. Xception is claimed to outperform the Inception V3 architecture, which became its predecessor in the ImageNet dataset. The number of parameters used is the same as Inception V3. The main difference is that Xception could achieve fewer model parameters and still maintain the results [11].

From Figure 3, it can be seen that the pointwise convolution (1 x 1 Convolution) is performed to change the dimension into a new one before the depth wise convolution (n x n Spatial Convolution). Thus, the model can extract more features in one step.

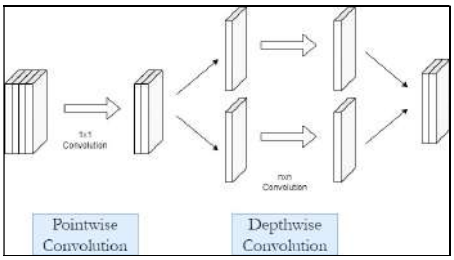


Fig. 3. Modified Depthwise Separable Convolution in Xception

3) EfficientNet

EfficientNet was published by Google in 2019. This model not only increases accuracy but also improves model accuracy by reducing parameters and Floating Point Operations Per Second (FLOPS) as was done in the GPipe architecture where GPipe itself uses Pipeline Parallelism [12].

EfficientNet scales uniformly from the width, depth, and resolution aspects with a fixed coefficient defined as the ϕ symbol. This method is called Compound Scaling. The formula can be written as follows:

$$depth : \delta = \alpha^{\phi} \quad (1)$$

$$width : \omega = \beta^{\phi} \quad (2)$$

$$resolution : \tau = \gamma^{\phi} \quad (3)$$

These fixed coefficients are the coefficients that control the computing resources. For example, if we want to use 2^{ϕ} more computing resources, then we can increase the network depth by α^{ϕ}, the width by β^{ϕ}, and the image size by γ^{ϕ}. The values are constant coefficients determined by tracing the small tile in the original mini model.

In this study, the evaluation metrics that will be used to compare between models are accuracy. Accuracy is the most popular, which usually be the first metric to be calculated in all classification problems. It tells the ratio of accurately classified data items to the total number of observations based on Formula 4 [18].

Using accuracy as the only evaluation criteria is not a good idea, especially when there is a scarcity of data in the actual world and the data is likely to be imbalanced. The nature of accuracy, which implies equal relevance of classes in terms of the number of instances and the level of importance, requires the calculation of the F1 score evaluation metric to account for these concerns [19]. The F1 Score indicates whether the results are biased in favor of the positive or negative class [20].

B. Callback

An overfitting model is not helpful in machine learning research. Overfitting occurs when a learning model is overly focused on the training data, resulting in poor performance when evaluating new data that has not been previously assessed [21]. Several callbacks are implemented at the end of each epoch to avoid constructing the overfitting model, namely:

1) Early Stopping

Early Stopping is a callback that stops the model training process when the current model provides the same outcome as the smaller model [22]. As a result, training time can be reduced while ensuring that the model does not become overfit.

2) Reduce Learning Rate on Plateau

Machine learning models use the learning rate to determine how much the weights can change when the training data is evaluated incorrectly. When the learning rate is too high, the model cannot adjust its weight when new data is provided. As a result, as the model plateaus, it is necessary to reduce the learning rate [23].

C. Confusion Matrix

Confusion matrix is a performance measurement of the classification task for machine learning [24]. True Positive (TP) tells that the prediction result gives a positive value and the actual result is positive. This means that the prediction results are correct. False Positive (FP) tells that the predicted result gives a positive value and the actual result is negative. This means that the prediction result is false. True Negative (TN) tells that the predicted result gives a negative value and the actual result is negative. This means that the prediction results are correct. False Negative (FN) tells that the predicted result gives a negative value and the actual result is positive. This means that the prediction result is false. The confusion matrix itself could be presented in the form of a table with a combination of predicted and actual results as shown in Figure 4.

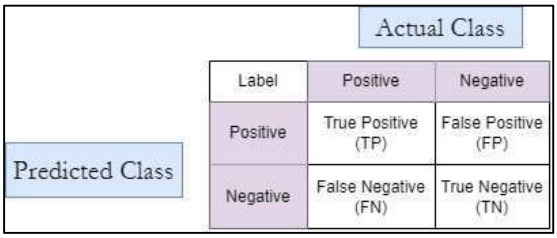


Fig. 4. Confusion Matrix Table

From the value generated by the confusion matrix, the accuracy value can be calculated. The accuracy formula is as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

D. Classification Report

The classification report has four main values that will be displayed when used. Indirectly, this value requires values from *confusion matrix*. The four main values in the classification report are:

1) *Precision*

The precision value describes the model's ability not to label positive images that are negative. This means a value representing the accuracy of all positive predictive results obtained. This value is calculated by the value of True Positive divided by the number of True Positive by False Positive. The formula is as follows:

$$Precision = TP/(TP + FP) \quad (5)$$

2) *Recall*

The recall value describes the model's ability to find all positive images. This means a value representing the accuracy of all positive image results obtained. This value is calculated by the value of True Positive divided by the number of True Positive by False Negative. The formula is as follows:

$$Recall = TP/(TP + FN) \quad (6)$$

3) *F Measure / F1 Score*

The value of the F1 Score is the weighted harmonic mean of precision and recall so that the highest value is one and the smallest is zero. . F1 Score emphasizes a balanced value between precision and recall. The formula is as follows:

$$F1Score = 2 * (Recall * Precision)/(Recall + Precision) \quad (7)$$

4) *Support*

This value is the actual number of class occurrences in the dataset. An unbalanced number between classes will worsen the prediction results of the classification [25].

III. Result and Discussion

This research leverages several libraries to support the data analysis process:

1) Pandas: Used for loading and preprocessing the pre-trained data.

2) Scikit-learn: Used to split the dataset into training and testing data.

3) Keras: Used for image augmentation with ImageDataGenerator and to import the pre-trained model.

4) TensorFlow: Used for fine-tuning the pre-trained model before training.

5) NumPy: Used for preprocessing the prediction results.

6) Matplotlib: Used to create visualizations like the confusion matrix, model accuracy plot, and model loss plot.

7) Seaborn: Used to create a heatmap on top of the visualizations generated by Matplotlib.

The results of the pre-trained models consist of the total time needed to train the model and prediction results based on the confusion matrix, accuracy, and f1 score from each model.

The test scenarios carried out are as follows:

1) ResNet50V2 pre-trained model with a batch size of 64
2) Xception pre-trained model with a batch size of 64
3) EfficientNetB0 pre-trained model with a batch size of 64
4) ResNet50V2 pre-trained model with a batch size of 32
5) Xception pre-trained model with a batch size of 32
6) EfficientNetB0 pre-trained model with a batch size of 32

A. Comparison of Results between Models with Batch Size 64

After doing scenario 1 until 3, results between models with a batch size of 64 by seeing from the perspective of the correctness of the prediction results can be seen in Table II. EfficientNetB0 is more accurate than other models.

