ISSN 2085-4552

ULTIMATICS

Jurnal Teknik Informatika

Engineering | Vol. 14 | No. 2 | Page 51 - 118 | December 2022 | 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 December 2022 edition, Ultimatics enters the 2nd Edition of Volume 14. 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: COVID-19 Fake News Detection With Pre-trained Transformer Models, The Social Impact of VR Technology on Society: A Systematic Literature Review, Design and Build a Website for Catering Sales and Orders Using Web Engineering, Automatic Portal Access Application Using Static QR Code Reading, Automatic Portal Access Application Using Static QR Code Reading, Use of eGovernment in Public Services Area Case Study: Bojonegoro Regency, Implementing the Chaotic Permutation Multicircular Cryptography Technique using Asymmetric Key, Topic Modelling Using VSM-LDA For Document Summarization, Crude Oil Price Forecasting Using Long Short-Term Memory and Support Vector Regression, Air Temperature Sensor Estimation on Automatic Weather Station Using ARIMA and MLP and Analyzing Level of International Humanitarian Law Knowledge and its Compliance Through Military Simulation Game.

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 2022 Edition of Ultimatics. We hope that scientific articles from research in this journal can be useful and contribute to the development of research and science in Indonesia.

December 2022,

M.B.Nugraha, S.T., M.T.
Editor-in-Chief

ISSN 2085-4552

TABLE OF CONTENT

ISSN 2085-4552

COVID-19 Fake News Detection With Pre-trained Transformer Models

Bakti Amirul Jabar[1], Seline[2], Bintang[3], Cameron Jane Victoria[4], Rio Nur Arifin[5]
[1,2,3,4]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[5]Data Scientist, Telkomsel, Jakarta, Indonesia
[1]bakti.jabar@binus.ac.id, [2]seline@binus.ac.id, [3]bintang001@binus.ac.id, [4]cameron.victoria@binus.ac.id, [5]rioarifin@gmail.com

Accepted 06 September 2022
Approved 06 October 2022

***Abstract*— COVID-19 is a new virus that first appeared in the year 2020 and is still currently plaguing our world. With the emergence of this virus, much information, both fake and real, has circulated in the internet. Fake information can lead to misleading information and cause a riot in society. In this paper, we aim to build a hoax detection system using the pre-trained transformer models BERT, RoBERTa, DeBERTa and Electra. From these four models, we will find which model gives the most accurate results. BERT gives a validation accuracy of 97.15% and test accuracy of 97.01%. RoBERTa gives a validation accuracy of 97.34% and test accuracy of 97.15%. DeBERTa gives a test accuracy of 97.48% and a test accuracy of 97.25%. Lastly, Electra gives a validation accuracy of 97.95% and a test accuracy of 97.76%. Electra is one of the newer models and is proven to be the most accurate model in our experiment and the one we will choose to implement fake news detection.**

***Index Terms*— deep learning; fake news; fake news detection; hoax; hoax detection; pre-trained transformer model; transformer model**.

I. Introduction

Since the start of the pandemic, millions of people have fallen victim and lost their lives. As of 17 June 2022, according to WHO, a total of 535,863,950 confirmed cases and 6,314,972 deaths of COVID-19 were reported [1]. In these two years, cases of COVID-19 have stagnated and surged continuously, contributing to many deaths. A significant factor that has also played a part in the surge is the existence of hoaxes or misinformation that has prevented individuals from following the correct health protocols or treatment for the virus [2].

Some of the most persistent hoaxes on the internet are about the COVID-19 vaccines. People claim that a person will die within three years of receiving the vaccine. Another claim is that the vaccines contain magnetic chips to track our location. They even create videos showing proof that metals stick to them after vaccination. Although these informations are most definitely false, a lack of, changing, or conflicting information also gives birth to these misinformation [2].

In modern natural language processing, pre-training has become the standard approach. This is done by pre-training the model on large amounts of unlabelled data followed by fine-tuning using small and specific data sets [4]. A research paper on machine learning implementations for hoax detection [5] has shown that pre-trained transformer models such as BERT and RoBERTa have outperformed traditional deep learning models and neural networks such as CNN (Convolutional Neural Network).

As such, this paper will focus on designing a COVID-19 hoax detection system by experimenting with four pre-trained deep learning models: BERT, RoBERTa, DeBERTa, and Electra. These four pre-trained models will be compared, and one of the models will be chosen for implementation based on their best validation accuracy and test accuracy.

II. Literature Review

Based on the research of Oberiri Destiny Apuke and Bahiyah Omar (2020, 18 October), there are four factors of news sharing in the COVID-19 pandemic [3]. The first factor is altruism. Altruism refers to the human activity of presenting something to others without expecting something in return. By sharing the news, people feel the satisfaction of successfully contributing to their social surroundings. Social media's primary function is to create and maintain social interactions between its users. Content sharing is considered to be a form of participation in building relationships. Therefore, socialization is considered the second factor.

A practice of making a perfect image on social media is called self-promotion. By sharing news they discover with others, they want to prove they are talented, capable and/or intelligent. To maintain this image, people develop the critical thinking of choosing which news is true and can be shared. The last factor is instant news sharing. People often share news on different platforms on social media without ensuring its factuality. This can be driven by fear and anxiety due to numerous news in the tense global conditions.

Three of the four factors mentioned above (altruism, socialization, and instant news sharing) are positively associated with fake news sharing. These factors are those that we cannot control; however, we can also use these attributes of society to our advantage. By propagating our fake news detector system, anyone can verify the news they hear on the internet and further spread the credibility to others.

A hoax can be classified into seven categories: satire or parody, misleading content, imposter content, fabricated content, false connection, fake news, and manipulated content. The top three types of the most widely spread hoax are 1). Manipulated content (30%), 2). Misleading content (28%), and 3). Fake news (18%) [6].

Fake news is low-quality news containing intentionally false information [3]. Health has become the most widely spread fake news topic (37% of the fake news takes up health as the topic) [6]. Fake news on health is very dangerous. It will lead people to improper ways of healthcare or/and will cause discredit to the medical world.

The COVID-19 pandemic has caused thousands of deaths in various parts of the world. At the same time, this pandemic has brought another disaster in our society, called the 'infodemic' or an abundance of false information circulating during the pandemic. The misinformation has proliferated widely on social media. This false information ranges from fake cures, conspiracies, or dangerous health advice. The large number of deaths caused by the virus is attributed to the virus itself and to the wrong or even late medical treatment due to hoaxes spreading online [7]. This is why we need to be able to distinguish between fake and factual information. However, we humans can be influenced by emotions such as fear, panic, and sadness, making rational decisions difficult, particularly during these difficult times. As such, by creating a hoax detection system, we can rely on it to find out whether particular information is credible without further questioning the reliability of choice.

Since the deep learning concept was created, researchers have designed multiple systems for fake news detection. At first, research on fake news detection mainly used traditional machine learning models such as Support Vector Machine, Naive Bayes, and Logistic Regression [8] [9] [10]. Later on, deep learning models such as Convolutional Neural Network and Long Short Term Memory Network were used [8] [9] [11] [12] [13] [14] [15]. Recently, pre-trained transformer models were developed and found to be more accurate than the other models. Pre-trained models were created to be more efficient than other deep learning models. They were also proven to be more accurate when trained using small datasets, making them suitable for fake news detection [16].

Therefore, we have concluded that most researchers utilize pre-trained transformer models for fake news detection. Although old neural network models such as CNN and RNN are still used, pre-trained transformer models have proven to be more efficient and accurate due to their improvements [17]. Transformer models started with the invention of BERT [13] [18] [19] [20] [21] [22] during 2018 [23], followed by its variations such as BART [24], ROBERTa [5] [18] [22] [25] [26], DeBERTa [18] [26] and Electra [21] [22] [27].

In this paper, we will conduct our experiment using BERT, RoBERTa, DeBERTa and Electra. After comparing the results, we will opt for the most accurate model.

III. Material and Method

A. Data Set

The data set consists of English tweets about COVID-19 taken in 2020, which have been checked in advance for their correctness. The dataset compiled has been taken from a popular dataset website, Kaggle. The dataset has been classified into two classes that are - real and fake. The total data consists of 10700 news items, 37050 words, and 5141 common words in fake and real news data. From the data collected, 52% are classified as real news and 48% as fake news. The data has been collected from 880 unique usernames. Table I contains 2 sample data from the COVID-19 news dataset.