TABLE II. Prediction Result of each Model on the Test Dataset based on Confusion Matrix with a Batch Size of 64

	ResNet50 V2	Xception	Efficient NetB0
True Normal	195	185	196
False COVID-19	5	15	4
False Normal	28	26	13
True COVID-19	172	174	187

B. Comparison of Results between Models with Batch Size 32

After doing the scenario 4 to 6, results between models with a batch size of 32 by seeing from the perspective of the correctness of the prediction results

can be seen in Table III. Using batch size 32, EfficientNetB0 also outperformed other models.

TABLE III. PREDICTION RESULT OF EACH MODEL ON THE TEST DATASET BASED ON CONFUSION MATRIX WITH A BATCH SIZE OF 32

	ResNet50 V2	Xception	Efficient NetB0
True Normal	190	187	193
False COVID-19	10	13	7
False Normal	51	79	35
True COVID-19	149	121	165

C. *Evaluation of Comparison Models with Batch Size 64 and 32*

The results of the comparison of the model with batch size of 64 and 32 can be seen in Table IV.

TABLE IV. COMPARISON OF MODEL RESULTS WITH BATCH SIZE 64 AND 32

Model	Batch Size	Accuracy	Training Time	F1 Score
2*ResNet50 V2	64	0.9175	3628s	0.9125
	32	0.8475	2252s	0.83
2*Xception	64	0.8975	7704s	0.8946
	32	0.77	3694s	0.7245
2*EfficientNetB0	64	0.9575	2553s	0.9565
	32	0.895	1923s	0.8871

From all of the evaluation metrics, the pre-trained model EfficientNetB0 with batch size 64 has the best accuracy of 0.9575 and F1 Score of 0.9565. The model with the fastest training time is EfficientNetB0 with a batch size of 32 for 1923 seconds. These results indicate that EfficientNetB0, which scales the width, depth, and resolution of convolutional neural networks with a fixed coefficient, is the best choice for COVID-19 infection diagnosis among ResNet50V2 and Xception.

IV. CONCLUSION

According to the research findings, the fine-tuning model with the best accuracy with a batch size of 64 is EfficientNetB0, which has an accuracy value of 0.9575, a training time of 2553 seconds, and an F1 score of 0.9565. Likewise, the best accuracy for the pre-trained model with a batch size of 32 is EfficientNetB0, which has an accuracy value of 0.895, a training time of 1923 seconds, and an F1 score of 0.8871.

As a result, it can be concluded that the EfficientNetB0 pretrained model with a batch size of 64 is the best application of the CNN algorithm among ResNet50V2, Xception, and EfficientNetB0 for the classification of COVID-19 infections in the COVIDx-CXR 2 dataset. The top outcomes are chosen because accuracy and F1 Score are more critical than training time. In the real-world scenario, training time only affects when the model is trained where accuracy and F1 Score affect image diagnosis at the time model will be used.

ACKNOWLEDGMENT

This research was fully funded by Universitas Multimedia Nusantara as a part of undergraduate thesis work.

V. CONCLUSIONS

In the results of research calculations that have been completed related to Indihome customer sentiment on Indihome services using the Naïve Bayes classification algorithm and Support Vector Machine to get the accuracy value, namely the accuracy of the Support Vector Machine algorithm is greater than the Naïve Bayes classification method. For this reason, in this study using 1000 Indihome customer datasets on the Twitter social media platform, the Support Vector Machine method is a better method than the Naïve Bayes method. Data is collected for 3 months starting from February 2021 to April 2021

However, this research still has several shortcomings, namely the process of labeling positive and negative sentiments is done manually which produces more negative sentiments than positive sentiments. There are differences from the data labeling that is applied manually to test the model using the class prediction results from the model classification results. In addition, this study only uses 1000 datasets. The accuracy of the Naive Bayes method is 82% while the Support Vector Machine is 84%

REFERENCES

[1] A. Japutra and R. Situmorang, "The repercussions and challenges of COVID-19 in the hotel industry: Potential strategies from a case study of Indonesia," International Journal of Hospitality Management, vol. 95, p. 102890, May 2021.

[2] D. Tjahjana, D. Dwidienawati, A. H. Manurung, and D. Gandasari, "Does people's wellbeing get impacted by COVID-19 pandemic measure in Indonesia?" Studies of Applied Economics, vol. 39, no. 4, May 2021. [Online]. Available: https://ojs.ual.es/ojs/index.php/eea/article/view/4873

[3] H. Putranto, "Covid-19 and the Crisis of Creative Industries in Digital Capitalism: Commodification of Digital Media Workers in the Framework of Data as Labor" Jurnal Politik, vol. 25, no. 2, pp. 9-48, April 2021. [Online]. Available: https://ejournal.atmajaya.ac.id/index.php/respons/article/view/2461

[4] E. Hidayat and D. Susilo, "Refusing to Die: Programmatic Goods in the Fight against COVID-19 in Sampang Regency" Respons: Jurnal Etika Sosial, vol. 7, no.1, pp. 47-73, March 2021. [Online]. Available: https://doi.org/10.7454/jp.v7i1.1001

[5] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," Informatics in Medicine Unlocked, vol. 19, p. 100360, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914820302537

[6] WHO, "Coronavirus." [Online]. Available: https://www.who.int/ westernpacific/health-topics/coronavirus

[7] I. Mporas and P. Naronglerdrit, "COVID-19 Identification from Chest X-Rays," in 2020 International Conference on Biomedical Innovations and Applications (BIA). Varna, Bulgaria: IEEE, Sep. 2020, pp. 69–72. [Online]. Available: https://ieeexplore.ieee.org/document/9244509/

[8] Y. Laurensia, J.C. Young, A. Suryadibrata, "Early Detection of Diabetic Retinopathy Cases using Pre-trained EfficientNet and XGBoost," International Journal of Advances in Soft Computing and its Applications, vol. 12, no. 3, pp 101-111, November 2020. [Online]. Available: https://www.i-csrs.org/Volumes/ijasca/2020.3.8.pdf

[9] K. A. Prayogo, A. Suryadibrata, J. C. Young, "Classification of pneumonia from X-ray images using siamese convolutional network," TELKOMNIKA Telecommunication, Computing, Electronics and Control, vol. 18, No. 3, pp 1302-1309, June 2020. [Online]. Available : http://telkomnika.uad.ac.id/index.php/TELKOMNIKA/article/viewFile/14751/8474#:~:text=In%20this%20research%2C%20we%20used,bacterial%20pneumonia%2C%20and%20viral%20pneumonia.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," arXiv:1512.03385 [cs], Dec. 2015, arXiv: 1512.03385. [Online]. Available: http://arxiv.org/abs/1512.03385

[11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," arXiv:1610.02357 [cs], Apr. 2017, arXiv: 1610.02357. [Online]. Available: http://arxiv.org/abs/1610.02357

[12] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," arXiv:1905.11946 [cs, stat], Sep. 2020, arXiv: 1905.11946. [Online]. Available: http://arxiv.org/abs/1905.11946

[13] E. Martinez Chamorro, A. Diez Tascon, L. Ibanez Sanz, S. Ossaba Velez, and S. Borruel Nacenta, "Radiologic diagnosis of patients with COVID-19," Radiologia (English Edition), vol. 63, no. 1, pp. 56–73, Jan. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2173510721000033

[14] H. Gunraj, A. Sabri, D. Koff, and A. Wong, "COVID-Net CT-2: Enhanced Deep Neural Networks for Detection of COVID-19 from Chest CT Images Through Bigger, More Diverse Learning," arXiv:2101.07433 [cs, eess], Jan. 2021, arXiv: 2101.07433. [Online]. Available : http://arxiv.org/abs/2101.07433

[15] S. T. Krishna and H. Kalluri, "Deep learning and transfer learning approaches for image classification," 06 2019. [Online]. Available : https://www.semanticscholar.org/paper/Deep-learning-and-transfer-learning-approaches-for-Krishna-Kalluri/6c2a0e3a798d59655732980d2b03b9e89f586548ss

[16] Y. Liu and S. Ji, "A Multi-Stage Attentive Transfer Learning Framework for Improving COVID-19 Diagnosis," arXiv:2101.05410 [cs, eess], Jan. 2021, arXiv: 2101.05410. [Online]. Available: http://arxiv.org/abs/2101.05410

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," arXiv:1603.05027 [cs], Jul. 2016, arXiv: 1603.05027. [Online]. Available: http://arxiv.org/abs/1603.05027

[18] M. Vakili, M. Ghamsari, and M. Rezaei, "Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification," arXiv:2001.09636 [cs, stat], Jan. 2020, arXiv: 2001.09636. [Online]. Available: http://arxiv.org/abs/2001.09636

[19] J. D. Novakovic, A. Veljovic, S. S. Ilic, A. Papic, and T. Milica, "Evaluation of Classification Models in Machine Learning," Theory and Applications of Mathematics & Computer Science, vol. 7, no. 1, pp. 39–46, Apr. 2017. [Online]. Available: https://www.uav.ro/applications/se/journal/index.php/TAMCS/article/view/158

[20] S. A. Hicks, I. Strumke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, and S. Parasa, "On evaluation metrics for medical applications of artificial intelligence," medRxiv, Tech. Rep., Apr. 2021, type: article. [Online]. Available: https://www.medrxiv.org/content/10.1101/2021.04.07.21254975v1

[21] A. C. Muller and S. Guido, Introduction to machine learning with Python: a guide for data scientists, first edition, Sebastopol, CA, 2016, oCLC: ocn895728667.

[22] R. Caruana, S. Lawrence, and L. Giles, "Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping," in Proceedings of the 13th International Conference on Neural Information Processing Systems, ser. NIPS'00. Cambridge, MA, USA: MIT Press, Jan. 2000, pp. 381–387.

[23] D. Wilson and T. Martinez, "The need for small learning rates on large problems," in IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222), vol. 1. Washington, DC, USA: IEEE, 2001, pp. 115–119. [Online]. Available: http://ieeexplore.ieee.org/document/939002/

[24] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, Jun. 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X

[25] D. M. W. Powers, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness correlation," Journal of Machine Learning Technologies, pp. 37–63, 2011. [Online]. Available: https://www.researchgate.net/publication/228529307

ISSN 2085-4552

Public Sentiment Analysis on the Transition from Analog to Digital Television Using the Random Forest Classifier Algorithm

Elfajar Bintang Samudera[1], Alexander Waworuntu[2], Ester Lumba[3]
[1,2] Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]elfajar.bintang@student.umn.ac.id, [2]alex.wawo@umn.ac.id
[3] Department of Informatics, Bunda Mulia University, Jakarta, Indonesia
[3]l0178@lecturer.ubm.ac.id

Accepted 3 June 2024
Approved 26 June 2024

***Abstract*— Television is one of the most popular media for entertainment and information. Analog television is the most widely used type among the public. However, with technological advancements, analog television is becoming obsolete and is being replaced by digital television, which offers better performance. On November 2, 2022, the Government officially mandated the transition from analog to digital broadcasting. This Analog Switch Off program has elicited various pro and con opinions among the public. X, a widely used social media platform, facilitates rapid communication and information dissemination among users. This study aims to classify public sentiment regarding the Analog Switch Off policy as either positive or negative. The classification model used is the Random Forest algorithm, with the Lexicon Inset for data labeling, Count Vectorizer and TF-IDF Vectorizer for data vectorization and weighting, and various train-test data splits. The study achieved the best classification performance using the Count Vectorizer method, with an 80%:20% train-test data ratio, yielding an accuracy of 88%, precision of 88%, recall of 88%, and an F1-score of 88%.**

***Index Terms*— Analog Television; Digital Television; Sentiment; Random Forest; X**.

I. INTRODUCTION

Television is a widely favored medium for entertainment and information. Its audiovisual nature enables it to offer various forms of entertainment such as movies, music, reality shows, sports, soap operas, and celebrity-related programs. In Indonesia, television is the most popular medium and serves as the primary promotional tool for industries to market products and services [1].

One of the oldest types of television among the public is analog television. Analog television uses analog signals to transmit images and sound, broadcasting a range of voltages and signal frequencies [1]. It requires a signal receiver called an antenna. The initial development of analog television involved the use of a disc with specific patterns for scanning images, known as mechanical television. Analog television programs are broadcasted by various national stations and are available for free.

Digital television, on the other hand, operates using digital modulation, where the carrier wave's properties are modified into bits (0 or 1) [2]. Digital television offers much higher resolution compared to analog television and provides more efficient use of the radio frequency spectrum. Digital television reception is also versatile, as it can receive broadcasts from regular transmission stations, regular antennas, cable television, and satellite television [3].

On November 2, 2022, the government mandated the migration from analog to digital television broadcasting. The cessation of analog television broadcasting is part of the digital transformation in the broadcasting sector, in accordance with the mandate of Law No. 11 of 2020 (Undang-undang No. 11 Tahun 2020 tentang Cipta Kerja). The implementation of this decision considers various benefits, such as reducing broadcast traffic congestion, optimizing spectrum usage, more efficient utilization of frequency resources, and providing more frequency space to accelerate internet access in Indonesia [1].