TABLE I. Sample News Data

ID	Tweet	Label
1	Our daily update is published. States reported 734k tests, 39k new cases, and 532 deaths. Current hospitalizations fell below 30k for the first time since June 22. https://t.co/wzSYMe0Sht	Real
2	Alfalfa is the only cure for COVID-19.	fake

B. Methodology

1) Text Pre-Processing

Text pre-processing is a process in which text data is converted to be more structured so that the data can be used for analysis or prediction. In this paper, we use nltk, pre-processor, and tweet-preprocessing library from python. Basically, those libraries did data cleaning such as removing stopwords, stemming, lemmatization, remove punctuation, lower-casing, remove numbers, and over spaces or ticks.

2) Tokenization

Tokenization is an activity of splitting an entire text into small units, also known as tokens. In this paper, we will use the tokenization concept already pre-trained from each transformer model we will use.

3) Text Pre-Processing

The deep learning model used will be the transformer models BERT, DEBERTa, ROBERTa, and Electra, which have been pre-trained. The dataset will be grouped into 32 batches and ten epochs with 0.00002 learning rate.

- *BERT (Bidirectional Encoder Representations from Transformers)*

BERT is a transformer model designed to pre-train deep bidirectional representations from the dataset by reading the entire sequence of text at once. In this way, the model will be able to learn the context of the word based on the left and right context. Correspondingly, the model can be fine-tuned based on the dataset provided to be used for many different tasks, such as language translation, question answering, or in our case, fake news detection [28].

- *RoBERTa (Robustly Optimized BERT Pre-training Approach)*

RoBERTa is a pre-trained encoder model that was built on BERT's language masking strategy. This model further optimizes BERT's architecture to shorten the pre-training time. RoBERTa is implemented in PyTorch, in which the main hyperparameter of BERT is modified, BERT's next-sentence pre-training objective is deleted and training is done with bigger batches and longer sequences. This enables RoBERTa to focus more on the language masking strategy objective compared to BERT [29].

- *DeBERTa (Decoding-enhanced BERT with disentangled attention)*

DeBERTa is a transformer model that further optimises both the BERT and RoBERTa model using two techniques. First, the disentangled attention mechanism is used, in which the content and position of every word are converted and stored into two vectors, respectively. Next, the attention weights are calculated using disentangled matrices based on the acquired contents and relative position. The following technique incorporates absolute positions in the decoding layer using an enhanced mask decoder. This is done to predict the masked tokens during the pre-training process. Furthermore, to improve the models' generalization, a newer virtual adversarial training method is implemented for fine-tuning [30].

- *Electra*

Google's engineers develop Electra. Electra uses a different pre-training approach that takes advantage of BERT, but more effectively [31].

IV. Result and Discussion

As mentioned before, the models used will be BERT, DEBERTa, ROBERTa, and Electra. The final model that will be implemented will be the model with the best validation accuracy from the ten epochs. The model will then be run with the dataset and compared with the actual result.

A. *BERT (Bidirectional Encoder Representations from Transformers)*

Figure 1 and Figure 2 illustrate the training loss graph and the validation accuracy graph respectively, of the dataset using the BERT model. The training loss graph indicates how well the model fits with the training data, shown by each epoch. The validation accuracy graph indicated how accurate the model is when tested with the validation data, also shown by each epoch.

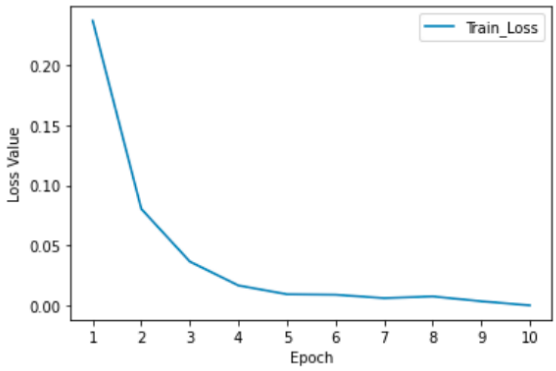


Fig. 1. Training Loss graph of BERT

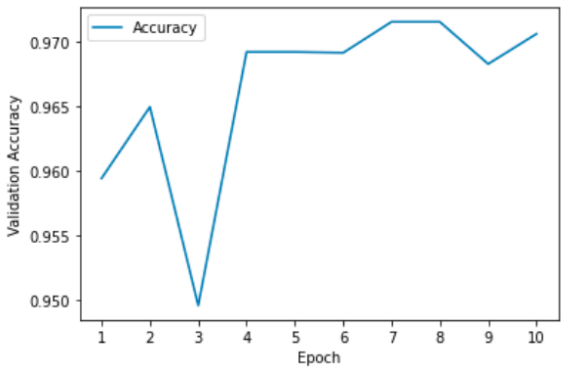


Fig. 2. Validation Accuracy graph of BERT

From Figure 2, we can infer that the best validation accuracy reaches a value of 97.15485074626866%. The test accuracy is then calculated by dividing the number of correct predictions by the total number of data, which results in 97.00826226012792%.

B. *RoBERTa (Robustly Optimized BERT Pre-training Approach)*

Figure 3 and Figure 4 define the training loss graph and the validation accuracy graph, respectively, of the dataset using the RoBERTa model. The training loss graph indicates how well the model fits with the training data, shown by each epoch. The validation accuracy graph indicated how accurate the model is when tested with the validation data, also shown by each epoch.

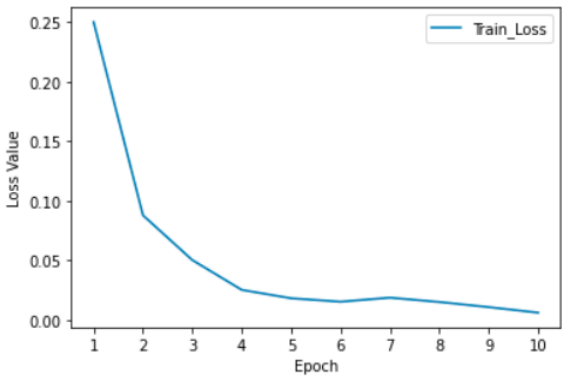


Fig. 3. Training Loss graph of RoBERTa

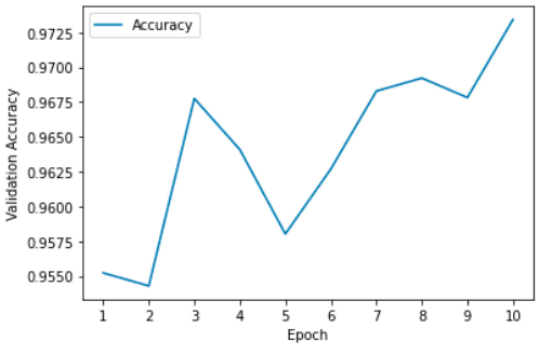


Fig. 4. Validation Accuracy graph of RoBERTa

From Figure 4, we can infer that the best validation accuracy reaches a value of 97.34141791044776%. The test accuracy is then calculated by dividing the number of correct predictions by the total number of data, which results in 97.15485074626866%.

C. *DeBERTa (Decoding-enhanced BERT with disentangled attention)*

Figure 5 and Figure 6 illustrate the training loss graph and the validation accuracy graph, respectively, of the dataset using the DeBERTa model. The training loss graph indicates how well the model fits with the training data, shown by each epoch. The validation accuracy graph indicated how accurate the model is when tested with the validation data, also shown by each epoch.

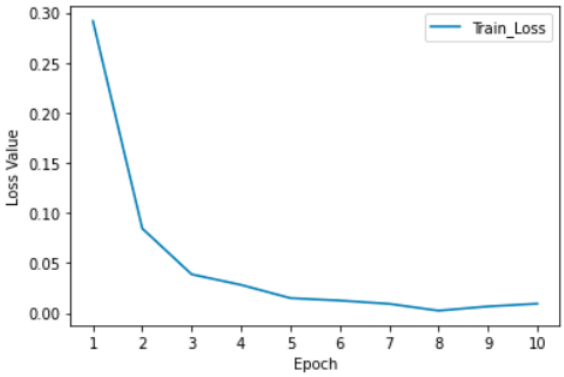


Fig. 5. Training Loss graph of DeBERTa

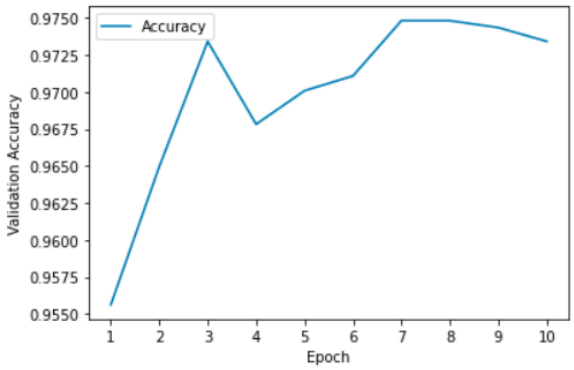


Fig. 6. Validation Accuracy graph of DeBERTa

From Figure 6, we can infer that the best validation accuracy reaches a value of 97.48134328358209%. The test accuracy is then calculated, by dividing the number of correct predictions by the total number of data, which results in 97.2481343283582%.