The implementation of the Analog Switch Off program across various regions has elicited a range of public opinions, both supportive and opposing. Director General Usman Kansong stated that transitioning to digital television offers benefits such as clearer pictures, better sound quality, and more advanced technology [1]. Conversely, Nailul Huda, a Digital Economy Analyst from the Institute for Development of Economics and Finance (INDEF), pointed out that switching to digital television could negatively impact industries by causing a loss of market share, affecting advertisements that would not reach those still using analog television [1]. Given these diverse sentiments, sentiment analysis plays a crucial role in this research. Sentiment analysis is the process of processing data to track public responses to a specific topic on the internet. With the advancement of information technology parallel to the broadcasting sector, platforms like X are

used for public responses regarding the transition from analog to digital television.

Sentiment analysis requires a method, and one frequently used is the Random Forest Classifier. The Random Forest Classifier is employed for classification processes. It works by constructing multiple decision trees and combining them to produce stable and accurate predictions [4]. The advantages of the Random Forest Classifier include its ability to handle noise and its suitability for classifying large datasets [4].

Several previous studies relevant to this research include Aisyah Nurul Izza et al.'s analysis of tourist attraction sentiments in South Sulawesi Province based on visitor reviews using the Random Forest Classifier method, which achieved an accuracy of 82% [5]. Hana Chyntis Morama et al. analyzed aspect-based sentiments regarding Hotel Tenrem Yogyakarta reviews using the Random Forest Classifier algorithm, achieving an accuracy and F1-score of 90% [6]. Evita Fitri et al. analyzed sentiments towards the Ruangguru application using the naive bayes, random forest, and support vector machine algorithms, achieving an accuracy of 97.16% and an AUC value of 0.996, with the best accuracy and performance obtained using the Random Forest algorithm [7]. Therefore, this research will analyze public sentiment regarding the transition from analog to digital television using data from X and the Random Forest Classifier algorithm. The aim of this research is to understand public sentiment towards the transition from analog to digital television and to determine the accuracy, precision, recall, and F1-score values produced by using the Random Forest Classifier algorithm on X data.

II. THEORETICAL FRAMEWORK

A. Sentiment Analysis

Sentiment analysis is the process of determining the sentiment of a text and categorizing it as either positive or negative [8]. It is often equated with opinion mining [9] because it focuses on opinions that express either positive or negative sentiments. Textual data such as product reviews, services, phenomena, and individuals can be the subject of sentiment analysis research [10].

B. Analog Switch Off

Analog Switch Off (ASO) is a policy that mandates the migration from analog to digital television broadcasting [3]. This policy aligns with technological and informational advancements in Indonesia. Government Regulation No. 46 of 2021 concerning Post, Telecommunications, and Broadcasting, in Article 72, paragraph 8, states that the migration from analog to digital terrestrial broadcasting must be completed no later than two years after its enactment. Consequently, the ASO policy was to be implemented by November 2, 2022 [2].

C. X

X is a social media platform that allows users to write and publish short, concise texts expressing their opinions [11]. In this research, the snscrape library will be used. Snscrape is a scraping tool for social networking services (SNS). This library can scrape data such as users, user profiles, hashtags, searches, and posts without using the X API.

D. Text Preprocessing

Text preprocessing is a crucial step aimed at preventing significant degradation in the performance of analysis [12]. Generally, text preprocessing is divided into several stages, such as data cleaning, data labeling, case folding, stemming, and tokenizing [13]. The details of these stages are as follows:

1) Data Cleaning: In this stage, the data is cleaned by removing characters such as symbols. Additionally, emojis, URLs, and usernames are also removed. This step aims to reduce and avoid disruptions in the classification results [14].
2) Data Labeling: In this stage, a process is conducted to label the tweets. This is done using existing libraries to classify the data into positive or negative sentiments.
3) Case Folding: This stage involves converting text into a uniform format, specifically by converting all text to lowercase [14].
4) Stemming: Stemming is the process of reducing words to their base form, known as the stem. This process helps in reducing the vocabulary size that the NLP model needs to process, thereby enhancing the model's efficiency [14].
5) Tokenizing: In this stage, sentences are broken down into individual words, known as tokens [14].

E. Decision Tree

The Decision Tree is a popular classification method due to its ease of interpretation by humans. It employs a tree or hierarchical structure to create a predictive model [15]. The concept behind this algorithm is to transform data into a decision tree and decision rules. It simplifies complex decision-making processes into a more straightforward form.

The Decision Tree is named so because the rules formed resemble a tree structure. The data within the decision tree are interpreted in a table format with attributes and records. The Decision Tree has nodes representing attributes, where each branch represents the result of a test, and leaf nodes represent the classes [16]. The construction of a decision tree involves three types of nodes: the root node, internal nodes, and leaf nodes. The root node is the top node with no input and more than one output. Internal nodes are branching nodes with one input and at least two outputs. Leaf

nodes are the terminal nodes with one input and no output.

The decision tree construction starts by calculating the entropy to determine the impurity of the attributes and the information gain [17].The formula for calculating entropy is: (1)

$$\text{Entropy } (S) = \sum_{i=1}^{c} - p_i \log_2 p_i \quad (1)$$

where ***pi*** represents the proportion of samples in subset ***S*** with the ***i***-th attribute value. Information gain is a metric used in the segmentation process [18]. The formula for calculating information gain is: (2)

$$\text{Gain } (S, A) = \sum_{V \in V(A)} \frac{|S_V|}{|S|} \text{Entropy } (S_V) \quad (2)$$

where ***V(A)*** is the range of attribute ***A***, and S_v is the subset of ***S*** that has the same value as attribute ***V***.

F. Random Forest Classifier

The Random Forest Classifier is an algorithm used for classifying large datasets. It combines multiple decision trees into a single model, enhancing accuracy with an increasing number of trees [5]. Each tree in the Random Forest model contains a collection of random variables, and the aggregation of these trees leads to improved accuracy.

The Random Forest algorithm combines each decision tree model into one comprehensive model. The number of trees used significantly impacts the accuracy, precision, and recall of the Random Forest model. The selection of trees from the decision tree model begins with calculating the entropy value to find the best trees for use in the Random Forest model [19].

The steps involved in the Random Forest algorithm begin with selecting random samples from the provided dataset. For each selected sample, a decision tree is constructed, and predictions are obtained from these decision trees. A voting process is then conducted for each prediction, where the most frequent value (mode) is used for classification problems, and the average value (mean) is used for regression problems. Finally, the algorithm determines the final prediction by selecting the prediction with the majority vote.

G. Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) is a technique used to assign a weight or value to a word within a document [20]. TF-IDF can be applied to various tasks, such as token extraction from articles, ranking determination, and calculating the similarity between documents. Term Frequency (TF) indicates how often a word appears in a document [20], while Inverse Document Frequency (IDF) signifies the importance of a word [20]. The formula for calculating TF-IDF is as follows [21]:

$$TF - \text{IDF}_{t,d} = \text{TF}_{t,d} \times iDF_t \quad (3)$$

$$IDF_t = \log \frac{N}{DF_t} \quad (4)$$

$$TF_{t,d} = \frac{n_{i,j}}{\sum_k n_{i,j}} \quad (5)$$

where:

- ***TF-IDF$_{t,d}$*** represents the weight of a word (***t***) in a document (***d***).
- ***TF$_{t,d}$*** is the Term Frequency.
- ***IDF$_t$*** is the Inverse Document Frequency.
- ***DFt*** is the number of documents containing the word.
- ***$n_{i,j}$*** is the total occurrences of the term in one document.
- $\sum_k n_{i,j}$ is the total number of words in one document.
- ***N*** is the total number of documents.
- ***t*** denotes the words being calculated.
- ***d*** is the document weight.

H. Hyperparameter Tuning

Hyperparameter tuning is the process of finding the optimal combination of parameters for a machine learning model [22]. The goal of this process is to identify the best hyperparameter combination to enhance performance and reduce the risk of overfitting and underfitting [22]. In this study, the following parameters will be optimized using the random search method:

1) **n_estimators:** This parameter refers to the number of decision trees to be created by the algorithm. A higher value of n_estimators results in more trees being formed [22].
2) **min_samples_split:** This parameter indicates the minimum number of samples required to split an internal node in a tree [22].
3) **min_samples_leaf:** This parameter specifies the minimum number of samples that must be present in a leaf node [22].
4) **max_depth:** This parameter denotes the maximum depth of each decision tree within the ensemble method [22].

5) **max_features:** This parameter controls the number of features randomly selected to build each decision tree [22].

6) **bootstrap:** This parameter manages the use of bootstrapping when constructing each decision tree. If set to True, each tree is built using bootstrap samples. If set to False, each tree is constructed using the original training dataset [22].

I. Confusion Matrix

The Confusion Matrix is a method used to calculate the accuracy of a classification model. It is represented as a table that shows the number of correct and incorrect classifications for the test data [10]. Several metrics can be derived from the Confusion Matrix to evaluate a classification model, including accuracy, recall, precision, and F1-score [4].

Accuracy represents the percentage of correctly classified tuples in the test data. It is calculated using the formula:

$$\text{accuracy} = \frac{TP+T}{TP+TN+FP+F} \quad (6)$$

Recall indicates the rate at which positive tuples are correctly identified, measuring how well the model identifies true positives. It is calculated using the formula:

$$\text{recall} = \frac{TP}{TP+FN} \quad (7)$$

Precision represents the percentage of tuples labeled as positive that are actually positive. It is calculated using the formula:

$$\text{precision} = \frac{TP}{TP+FP} \quad (8)$$

The F1-score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is calculated using the formula:

$$f1 - \text{score} = \frac{2x \text{ Precision } x \text{ Recall}}{\text{Recall + Precision}} \quad (9)$$

III. Research Methodology

This study involves several methodologies and processes, as outlined below:

1) **Literature Review:** The literature review process begins with studying relevant literature and theories pertinent to this research. The theories covered include sentiment analysis, analog switch off, X, text preprocessing, decision tree, random forest, Term Frequency-Inverse Document Frequency (TF-IDF), and confusion matrix.

2) **Data Collection:** In this stage, data collection involves gathering tweets from Indonesian-speaking users about their opinions on the transition from analog to digital television. The collected data will form a dataset that will be processed in subsequent stages. Data collection is conducted using the snscrape library. Figure 1 illustrates the data collection process in a flowchart format.

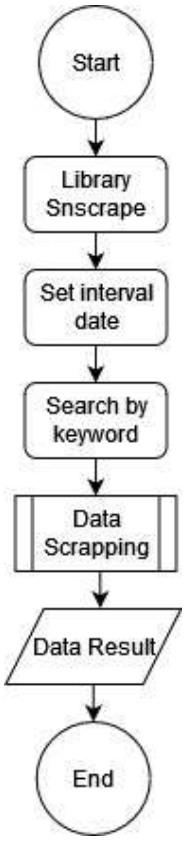


Fig. 1. Data Scraping Flowchart

3) **Data Processing:** This stage, also known as text preprocessing, involves several critical steps. First, data cleaning is performed to remove non-alphabet characters, symbols, emoticons, URLs, and usernames to prevent disruptions in classification results. Next, casefolding converts all words and sentences to lowercase, ensuring a uniform text format and eliminating any capitalized words or sentences. Tokenizing then segments the text into smaller parts, or tokens. In the stopword removal step, commonly used words that have little impact on the sentence's meaning, such as "and," "or," and "which," are removed. Normalizing involves

identifying and replacing incorrect word forms with their correct versions according to the Kamus Besar Bahasa Indonesia (KBBI), ensuring all spellings and abbreviations match KBBI standards. Finally, stemming processes words with affixes to convert them to their root forms by removing the affixes. Figure 2 illustrates these data processing steps in a flowchart format.

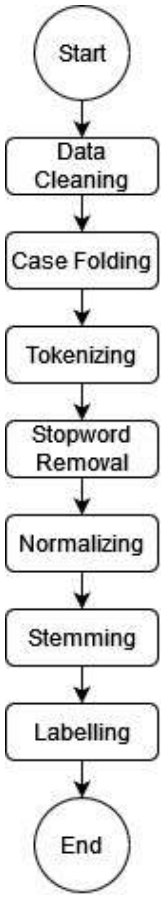


Fig. 2. Preprocessing Flowchart

4) **Applying TF-IDF Features:** TF-IDF is used for various tasks such as token extraction from articles, ranking, and calculating document similarity. Term Frequency (TF) indicates how often a word appears in a document [20], while Inverse Document Frequency (IDF) indicates the importance of a word [20]. Figure 3 illustrates the process of applying TF-IDF features in a flowchart format.

GAMBAR

In addition to TF-IDF vectorization, this research also employs the Count Vectorizer method. The difference between these methods lies in their output and process. Count Vectorizer uses the concept of term frequency alone, where the weight of each word appearing in a document is calculated, resulting in a bag-of-words model [23]. Unlike TF-IDF vectorization, Count Vectorizer does not calculate the IDF score for each feature, thus not assessing the importance of a feature within a document.

5) **Implementation of the Random Forest Classifier Algorithm:** In this stage, a model is created using data that has undergone the text preprocessing steps. The output from text preprocessing is processed using the Random Forest algorithm. The Random Forest model is trained to provide classification predictions for positive and negative sentiments.

6) **Evaluation Testing:** This stage involves testing the model by measuring the values of accuracy, recall, precision, and F1-score. These measurements are obtained by evaluating the results of the Random Forest model using a confusion matrix.

IV. Results and Discussion

A. Testing Scenario

In the model testing phase using Random Forest, various scenarios were implemented. These scenarios included the use of Count Vectorizer and TF-IDF Vectorizer, testing the model on different dataset sizes split with test sizes of 20%, 30%, 40%, and 50%, and employing hyperparameter tuning with the random search method to identify optimal parameters. The various scenarios implemented in the machine learning model are presented in Table I. After multiple processes and testing scenarios, the evaluation results for the best model performance for each vectorization method are shown in Table II.