D. *Electra*

Figure 7 and Figure 8 illustrate the training loss graph and the validation accuracy graph respectively, of the dataset using the Electra model. The training loss graph indicates how well the model fits with the training data, shown by each epoch. The validation accuracy graph indicated how accurate the model is when tested with the validation data, also shown by each epoch.

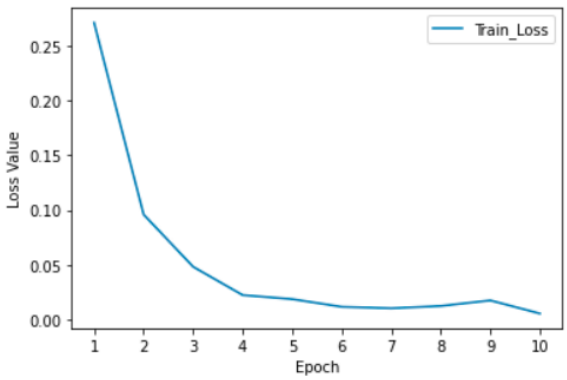


Fig. 7. Training Loss graph of Electra

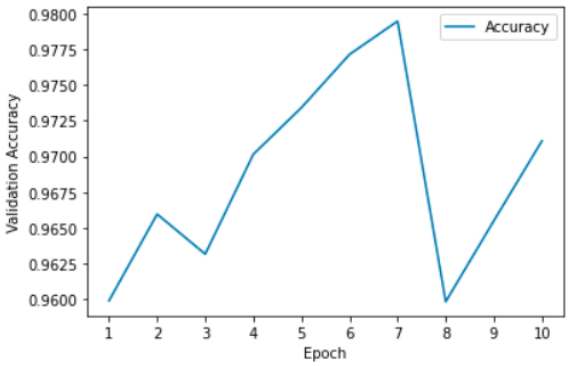


Fig. 8. Validation Accuracy graph of Electra

From Figure 8, we can infer that the best validation accuracy reaches a value of 97.94776119402985%. The test accuracy is then calculated by dividing the number of correct predictions by the total number of data, which results in 97.76119402985076%.

E. Discussion

TABLE II. VALIDATION ACCURACY AND TEST ACCURACY

Model Accuracy	Best Validation	Best Test Accuracy
BERT	97.15%	97.01%
DeBERTa	97.48%	97.25%
RoBERTa	97.34%	97.15%
Electra	97.95%	97.76%

Table II summarizes the validation accuracy and test accuracy results obtained from the experiments. From the table, we can conclude that Electra provides the best validation accuracy and test accuracy compared to BERT, DEBERTa, and ROBERTa. This is because Electra implements a pre-training task called Replaced Token Detection (RTD) which will train the model towards two outputs - real and fake. This task makes Electra a better, more suitable, and efficient model for implementing a fake news detection model. Although our experiments have shown that pre-trained transformer models are highly accurate, note that we have chosen to focus on a specific topic for hoax detection. Future researchers should aim to create a more general hoax detector system that can detect false information from any category or topic.

V. CONCLUSIONS

Of the four models, the best results on validation and test data were obtained using the Electra model with epoch 7. This model can still be improved by increasing the number of iterations on the epoch to get a more stable training loss, then validation is carried out at each epoch with cross validation. Future researchers should aim to create a more general hoax detector system that can detect false information from any category or topic.

REFERENCES

[1] "WHO Coronavirus (COVID-19) Dashboard". Covid19.who.int. https://covid19.who.int (accessed Jun. 17, 2022).

[2] J. L. Ravelo. "'Hoax killed my father': Indonesia's other pandemic". www.devex.com. https://www.devex.com/news/hoax-killed-my-father-indonesia-s-other-pandemic-100488 (accessed Jun. 17, 2022).

[3] O. D. Apuke and B. Omar. "User motivation in fake news sharing during the COVID-19 pandemic: an application of the uses and gratification theory", Online Information Review, vol. 45, no. 1, pp. 220-239, 2020, doi: 10.1108/OIR-03-2020-0116.

[4] K. S. Kalyan, A. R. and S. Sangeetha, "AMMUS : A Survey of Transformer-based Pretrained Models in Natural Language Processing", arXiv:2108.05542v2 [cs.CL], Aug. 2021.

[5] S. Malla and P. Alphonse, "Fake or real news about COVID-19? Pretrained transformer model to detect potential misleading news", The European Physical Journal Special Topics, 2022, doi : 10.1140/epjs/s11734-022-00436-6.

[6] Fardiah and F. Darmawan, "Hoax Digital Literacy on Instagram", Jurnal Komunikasi Ikatan Sarjana Komunikasi Indonesia, vol. 6, no. 2, pp. 171-186, 2021. doi : 10.25008/jkiski.v6i2.581

[7] S. van der Linden, J. Roozenbeek and J. Compton, "Inoculating Against Fake News About COVID-19", Frontiers in Psychology, vol. 11, 2020. doi : 10.3389/fpsyg.2020.566790

[8] S. Vijayaraghavan et al., "Fake News Detection with Different Models", arXiv:2003.04978v1 [cs.CL], Feb. 2020.

[9] B. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models", Procedia Computer Science, vol. 179, pp. 704-712, 2021. doi : 10.1016/j.procs.2021.01.059

[10] A. Kumar, S. Singh and G. Kaur, "Analyzing And Detecting The Fake News Using Machine Learning", International Journal of Computer Sciences and Engineering, vol. 7, no. 5, pp. 1044-1050, 2019. doi : 10.26438/ijcse/v7i5.10441050

[11] J. Nasir, O. Khan and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach", International Journal of Information Management Data Insights, vol. 1, no. 1, pp. 100007, Apr. 2021. doi : 10.1016/j.jjimei.2020.100007

[12] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake News Detection: A Deep Learning Approach," SMU Data Science Review, vol. 1, no. 3, 2018, Art. no. 10.

[13] Á. I. Rodríguez and L. L. Iglesias, "Fake News Detection Using Deep Learning", arXiv:1910.03496v2 [cs.CL], Sep. 2019.

[14] Y. Yang, et al.,"TI-CNN: Convolutional Neural Networks for Fake News Detection", arXiv:1806.00749v1 [cs.CL], Jun. 2018.

[15] J. Tembhurne, M. Almin and T. Diwan, "Mc-DNN: Fake News Detection Using Multi-Channel Deep Neural Networks", International Journal on Semantic Web and Information Systems, vol. 18, no. 1, pp. 1-20, 2022. doi : 10.4018/ijswis.295553

[16] J. Khan, M. Khondaker, S. Afroz, G. Uddin and A. Iqbal, "A benchmark study of machine learning models for online fake news detection", Machine Learning with Applications, vol. 4, pp. 100032, 2021. doi : 10.1016/j.mlwa.2021.100032

[17] D. Fox, "4 Reasons Transformer Models are Optimal for NLP". https://www.eweek.com/big-data-and-analytics/reasons-transformer-models-are-optimal-for-handling-nlp-problems/#:~:text=Transformer%20models%20are%20excellent%20at,by%20way%20of%20the%20decoder (Accessed: Jun. 19, 2022).

[18] S.M.S. Shifath., M. Faiyaz Khan and S.Md. Islam. "A transformer based approach for fighting COVID-19 fake news", arXiv:2101.12027v1 [cs.CL], Jan. 2021.

[19] H. Jwa, D. Oh, K. Park, J. Kang and H. Lim, "exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT)", Applied Sciences, vol. 9, no. 19, pp. 4062, 2019. doi : 10.3390/app9194062

[20] Y. Wang, Y. Zhang, X. Li and X. Yu, "COVID-19 Fake News Detection Using Bidirectional Encoder Representations from Transformers Based Models", arXiv:2109.14816v2 [cs.CL], Sep. 2021.

[21] P. Singh, R. Srivastava, K.P.S.Rana and V. Kumar, "SEMI-FND: Stacked Ensemble Based Multimodal Inference For Faster Fake News Detection", arXiv:2205.08159v1 [cs.CL], May. 2022.

[22] E. Shushkevich, M. Alexandrov and J. Cardiff, "Covid-19 Fake News Detection: A Survey", Computación y Sistemas, vol. 25, no. 4, Dec. 2021. doi : 10.13053/cys-25-4-4089.