B. Discussion

Based on the evaluation results in the table above, the model that produced the best performance was in the first scenario, where the training and testing data split ratio was 80%:20%, and the text data vectorization method used was Count Vectorizer. This model achieved an accuracy of 88%, with identical values for recall, precision, and F1-score. As the dataset's split ratio for training and testing data was adjusted, the accuracy significantly decreased, while the other metrics also declined, though less dramatically. The lowest performance for the Count Vectorizer method occurred in the eighth scenario, with a dataset split ratio of 50%:50% and hyperparameter tuning using random search. Despite the tuning, the results were not as good as the previous models due to the random nature of parameter selection, which did not consider inter-parameter relationships.

In the scenarios using the TF-IDF Vectorizer method, the best model performance was obtained in the tenth scenario. Here, the training and testing data split ratio was 80%:20%, with fine-tuning involving

600 trees, a minimum of 10 samples required for splitting, a minimum of 1 sample per leaf, automatic feature selection for splitting, a maximum tree depth of 90, and each tree built using the entire training dataset without bootstrap sampling. This model achieved an accuracy of 87.75%, with recall, precision, and F1-score values of 88%. The worst performance for the TF-IDF Vectorizer method occurred in the fifteenth scenario, with a training and testing data split ratio of 50%:50%. Despite using two different text data vectorization methods, the worst performance occurred with a split ratio of 50%:50% for training and testing data.

TABLE I. TABLE OF MACHINE LEARNING SCENARIO TEST

No.	Scenario
1	CountVectorizer()
	test_size 20%
2	CountVectorizer()
	test size 20% + Hyperparameter Tuning
3	CountVectorizer()
	test size 30%
4	CountVectorizer()
	test size 30% + Hyperparameter Tuning
5	CountVectorizer()
	test size 40%
6	CountVectorizer()
	test size 40% + Hyperparameter Tuning
7	CountVectorizer()
	test size 50%
8	CountVectorizer()
	test size 50% + Hyperparameter Tuning
9	TFIDFVectorizer()
	test size 20%
10	TFIDFVectorizer()
	test size 20% + Hyperparameter Tuning
11	TFIDFVectorizer()
	test size 30%
12	TFIDFVectorizer()
	test size 30% + Hyperparameter Tuning
13	TFIDFVectorizer()
	test size 40%
14	TFIDFVectorizer()
	test size 40% + Hyperparameter Tuning
15	TFIDFVectorizer()
	test size 50%
16	TFIDFVectorizer()
	test size 50% + Hyperparameter Tuning

TABLE II. TABLE OF TEST RESULT WITH HIGHEST AND LOWEST PERFORMANCE FOR EACH VECTORIZATION METHOD

Model	S	A	P	R	F1
CountVectorizer 80%:20%	1	88%	88%	88%	88%
CountVectorizer 50%:50%	8	80.68%	81%	81%	81%
TFIDF80%:20% Hyperparameter Tuning	10	87.75%	88%	88%	88%
TF-IDF 50%:50%	15	81.78%	82%	82%	82%

Legend:
S: Scenario; A: Accuracy; P: Precision; R: Recall; F1: F1 Score

V. CONCLUSION

Based on the research conducted on sentiment analysis using the Random Forest algorithm, several conclusions can be drawn. The sentiment analysis of public opinion regarding the transition from analog to digital television using the Random Forest algorithm was successfully implemented. This study utilized Count Vectorizer and TF-IDF methods for word weighting, and Random Search for hyperparameter tuning. The machine learning model that achieved the highest accuracy, precision, recall, and F1-score was the one using Count Vectorizer for word weighting, with a train-test split ratio of 80%:20%, without undergoing hyperparameter tuning. The results obtained were 88.00% accuracy, 88% precision, 88% recall, and 88% F1-score. Model performance significantly decreased, particularly in terms of accuracy, with an increase in test data size. The lowest performance was observed in the model using the Count Vectorizer method, with a train-test split ratio of 50%:50% that had undergone hyperparameter tuning. The results were 80.60% accuracy, 81% precision, 81% recall, and 81% F1-score.

REFERENCES

[1] K. Alfarizi, "Siaran digital indonesia — gugus tugas migrasi siaran televisi analog ke digital," Nov. 2022.

[2] A. D. Gultom, "Digitalisasi penyiaran televisi di indonesia," Buletin Pos dan Telekomunikasi, vol. 16, pp. 91–100, Dec. 2018. [Online]. Available: https://bpostel.kominfo.go.id/index.php/bpostel/article/view/160202

[3] M. Mubarok and M. D. Adnjani, "Kesiapan industri tv lokal di jawa tengah menuju migrasi penyiaran dari analog ke digital," Communicare: Journal of Communication Studies, vol. 7, no. 1, pp. 18–32, 2020.

[4] M. Y. Aldean, P. Paradise, and N. A. S. Nugraha, "Analisis sentimen masyarakat terhadap vaksinasi covid-19 di twitter menggunakan metode random forest classifier (studi kasus: Vaksin sinovac)," Journal of Informatics Information System Software Engineering and Applications (INISTA), vol. 4, pp. 64–72, Jun. 2022. [Online]. Available: https://journal.ittelkom-pwt.ac.id/index.php/inista/article/view/575

[5] A. N. Izza, D. E. Ratnawati, W. Hayuhardhika, N. Putra, and P. Korespondensi, "Analisis sentimen objek wisata di provinsi sulawesi selatan berdasarkan ulasan pengunjung menggunakan metode random forest classifier," Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi, vol. 3, pp. 97–105, Oct. 2022. [Online]. Available: https://just-si.ub.ac.id/index.php/just-si/article/view/77

[6] H. C. Morama, D. E. Ratnawati, and I. Arwani, "Analisis sentimen berbasis aspek terhadap ulasan hotel tentrem yogyakarta menggunakan algoritma random forest classifier," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2548, p. 964X, 2022.

[7] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis sentimen terhadap aplikasi ruangguru menggunakan algoritma naive bayes, random forest dan support vector machine," Jurnal Transformatika, vol. 18, pp. 71–80, Jul. 2020. [Online]. Available: https://journals.usm.ac.id/index.php/transformatika/article/view/2317

[8] A. Perdana, A. Hermawan, and D. Avianto, "Analisis sentimen terhadap isu penundaan pemilu di twitter menggunakan naive bayes classifier," Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 11, pp. 195–200, Jul. 2022. [Online].

Available: http://jurnal.atmaluhur.ac.id/index.php/sisfokom/article/view/1412

[9] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, 2012.

[10] A. M. Zuhdi, E. Utami, and S. Raharjo, "Analisis sentiment twitter terhadap capres indonesia 2019 dengan metode k-nn," Jurnal Informa: Jurnal Penelitian dan Pengabdian Masyarakat, vol. 5, pp. 1–7, Aug. 2019. [Online]. Available: http://www.poltekindonusa.ac.id/SUBDOMAIN/informa/index.php/informa/article/view/73

[11] A. K. Fauziyyah and D. H. Gautama, "Analisis sentimen pandemi covid19 pada streaming twitter dengan text mining python," Jurnal Ilmiah SINUS, vol. 18, pp. 31–42, Jul. 2020. [Online]. Available: https://p3m.sinus.ac.id/jurnal/index.php/ejurnal SINUS/article/view/491

[12] M. Syarifuddinn, "Analisis sentimen opini publik terhadap efek psbb pada twitter dengan algoritma decision tree, knn, dan naive bayes," INTI Nusa Mandiri, vol. 15, pp. 87–94, Aug. 2020. [Online]. Available: http://ejournal.nusamandiri.ac.id/index.php/inti/article/view/1433

[13] R. T. Vulandari, "Data mining: Teori dan aplikasi rapidminer," 2017.

[14] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan algoritma naive bayes untuk analisis sentimen review data twitter bmkg nasional," Jurnal Tekno Kompak, vol. 15, pp. 131–145, Feb. 2021. [Online]. Available: https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/744

[15] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi dengan pohon keputusan berbasis algoritme c4.5," in PRISMA, Prosiding Seminar Nasional Matematika, vol. 3, 2020, pp. 64–71.

[16] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Perbandingan algoritma random forest, naive bayes, dan support vector machine pada analisis sentimen twitter mengenai opini masyarakat terhadap penghapusan tenaga honorer," Seminar Nasional Official Statistics, vol. 2022, pp. 563–572, Nov. 2022. [Online]. Available: https://prosiding.stis.ac.id/index.php/semnasoffstat/article/view/1410

[17] B. B. Baskoro, I. Susanto, S. Khomsah, P. Informatika, P. S. Data, I. D. T. T. P. J. Panjaitan, and J. Tengah, "Analisis sentimen pelanggan hotel di purwokerto menggunakan metode random forest dan tf-idf (studi kasus: Ulasan pelanggan pada situs tripadvisor)," Journal of Informatics Information System Software Engineering and Applications (INISTA), vol. 3, pp. 21–29, Jun. 2021. [Online]. Available: https://journal.ittelkom-pwt.ac.id/index.php/inista/article/view/218

[18] B. T. Jijo and A. M. Abdulazeez, "Classification based on decision tree algorithm for machine learning," vol. 02, pp. 20–28, 2021. [Online]. Available: www.jastt.org

[19] M. Huljanah, Z. Rustam, S. Utama, and T. Siswantining, "Feature selection using random forest classifier for predicting prostate cancer," IOP Conference Series: Materials Science and Engineering, vol. 546, p. 052031, Jun. 2019. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052031 https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052031/meta

[20] Imamah and F. H. Rachman, "Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regression," 2020 Information Technology International Seminar (ITIS), pp. 238–242, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9320958

[21] M. Guia, R. R. Silva, and J. Bernardino, "Comparison of Naive Bayes, Support Vector Machine, Decision Trees and Random Forest on Sentiment Analysis," 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), pp. 525–531, 2019.

[22] R. Ahuja, K. Vats, C. Pahuja, T. Ahuja, and C. Gupta, "Pragmatic analysis of classification techniques based on hyper-parameter tuning for sentiment analysis," EasyChair, Tech. Rep., 2020.

[23] H. P. Doloksaribu and Y. T. Samuel, "Komparasi algoritma data mining untuk analisis sentimen aplikasi pedulilindungi," Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika, vol. 16, no. 1, pp. 1–11, 2022.

AUTHOR GUIDELINES

1. Manuscript criteria

- The article has never been published or in the submission process on other publications.
- Submitted articles could be original research articles or technical notes.
- The similarity score from plagiarism checker software such as Turnitin is 20% maximum.
- For December 2021 publication onwards, Ultimatics : Jurnal Teknik Informatika will be receiving and publishing manuscripts written in English only.

2. Manuscript format

- Article been type in Microsoft Word version 2007 or later.
- Article been typed with 1 line spacing on an A4 paper size (21 cm x 29,7 cm), top-left margin are 3 cm and bottom-right margin are 2 cm, and Times New Roman's font type.
- Article should be prepared according to the following author guidelines in this template. Article contain of minimum 3500 words.
- References contain of minimum 15 references (primary references) from reputable journals/conferences

3. Organization of submitted article

The organization of the submitted article consists of Title, Abstract, Index Terms, Introduction, Method, Result and Discussion, Conclusion, Appendix (if any), Acknowledgment (if any), and References.

- Title
  The maximum words count on the title is 12 words (including the subtitle if available)
- Abstract
  Abstract consists of 150-250 words. The abstract should contain logical argumentation of the research taken, problem-solving methodology, research results, and a brief conclusion.
- Index terms
  A list in alphabetical order in between 4 to 6 words or short phrases separated by a semicolon (;), excluding words used in the title and chosen carefully to reflect the precise content of the paper.
- Introduction
  Introduction commonly contains the background, purpose of the research, problem identification, research methodology, and state of the art conducted by the authors which describe implicitly.
- Method
  Include sufficient details for the work to be repeated. Where specific equipment and materials are named, the manufacturer's details (name, city and country) should be given so that readers can trace specifications by contacting the manufacturer. Where commercially available software has been used, details of the supplier should be given in brackets or the reference given in full in the reference list.
- Results and Discussion
  State the results of experimental or modeling work, drawing attention to important details in tables and figures, and discuss them intensively by comparing and/or citing other references.
- Conclusion
  Explicitly describes the research's results been taken. Future works or suggestion could be explained after it
- Appendix and acknowledgment, if available, could be placed after Conclusion.
- All citations in the article should be written on References consecutively based on its' appearance order in the article using Mendeley (recommendation). The typing format will be in the same format as the IEEE journals and transaction format.

4. Reviewing of Manuscripts

Every submitted paper is independently and blindly reviewed by at least two peer-reviewers. The decision for publication, amendment, or rejection is based upon their reports/recommendations. If two or more reviewers consider a manuscript unsuitable for publication in this journal, a statement explaining the basis for the decision will be sent to the authors within six months of the submission date.

5. Revision of Manuscripts

Manuscripts sent back to the authors for revision should be returned to the editor without delay (maximum of two weeks). Revised manuscripts can be sent to the editorial office through the same online system. Revised manuscripts returned later than one month will be considered as new submissions.

6. Editing References

- **Periodicals**
  J.K. Author, “Name of paper,” Abbrev. Title of Periodical, vol. x, no. x, pp. xxx-xxx, Sept. 2013.

- **Book**
  J.K. Author, “Title of chapter in the book,” in Title of His Published Book, xth ed. City of Publisher, Country or Nation: Abbrev. Of Publisher, year, ch. x, sec. x, pp xxx-xxx.