[23] A. Chernyavskiy, D. Ilvovsky, and P. Nakov, "Transformers:'The End of History' for Natural Language

Processing?", in Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2021, pp. 677-693. doi: 10.1007/978-3-030-86523-8_41

[24] K. Huang, K. McKeown, P. Nakov, Y. Choi, and H. Ji, "Faking Fake News for Real Fake News Detection: Propaganda-loaded Training Data Generation", arXiv:2203.05386v1 [cs.CL], Mar. 2022.

[25] A. Pritzkau, "Multi-class fake news detection of news articles and domain identification with RoBERTa - a baseline model", in CEUR Workshop Proc., Sept. 21-24, 2021. [Online]. Available: http://ceur-ws.org/Vol-2936/paper-46.pdf

[26] S. D. Das, A. Basak, and S. Dutta, "A Heuristic-driven Uncertainty based Ensemble Framework for Fake News Detection in Tweets and News Articles", arXiv:2104.01791v2 [cs.CL], Dec. 2021.

[27] K. A. Das, A. Baruah , F. A. Barbhuiya, and K. Dey, "Ensemble of ELECTRA for Profiling Fake News Spreaders", in CEUR Workshop Proc., Sept. 22-25, 2020. [Online]. Available: http://ceur-ws.org/Vol-2696/paper_193.pdf

[28] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv:1810.04805v2 [cs.CL], May. 2019.

[29] Y. Lie, et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach", arXiv:1907.11692v1 [cs.CL], Jul. 2019.

[30] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-Enhanced BERT With Disentangled Attention", arXiv:2006.03654v6 [cs.CL], Oct. 2021.

[31] K. Clark and T. Luong. "More Efficient NLP Model Pre-training with ELECTRA". Google AI Blog. https://ai.googleblog.com/2020/03/more-efficient-nlp-model-pre-training.html (accessed Jun. 19, 2022).

ISSN 2085-4552

The Social Impact of VR Technology on Society: A Systematic Literature Review

Bakti Amirul Jabar[1], Millia Azmi[2], Graciel Hernando[3], Jurike V. Moniaga[4], Angelique Aurielle Alpaullivarez[5]

[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[1]bakti.jabar@binus.ac.id, [2]millia.azmi@binus.ac.id, [3]graciel.hernando@binus.ac.id, [4]jurike@binus.ac.id, [5]angelique.alpaullivarez@binus.ac.id

Accepted 06 September 2022
Approved 11 November 2022

***Abstract*— Humans are social creature who used to interact directly. As technology advanced, human interactions were deeply impacted. One of the many popular technologies in the 4.0 era is VR technology. An example of this technology is the metaverse, a combination of real world and virtual world. This paper presented a review to analyze the social impact of Virtual Reality technology, especially the metaverse. This paper used the systematic literature review (SLR) as the method. The research was conducted based on the RQ that had been made. There were 18 related articles generated from the initial searching attempt. This study concluded that Social VR has three impacts on society: Social Presence Illusion, Virtual Embodiment, and the Proteus Effect. In addition, this paper also explain some alternative ways to create a safe social VR environment, such as by playing with family and friends, educating the public about VR technology, and suggesting the platform to provide age limit for accessing VR environment.**

***Index Terms*— metaverse; social; technology; virtual reality**.

I. INTRODUCTION

Since the late 1970s and early 1980s, much of the tech community has envisioned a future state, if not a pseudo-successor, that is the Internet–the so-called "Metaverse" [1]. The word 'metaverse' was first coined by Neal Step-Henson. He described it as a giant virtual universe parallel to the real world. Neal took this term from the the dystopian and cyberpunk novel Snow Crash [2]. Since then, Facebook CEO Mark Zuckerberg has had a significant impact on the concept of the metaverse. And nowadays, the idea of the metaverse is very global and well known in wider community.

In this modern, digital era of 4.0, the metaverse has brought many changes in technology. Metaverse is a VR technology where a person can interact as an avatar. Virtual reality is a computer-generated reality simulating an advanced world where one can browse the artificial world virtually [2]. Social VR refers to a 3D multi-user immersive space where users are connected via a head-mounted display (HMD) and represented by a full or partial bodily-tracked virtual avatar. This is considered as the modernization of Neal Stephenson's concept of the metaverse [3]. According to Stephenson's theory, the metaverse has evolved to include affordances (i.e., what the environment offer to an individual) specific to a given system, whether through the chosen mode of communication, the promotion of group/collaborative behaviors, or the ability to design and personalize avatars [3]. A metaverse is a place for interaction between computers, computer to human and between humans. The development of this interaction technology was the development of augmented reality technology. This technology integrates the actual and virtual worlds [4]. These developments have enabled the development of new, increasingly realistic empathy technologies and experiences [5].

Many changes have brought about by this technology. From an end-user perspective, there are three main waves of innovation. The presence of personal computers, the internet, and mobile devices. Virtual and augmented reality are technologies driving today's fourth wave of computing innovation. Future waves will include Metaverse technology [6]. The VR technology has played a positive role as happened at science museums in Japan. Using VR, the museums were able to set up a 3D landscape simulation. Other example is the Dinosaur Exhibition that can show the Earth's history [7].

According to Mark Zuckerberg, the more sophisticated version of metaverse will focus on giving the users a more inclusive digital experience. Typically, two types of metaverse can be found: one that allow user to build the environment from zero and the other one that already has an embedded environment so the user can instantly browse the world. An example of the first model is NFT, a digital token that cannot be exchanged or conversed to money but can be used as proof of ownership of certain work. Examples of the second model are Fortnite and Roblox. Both Roblox and Fortnite are games that allow users to design their avatars based on their creativity.

The metaverse integrated a variety of emerging technologies. The 5G network provided ultra-high reliability and ultra-low latency connections for enormous metaverse devices, wearable sensors, and brain-computer interfaces (BCI). In general, the growth of the metaverse occurred in three sequential phases: digital twins, digital natives, and surrealist realm. In the first phase, the activity and the user are clones in which there were two parallel worlds. In the second phase, its digital content can transform and innovate in the real world. In the last stage, the virtual sphere changes, the reality will assimilate so that seamless integration will be present where many things can only be done in the virtual world [8].

This paper reviewed many fields related to VR use to enhance perspective. Moreover, many of these fields are related to each other. So, even though this paper aims to explain the impact of VR technology on social frameworks, the analysis also consists of a more general point of view. This is what differentiates this paper from other papers.

The social impact was chosen as the main topic because social frameworks have a more comprehensive approach. Moreover, it is related to many other fields that bind together as a coherent range of influence. In this sense, starting from this paper, future researchers can further develop the discussion on VR impact.

This paper was divided into several sections that aim to analyze the impact of VR technology. The theory was discussed in Section 2. Section 3 examined the systematic review of literature as the research method. In Section 4, the results of the literature review were discussed. In section 5, the conclusion was stated.

II. THEORY

A. VR Technology

Virtual reality is an activity that people can explore virtually online. VR is an artificial application created by various devices to create a real-life environment [9]. Furthermore, VR is an interactive experience wherein one can become immersed within a computer-generated environment, encompassing programs viewed on a flat screen [10]. This apps provides users with experience of altered reality in different ways by giving different stimulus to human major senses like sight and hearing. According to Ramos, virtual reality can detach experiences and interactions of the actual world or mix them with the real world, making the real and virtual worlds on a par. In recent years, the number of users has expanded significantly for this reason [11]. The purpose of Virtual Reality is to create an immersive experience using technology to modify reality. Virtual Reality display must be highly immersive, with a wide field of vision and realistic 3D visuals [12].

B. Social VR

Social VR can be defined as a user experience to socialize, interact, and prosper in a digital social arena where people meet, interact, and socialize in new and more engaging ways. The emergence of a new era of computing has been attributed to the increase in the number of commercial social VR apps, which has spawned a new research agenda in HCI and ushered in a new era of computing. Using these avatars, social VR users can enjoy real-life social activities such as wandering in public locations, playing a game, watching a movie, performing, or singing at a virtual stage, and throwing a party in highly realistic 3D replicated virtual surroundings [13]. In addition, virtual reality technology improves the creative performance of its learners, which can increase inspiration and imagination and make an immersive creative environment more conducive for creativity [20].

C. Metaverse

A metaverse can be defined as a collection of virtual reality spaces connected to a perceived virtual universe. Similar to the internet which allow people to access almost every bit of information. The word metaverse is composed of the prefix ‘meta-’, which indicates beyond, and the word ‘verse’, the final form of the word universe. Typically, the word metaverse represents a future version of the internet that consists of a shared 3D virtual world that is always connected to a perceived virtual cosmos. Metaverse, in a larger sense, may relate to the virtual world, the internet, and the full spectrum of augmented reality [1].