- **Report**
  J.K. Author, “Title of report,” Abbrev. Name of Co., City of Co., Abbrev. State, Rep. xxx, year.

- **Handbook**
  Name of Manual/ Handbook, x ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. xxx-xxx.

- **Published Conference Proceedings**
  J.K. Author, “Title of paper,” in Unabbreviated Name of Conf., City of Conf., Abbrev. State (if given), year, pp. xxx-xxx.

- **Papers Presented at Conferences**
  J.K. Author, “Title of paper,” presented at the Unabbrev. Name of Conf., City of Conf., Abbrev. State, year.

- **Patents**
  J.K. Author, “Title of patent,” US. Patent xxxxxxx, Abbrev. 01 January 2014.

- **Theses and Dissertations**
  J.K. Author, “Title of thesis,” M.Sc. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year. J.K. Author, “Title of dissertation,” Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

- **Unpublished**
  J.K. Author, “Title of paper,” unpublished. J.K. Author, “Title of paper,” Abbrev. Title of Journal, in press.

- **On-line Sources**
  J.K. Author. (year, month day). Title (edition) [Type of medium]. Available: http:// www.(URL) J.K. Author. (year, month). Title. Journal [Type of medium]. volume(issue), pp. if given. Available: http://www.(URL) Note: type of medium could be online media, CD-ROM, USB, etc.

7. Editorial Adress

Universitas Multimedia Nusantara
Jl. Scientia Boulevard, Gading Serpong
Tangerang, Banten, 15811
Email: **ultimatics@umn.ac.id**

Paper Title

Subtitle (if needed)

Author 1 Name[1], Author 2 Name[2], Author 3 Name[2]
[1] Line 1 (of affiliation): dept. name of organization, organization name, City, Country
Line 2: e-mail address if desired
[2] Line 1 (of affiliation): dept. name of organization, organization name, City, Country
Line 2: e-mail address if desired

Accepted on mmmmm dd, yyyy
Approved on mmmmm dd, yyyy

***Abstract*—This electronic document is a "live" template which you can use on preparing your ULTIMATICS paper. Use this document as a template if you are using Microsoft Word 2007 or later. Otherwise, use this document as an instruction set. Do not use symbol, special characters, or Math in Paper Title and Abstract. Do not cite references in the abstract.**

***Index Terms*—enter key words or phrases in alphabetical order, separated by semicolon (;)**

I. INTRODUCTION

This template, modified in MS Word 2007 and saved as a Word 97-2003 document, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. Margins, column widths, line spacing, and type styles are built-in here. The authors must make sure that their paper has fulfilled all the formatting stated here.

Introduction commonly contains the background, purpose of the research, problem identification, and research methodology conducted by the authors which been describe implicitly. Except for Introduction and Conclusion, other chapter's title must be explicitly represent the content of the chapter.

II. EASE OF USE

A. Selecting a Template

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

B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them.

III. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not add any kind of pagination anywhere in the paper. Please take note of the following items when proofreading spelling and grammar.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Abbreviations that incorporate periods should not have spaces: write "C.N.R.S.," not "C. N. R. S." Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units (SI units are encouraged).
- Do not mix complete spellings and abbreviations of units: "Wb/m2" or "webers per square meter," not "webers/m2." Spell units when they appear in text: "...a few henries," not "...a few H."
- Use a zero before decimal points: "0.25," not ".25."

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number the equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop.

$$\int_0^{r_2} F(r,\phi)\, dr\, d\phi = [\sigma r_2 / (2\mu_0)] \qquad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the

equation. Use "(1)," not "Eq. (1)" or "equation (1)," except at the beginning of a sentence: "Equation (1) is ..."

D. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o."
- In American English, commas, semi-/colons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset," not an "insert." The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively."
- In your paper title, if the words "that uses" can accurately replace the word using, capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect," "complement" and "compliment," "discreet" and "discrete," "principal" and "principle."
- Do not confuse "imply" and "infer."
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al."
- The abbreviation "i.e." means "that is," and the abbreviation "e.g." means "for example."

IV. Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention as below

ULTIMATICS_firstAuthorName_paperTitle.

In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper. Please take note on the following items.

A. Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

B. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include ACKNOWLEDGMENTS and REFERENCES, and for these, the correct style to use is "Heading 5."

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles, named "Heading 1," "Heading 2," "Heading 3," and "Heading 4", are prescribed.

C. Figures and Tables

Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1," even at the beginning of a sentence.

TABLE I. TABLE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy		

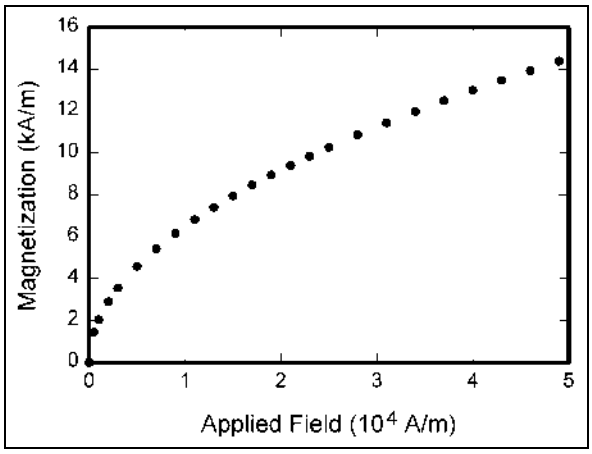


Fig. 1. Example of a figure caption

V. Conclusion

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

Appendix

Appendixes, if needed, appear before the acknowledgment.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments. Avoid expressions such as "One of us (S.B.A.) would like to thank ... ." Instead, write "F. A. Author thanks ... ." You could also state the sponsor and financial support acknowledgments here.

References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

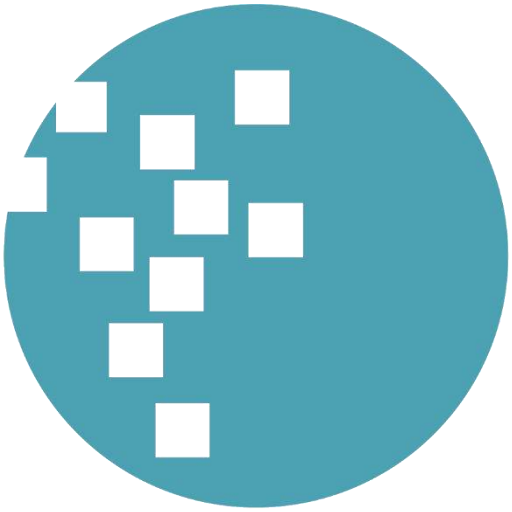
[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ISSN 2085-4552

Universitas Multimedia Nusantara
Scientia Garden Jl. Boulevard Gading Serpong, Tangerang
Telp. (021) 5422 0808 | Fax. (021) 5422 0800