III. RESEARCH METHODOLOGY

A. Systematic Literature Review

Systematic Literature Review (SLR) is a study that map, identify, critically evaluate, consolidate, and collect preliminary study results on a specific research issue to solve a problem. SLR has become the usual way of finding answers by doing a literature review based on past studies relevant to the question.

SLR aims to summarize previous research, identify research gaps between previous and current studies, publish a coherent report/synthesis, and develop a research framework. This research's literature review aimed to better understand the objective, tools, and user experience of virtual reality in social life and how it affect and give the optimal solution. This study examined various journal articles retrieved from search results to achieve comprehensive conclusion.

SLR consists of multiple stages: determining research questions, selecting relevant research, collecting the necessary data, analyzing, and reporting the results.

B. *Research Question*

The following table contain the research questions that has been carried out on this paper. Table I showed three questions that is the main focus of this paper.

TABLE I. RESEARCH QUESTION

ID	Research Question	Motivation
RQ1	How do people use social VR?	To analyze how people use social VR
RQ2	How does the use of social VR influence people's social life?	To identify if social VR influence the social lives of people
RQ3	What is the way to create a safe and comfortable social VR environment?	To explain what advice can be given for comfortable and safe environment to play VR technology

C. *Work Procedure*

Work Procedure includes literature search, selection, documentation, analysis, and drawing conclusions. According to Creswell, there are six steps in the procedure of Literature Review.

First, identifying key terms. Searching on journals or literature using keywords. The keywords used in this paper was: VR, social technology and metaverse. Second, determining the origin of literature (local literature) following the topics that have been found from the database or the internet to collect relevant journals or literature. Journals or literature in this study were obtained by accessing them online. Third, critically evaluating and selecting the literature for review. There are 48 works of literature that have been searched and collected through the internet. Following initial screening, only 18 journals were selected and will be reviewed. Fourth, organizing the literature that has been chosen. Information materials and data from previous researchers have been obtained, read, recorded, and reprocessed. At this stage, information will be collected regarding the topics discussed from 18 literatures that have been evaluated and read. Fifth, writing the literature review. After gathering information on the literature, a summary will be compiled regarding the topics discussed to answer the research questions.

And then the last, making results and conclusions. The results were produced to answer the information collected and have been analyzed to answer research questions. After the results have been made, then at the end, the conclusions will be drawn. Figure 1 is the flowchart of this paper's work procedure structure. There were sequential stages from the beginning to the end of this paper.

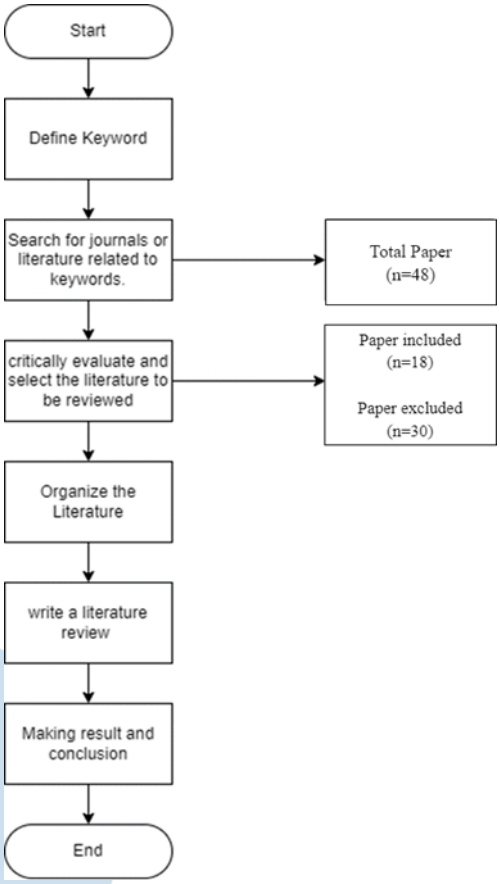


Fig. 1. Flowchart Work Procedure

IV. LITERATURE REVIEW

A. *The Impact of Metaverse in General*

The impacts and changes effected by the metaverse technology are manifold. Some have a negative impact. For example, in the environmental field, in the journal Metaverse Future Addiction Concerned for Human-Being (2022), it was stated that metaverse can harm the ecosystem [1]. Some say that the metaverse trend of implementing empathy tools has a limited effect on users [5]. In the field of psychology, where the governance model and privacy security were not handled properly, it will lead to a lifestyle that is full of worries and vigilance of others [12]. However, there are also positive impacts. For example, in the social field, in research conducted by Divine Maloney (2021), a metaverse with high level of interactivity can act as a tool for social development [3]. In the journal Metaverse: Future of the Internet (2022), it was also stated that with the metaverse, we can understand each other and communicate better [13]. From a technology perspective, the metaverse offers a great opportunity to use a human-centered approach to create additional high-value applications [14]. This technology can be used for speculation that is not yet possible [2]. Augmented reality technology can help humans solve social problems that is currently being studied, especially in terms of interaction [4].

Metaverse technology continues to develop with more sophisticated technology and creates dimensional virtual reality [15] by linking various factors that can affect Metaverse technology with its users [16]. Since the COVID-19 pandemic, it has affected the lives of multiple sectors, from Education to Industry, creating the concept of work and distance learning. Metaverse provides a VR/AR concept that simplifies remote concepts [17]. Metaverse has links in various industrial fields, including education and health. In the industrial sector, the metaverse can be used in making business strategies in the form of dimensions [18]. In the health sector, metaverse can create a virtual social reality for people with disabilities [19].

In education, metaverse technology functions as a central teaching system by examining the relationship between the surrounding environment and students' lives [6]. By changing the traditional (teacher-centered) learning system to VR (machine-centered) technology, many factors are involved in the transfer of the system, such as mindset and technology mastery [20]. It can be started by collaborating between virtual concepts and conventional learning, [21]. The creation of the metaverse, which the creators of Facebook started, still has shortcomings, the data used in user identification for VR/XR is insufficient for people with disabilities [7].

Regardless of the metaverse's sophistication and positive impact, there will certainly be a downside and security threats. There are seven dimensions to classify security threats in the metaverse: identity, data, privacy, network, economy, physical/social effects, and governance. In the identity-related dimension, there are threats of identity theft, impersonation attacks, and avatar authentication issue. In the data-related dimension, there are threats of data tampering attacks, false data injection attacks, issues in managing new data types, threats to data quality, and threats to gun ownership. On the privacy dimension, there are pervasive data collection, privacy leaks, compromised end devices, unauthorized data access, misuse of avatar data, and accountability. There are also SPOF, DDoS, and Sybil attacks on the network dimensions. On the economic-related dimension, there are issues of trust, threats to digital asset ownership, and threats to economic justice. On the dimension of physical/social effects, there are threats to infrastructure security and social impact. On the governance-related dimension, there are new laws for virtual crime, bad regulators, and threats to collaborative governance [8]. Table II show the summary of this section. Table II has several categories of metaverse benefits with their references.

TABLE II. LIST OF METAVERSE IMPACT IN MANY FIELDS

No	Impact	References
1	Psychology	[2]
2	Ecosystem	[3]
3	Empathy	[6]
4	Social	[1], [2], [4], [5], [8], [9], [13], [16]
5	Education	[7], [15]
6	Law and privacy	[10]
7	Technology	[14], [11]
8	Health	[16], [18]

B. Online Social VR Experience with Metaverse

Many studies have investigated how society utilize and experience metaverse, particularly among young people. Habbo Hotel, one of the largest social virtual environments for adolescents, is well-known for three qualities. The ability to co-create a social setting comes first. Second, utilizing the virtual world as a location for digital artifacts, such as avatars, goods, and experiences such as escapism, voyeurism, and self-image. Thirdly, anonymity and security in metaverse permit freedom of expression and experimentation with various avatar skins and identity-based social events [3].

This feature suggests that young people perceive certain aspects of the virtual social world to be less risky than the offline world. They primarily use this environment for social experimentation, creating exciting challenges and opportunities around identity construction because, for most young people, engaging in a virtual world may become their first experience to be truly free on expressing their identity. This process allows them to construct, reconstruct, and comprehend themselves and others in the online world. This process of self-discovery may not always be beneficial because sometimes undesired interactions such as harassment, cyberbullying, and sexual harassment also occurred in this virtual realm. This demand for higher security stringency [3].

C. Social Effect in Metaverse

There are three social effects that occur in the metaverse:

1) Social Presence Illusion

Experiences in social VR can be described as having a real sensations, similar to offline sensations with other people. This sensation can be referred to as social presence, which generally refers to feeling connected to other people [3]. Additionally, VR gives the users access to real-time experiences in a 3D world while allowing them to participate in interactive situations that give them the sensation of being in another place [21]. In the user's mind, there will be an illusionary sensation of being in another world. The illusion makes a person think he is somewhere else even though he is still there.

2) Virtual Embodiment

Virtual embodiment, often known as "presence self," refers to the user's representation in a virtual world. Ben-ford et al. describe embodiment as

providing users with a suitable body image to represent them to others (and themselves) in collaborative contexts. The illusion of body possession, which may be described as a method in which the user executes a movement to support their perception of ownership of a virtual creature, can be used to induce an embodiment [3]. A virtual embodiment is a feeling when playing a virtual game where one feel one's body is playing an avatar like in the real world where they can move their body as their wishes.

3) Proteus Effect

Many works have demonstrated the effect of virtual avatars on a person's perception, behavior, and cognition. This is called the Proteus effect, which describes how the characteristics of their avatars alter individuals' behavior in the virtual world. For instance, inhabiting a juvenile avatar may induce immature behavior and perceptual shifts in the user [3]. The nature and personality of a person when playing sometimes cause an effect in the form of a character that is different from the real world and created new traits both in the virtual and real world.

D. Create a Safe Social VR Environment

There are many worrying issues in the VR world. The VR environment may bring about a person's social negativity instead of fixing it. Therefore, to avoid this, we can create a safe environment in the following ways:

1) Experience Social VR With Loved Ones and Friends

Mentally unstable minors are better off playing with their closest friends or family to avoid the bad effects of playing with strangers, such as bullying, bad attitude, or mental down. Playing with strangers can also lead to feelings of discomfort and uneasy. For example, games such as PUBG, when played with strangers sometimes cause alertness and fear compared to playing with friends or family.

2) Educate People about VR Technology

Educate the people about Virtual Reality Digital Literacy. A more comprehensive and proper education on social VR and broader immersive technologies is required to provide a secure online social space for the general public, particularly for minors. There are two ways to educate people about technology; directly and indirectly. In a direct way, educator can held an event that talk about VR world routinely. They can also implement tools while on it. Meanwhile in an indirect way (online), educator can build a website that can be used by all ages that contain a complete information about Virtual Reality and provide space for discussion because the more people know about technology the more they will be able to use it effectively for their benefit.

3) Grouping Independently in The Environment

This medium means that the platform must support the age-centered design and practical activities that focus on the target age group because everyone is different. For example, like Fortnite, age restrictions are made to avoid improper socialization, such as harsh words, swearing, violence and sexual harassment. In addition, games can influence emotions and behavior in children's brain development.

V. CONCLUSIONS

The use of VR Metaverse has a social impact on society. Bad things such as bullying sometimes still happen in society. However, playing VR gradually can also cause social effects such as Social Presence Illusion, virtual embodiment, and the Proteus effect. Therefore, we must prevent this and create a safe environment in the VR world. Our study provides solutions such as building advance security, playing with loved ones or friends, and educating the public more about VR technology. We hope there will be a safe environment for social VR that does not have bad effect on society.

REFERENCES

[1] P. Cipresso, I. A. C. Giglioli, M. A. Raya, and G. Riva, "The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature," *Front Psychol*, p. 2086, 2018.

[2] U. v Ramesh, A. Harini, C. S. D. Gowri, K. V. Durga, P. Druvitha, and K. S. Kumar, "Metaverse: Future of the Internet," *Journal homepage: www.ijrpr.com ISSN*, vol. 2582, p. 7421. 2022.

[3] A. Kemec, "From Reality to Virtuality: Re-discussing Cities with the Concept of the Metaverse," *International Journal of Management and Accounting*, 4(1), 12-20, 2022.

[4] D. Maloney, "A Youthful Metaverse: Towards Designing Safe, Equitable, and Emotionally Fulfilling Social Virtual Reality Spaces for Younger Users," 2021.

[5] K. Teguh Martono, "Augmented Reality sebagai Metafora Baru dalam Teknologi Interaksi Manusia dan Komputer," *Jurnal sistem komputer*, vol. 1, no. 2, pp. 60–64, 2011.

[6] V. Paananen, M. S. Kiarostami, L.-H. Lee, T. Braud, and S. Hosio, "From Digital Media to Empathic Reality: A Systematic Review of Empathy Research in Extended Reality Environments," *arXiv preprint arXiv:2203.01375*, 2022.

[7] W. Suh and S. Ahn, "Utilizing the Metaverse for Learner-Centered Constructivist Education in the Post-Pandemic Era: An Analysis of Elementary School Students," *J Intell*, vol. 10, no. 1, p. 17, 2022.

[8] B. Egliston and M. Carter, "Critical questions for Facebook's virtual reality: data, power and the metaverse," *Internet Policy Review*, vol. 10, no. 4, 2021.

[9] Y. Wang *et al.*, "A survey on metaverse: Fundamentals, security, and privacy," *arXiv preprint arXiv:2203.02662*, 2022.

[10] N. Xi, J. Chen, F. Gama, M. Riar, and J. Hamari, "The challenges of entering the metaverse: An experiment on the effect of extended reality on workload," *Information Systems Frontiers*, pp. 1–22, 2022.

[11] P. Yagol, F. Ramos, S. Trilles, J. Torres-Sospedra, and F. J. Perales, "New trends in using augmented reality apps for smart city contexts," *ISPRS International Journal of Geo-Information*, vol. 7, no. 12, p. 478, 2018.

[12] M. Dutilleux and K.-M. Chang, "Future Addiction Concerned for Human-Being," *International Multilingual Journal of Science and Technology (IMJST)*, vol. 7 issue 2, 2022.

[13] R. Leenes, "Privacy in the Metaverse," in *IFIP International Summer School on the Future of Identity in the Information Society*, 2007, pp. 95–112.
[14] L.-H. Lee *et al.*, "All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda," *arXiv preprint arXiv:2110.05352*, 2021.
[15] A. M. Aburbeian, A. Y. Owda, and M. Owda, "A Technology Acceptance Model Survey of the Metaverse Prospects," *AI*, vol. 3, no. 2, pp. 285–302, 2022.
[16] M. Damar, "Metaverse Shape of Your Life for Future: A bibliometric snapshot," *Journal of Metaverse*, vol. 1, no. 1, pp. 1–8, 2021.
[17] J. H. Murray, "Virtual/reality: how to tell the difference," *Journal of Visual Culture*, vol. 19, no. 1, pp. 11–27, 2020.
[18] L. Appel *et al.*, "Virtual reality to promote wellbeing in persons with dementia: A scoping review," *Journal of Rehabilitation and Assistive Technologies Engineering*, vol. 8, pp. 20556683211053950–20556683211053950, 2021.
[19] J. Chen, C. Liu, R. Chang, P. Gui, and S. Na, "From traditional to VR-based online education platforms: A model of the mechanism influencing user migration," *Information*, vol. 11, no. 9, p. 423, 2020.
[20] S. Suzuki *et al.*, "Virtual Experiments in Metaverse and their Applications to Collaborative Projects: The framework and its significance," *Procedia Computer Science*, vol. 176, pp. 2125–2132, 2020.

ISSN 2085-4552

Design and Build a Website for Catering Sales and Orders Using Web Engineering

Rd. Erwin Gunadhi, M.T.[1], M. Rikza Nasrulloh, M.Kom.[2], Epril Mohamad Rizaludin[3]
[1,2,3] Program Studi Teknik Informatika, Jurusan Ilmu Komputer, Institut Teknologi Garut, Garut, Indonesia
[1]erwingunadhi@itg.ac.id, [2]rikza@itg.ac.id, [3]1806072@itg.ac.id

Accepted 20 September 2022
Approved 12 October 2022

***Abstract*— Ordering catering is a consumer activity to order products in the form of food packages. At this time, catering orders in the Garut area are still done manually. Consumers see a list of products through brochures, even though consumers do not necessarily know the shape of the product written in the brochure, it will make it difficult for consumers to determine which product to choose, and these difficulties will make consumers have to meet with the owner of the catering frequently which will hinder time and drain costs. The existence of technology encourages researchers to build a catering ordering website that is expected to facilitate consumers in the catering ordering process. Consumers can find out the list and pictures of the products provided, which can be facilitated in determining the product to be selected. Consumers can also save time and costs because there is no need to repeatedly meet with the catering owner to find out the products' details. The method used in the development of this application is the Web Engineering method with the stages of communication, planning, modeling, and deployment, using the PHP programming language, HTML, using the Codeigniter framework, and bootstrap and with MySql database management by testing the Black Box Testing application. The research results are in the form of a catering sales and ordering website, which can help consumers find out information about the products provided by the catering party, including prices, descriptions, and pictures. You can also order catering products for certain activities or events, and this website can manage the incoming products and orders.**

***Index Terms*— Catering; Codeigniter; PHP; Product; Website**.

I. INTRODUCTION

The business in the catering sector is currently growing. In the Garut area, for example, it was found that several catering entrepreneurs who previously worked in one of the leading caterers opened their catering businesses. Catering is a service sector business that provides or serves food orders for various purposes according to customer demand [1].

Ordering catering is a consumer activity to order products in the form of food packages. At this time, catering orders in the Garut area are still done manually. To order catering, consumers first see a list of products and packages through a brochure provided by the catering owner. This method must be more straightforward for consumers to determine which product to choose. The limitations of the images in the brochure can make consumers not know how the form of the product list written on the brochure. The limitations of the price list contained in the brochure can also make it difficult for consumers to choose what products to take. Consumers will need help to compare prices between products the catering party provides. This ignorance makes consumers frequently meet or communicate with catering owners to ask questions about the shape of the products in the brochure. This is quite time-consuming for consumers who have to prepare various things for the continuity of an event. In addition, meeting with catering owners can also cost money for some consumers who come from outside areas far from city roads. The distance traveled by consumers who are far outside the area makes consumers need to incur transportation costs or more costs to meet the catering owner.

The existence of technology encourages researchers to build a catering ordering website that is expected to facilitate consumers in the catering ordering process. The method used in the development of this application is the Web Engineering method, using the PHP programming language, HTML, using the Codeigniter framework, and bootstrap and MySql database management by testing the Black Box Testing application.

Researchers refer to several research journals about online catering services that existed previously, namely research conducted by [2]. This research aims to expand the sales information network and introduce Ummi Nisa's catering by creating a website. The second study was conducted by [3]. This study aims to find out the flow of catering orders at Cimahi catering at this time and to build a catering application based on Android Mobile at Cimahi catering. The third research was conducted by [4]. This study aims to expand the sales information network and several menu packages offered by Ratu Catering. The catering ordering system is expected to facilitate the ordering process by Riau Catering and consumers. The fourth research was conducted by [5]. This study aims to facilitate the recording and reporting of the catering business. The

fifth study was conducted by [1]. This research aims to make the work of catering management more effective and efficient. The sixth study was conducted by [6]. This research aims to create software selling typical fabrics that can increase the income of Palembang typical fabric artisans and promote Palembang typical fabrics in the eyes of the world. The seventh study was conducted by [7]. This research aims to provide information about culinary tourism starting from culinary information, the first location for tourists to access the application equipped with a route to culinary places and to help culinary business activists as a promotional medium for their culinary companies.

II. METHODOLOGY

This research has a framework of thought which is the activity stage in the application design process which is presented in the form of a diagram so that it is easy to read and understand the direction of the flow of this research. The following are the stages of the application design process activity in this study:

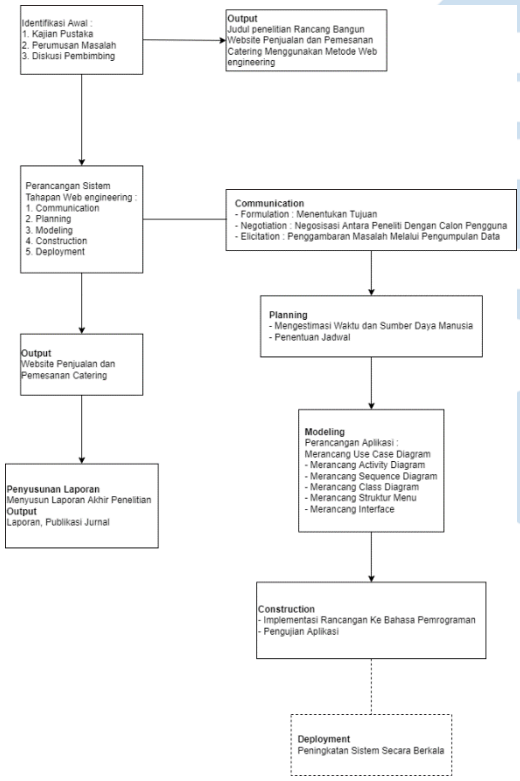


Fig. 1. Thinking Framework

The framework is a stage of activity in the application design process which is presented in the form of a diagram, so that it is easy to read and understand the direction of the flow of this research. The following is an explanation of the framework of Figure 1:

The first stages of research activities are as follows:

a. Literature review, at this stage the activity carried out is the collection of previous research journals related to the topic under study.

b. Problem formulation and data collection, at this stage the activity carried out is to formulate the problems found in the field based on a questionnaire.

c. Supervising discussion, at this stage the activity carried out is a discussion to determine the topic or method to be taken and collect data obtained from the questionnaire.

The output of the first stage is the research title, namely Design and Build a Website for Sales and Catering Orders Using the Web Engineering Method.

The second stage is application design using web engineering methods with the stages namely Communication, Planning, Modeling Construction, and Deployment:

a. Communication, at this stage the activity carried out is to determine the purpose of the application design and determine who will use the application. Furthermore, negotiations between developers and potential users, and describe the problem from the research data.

b. Planning, at this stage the activities carried out are estimating the time for application design and determining human resource requirements.

c. Modeling, at this stage the activities carried out are designing use case diagrams, activity diagrams, sequence diagrams, class diagrams, creating menu structures, and designing user interfaces.

d. Construction, at this stage the activities carried out are building applications or implementing designs into programming code and testing applications.

e. Deployment, at this stage the activities carried out are running applications that have been built.

The output of this second stage is a website for sales and catering orders.

The last stage is the preparation of the final report of the research, with the output at this stage is reports and publications of journals and posters.

As for database on websites designed using MySQL,. MySQL is a DBMS (Database Management System) using SQL (Structured Query Language) commands that are widely used today in making web-based applications [16].

III. RESULTS AND DISCUSSION

A. Communication

In the first stage, problems were found in the catering ordering business process.

1. Literature review, based on reference research, the process of ordering catering manually can consume customers' time in placing orders.
2. Questionnaire, after the assessment activity in the reference journal, then proceed with data collection through questionnaires or questionnaires in order to strengthen the source of the research conducted. In this study, there were 30 respondents from several areas with an average age of 40 years who were experienced in catering orders with details of the results 70% of customers knew about products and prices through brochures and 30% of customers found out from the catering owner. As many as 43% of customers immediately know the product variants in the brochure and 57% who cannot know it immediately. As many as 36% of customers feel that brochures are very easy to choose products and 64% consider brochures to be less helpful. As many as 50% of customers feel they are a burden by frequently meeting the catering owner and 50% do not make it a burden. As many as 80% of customers or catering owners often come to each other's places and 20% do not often have meetings. As many as 70% of customers feel that meetings are a time constraint and 30% are not time constrained. As many as 63% of customers find meetings to be cost-effective and 37% do not find them expensive. As many as 60% of customers feel that distance from the catering owner is an obstacle and 40% do not consider distance as an obstacle. Then 83% of customers agree with the catering booking website and 17% disagree with the website.

Based on a literature review and research questionnaires, it is found that the purpose of website development and actors who will use this catering sales and ordering website are obtained. The purpose of the development of this website is to build a catering booking website so that consumers can find out information about the products provided by the catering party which includes prices, descriptions, and pictures, and consumers can place orders for products, so that consumers do not have to often meet the catering owner in placing orders. Actors who will use this website are customers and catering owners.

B. *Planning*

At this stage the activities carried out are estimating the time for web design and determining human resource needs. This activity produces an estimated web design time table which is shown in Figure 2.

TABLE I. ESTIMATED WEB DESIGN TIME

No	Kegaiatan / Aktivitas	Februari				Maret				April				Mei				Juni				Juli				Agustus			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Perancangan Aplikasi																												
	a. Communication																												
	1) Identifikasi Masalah																												
	2) Menentukan Tujuan																												
	3) Menentukan Aktor																												
	b. Planning																												
	1) Estimasi Waktu																												
	2) Estimasi Biaya																												
	3) Menentukan Sumber Daya Manusia																												
	c. Modeling																												
	1) Merancang Use Case Diagram																												
	2) Merancang Activity Diagram																												
	3) Merancang Sequence Diagram																												
	4) Merancang Class Diagram																												
	5) Merancang Struktur Menu																												
	6) Merancang User Interface																												
	d. Construction																												
	1) Implementasi Ke Bahasa Pemrograman																												
	2) Pengujian Aplikasi																												
	e. Deployment																												
	1) Menjalankan Apliaksi																												

The communication stage is estimated to last for three weeks, the planning stage is estimated to last three weeks, the modeling stage is estimated to last one month and two weeks, the construction stage is estimated to last one month and one week, and report preparation is estimated to last three months.

The human resources in this study are:

1. Rd. Erwin Gunadhi R, M.T. as the first guide
2. M. Rikza Nasrulloh, M.Kom. as the second guide
3. Epril Mohamad Rizaludin as a researcher
4. Respondent.

C. *Modeling*

At this stage the activity carried out is the design of UML (Unified Modeling Language) to create an initial picture of the system based on the problems and results of data collection that has been done previously. The system depiction is implemented into use case diagrams, activity diagrams, sequence diagrams, class diagrams, and user interface designs.

1) *Use Case Diagram*

Tokenization is an activity of splitting an entire text into small units, also known as tokens. In this paper, we will use the tokenization concept already pre-trained from each transformer model we will use.

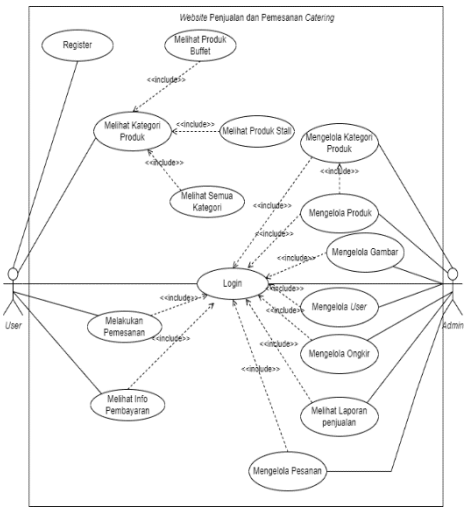


Fig. 2. Use Case Diagram

Use Case Diagram, used to find out what functions exist in the system and who can use these functions can be seen in Figure 2.

2) Activity Diagram

Activity diagrams explain the interaction between the user and the system to model the system according to the use case. The login activity diagram can be seen in Figure 3.

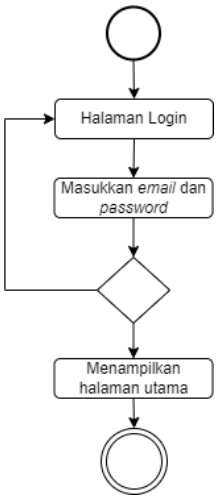


Fig. 3. Activity Diagram Login

3) Sequence Diagram

Sequence diagrams describe the interaction of objects in the use case by describing the life time of the object with messages sent and received between objects. The login sequence diagram can be seen in Figure 4.

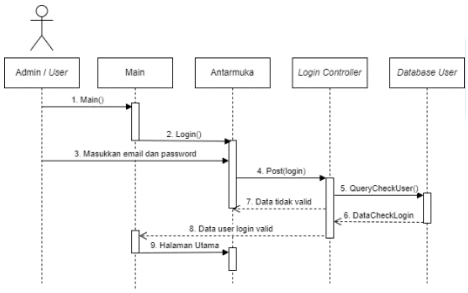


Fig. 4. Sequence Diagram Login

4) Class Diagram

Class diagrams are used to display several classes that exist in the software system to be developed. Class diagrams show the relationships between classes in the system and how they relate to each other to achieve a goal. Figure 5 is a class diagram of a catering sales and ordering website.

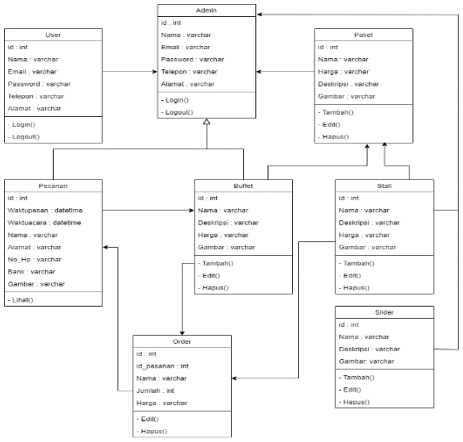


Fig. 5. Class Diagram

D. Construction

At this stage is the implementation of the design into a programming language and web testing.

1) Design Implementation

a) Register View

To place an order, the user must first create an account via the register page. In registering, the user fills in his full name, email, phone number, address, and password. On this page there is also a button in the form of text that can be selected if the user already has an account. The display of the register can be seen in Figure 6.

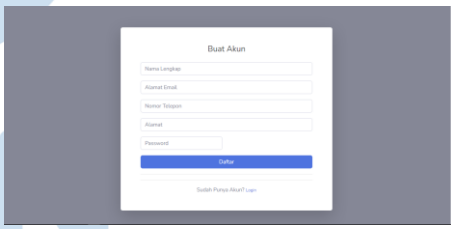


Fig. 6. Register View

b) Login View

After the user has an account, to be able to access the web and be able to place an order, the user must login by filling out the username and password form. On this page there is also a button in the form of text that can be selected if the user does not have an account or to create an account. The login display can be seen in Figure 7.

Fig. 7. Login View

c) Home View

This page is the main page of the web catering sales and reservations. This page contains a navbar menu that is interconnected with their respective pages, landing pages and several products with package categories. The home display can be seen in Figure 8.

Fig. 8. Home View

d) Product List View

This page contains a list of products provided by catering. The product includes the name, price, picture and product description. On this page the user can add products to the cart and can also view product details by selecting the detail button. Product display can be seen in Figure 9.

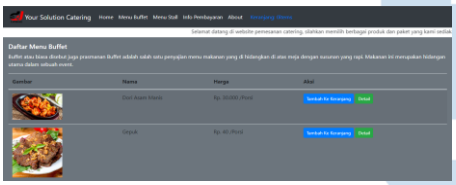


Fig. 9. Product List View

e) Cart View

This page contains the products that the user has ordered. The user can delete the filled cart, continue shopping and select checkout, when the user selects the checkout button, it will then go to the checkout page. The cart display can be seen in Figure 10.

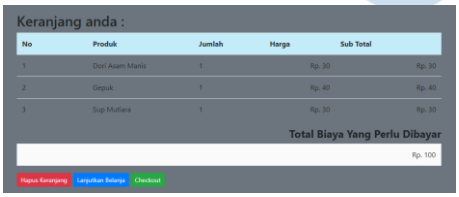


Fig. 10. Cart View

f) Checkout View

This page contains a bill or order form, the user fills in the order time, event time, name, phone number, address, and bank account. The checkout display can be seen in Figure 11.

Fig. 11. Checkout View

g) Dashboard Admin View

This page is a page that only admins can access. On this page the admin can manage products such as adding, changing, or, deleting products. Admin also accepts orders from users through this page. This page consists of user data, category data, buffet menu data, stall menu data, sales reports, orders, and pictures. The admin dashboard view can be seen in Figure 12.

Fig. 12. Dashboard Admin View

2) Black Box Testing

Blackbox testing tests based on existing activities on the web and this stage serves to find out whether the activities on the web can run according to their functions without any errors occurring. The test results can be seen in Table 2.

TABLE II. BLACK BOX TESTING

Activity	Test Class	Expected Results	Conclusion
Register	Register	Create an account by filling in your full name, email, phone number, address, and password.	In accordance
Login	Login	Access the website using the registered username and password through the register	In accordance
Showing Product	Product display by category	Product display by category	In accordance
	Product detail view	Displays the details of the product selected by the user	In accordance
Product Order	Add product to cart	Add ordered products to cart.	In accordance
	Show Cart Page	Displays a cart page with product order	In accordance

Activity	Test Class	Expected Results	Conclusion
		data that has been ordered	
	Showing Checkout Page	Showing checkout page	In accordance
	Uploading Proof of Payment	Save the image file uploaded by the user into the database	In accordance
Category Management	Showing Category	Showing Category data	In accordance
	Add Category	Add category data	In accordance
	Edit Category	Edit selected category data	In accordance
	Delete Category	Delete the selected category	In accordance
	Product Show	Display product data	In accordance
	Add Product	Add product data	In accordance
	Edit Product	Edit the selected product data	In accordance
	Delete Product	Delete the selected product data	In accordance
	Showing Image	Showing image data	In accordance
	Add Image	Add image data	In accordance
	Edit Image	Edit the selected image data	In accordance
	Delete Image	Delete the selected image data	In accordance
	Show Orders	Display order data	In accordance
	Show Order Details	Show order details	In accordance
	Edit Order	Edit the selected order data	In accordance
	Delete Order	Delete the selected order	In accordance

E. Deployment

This stage is running the application by the user, the following are the results of the analysis obtained after the user runs the website for sales and catering orders. This stage is carried out by consumers using the web and then filling out a questionnaire containing questions about the web. This test was carried out on 10 consumers with the result that 100% of users can find out the shape of the product through images available on the web, 100% of consumers know product details through detailed features and 100% of consumers can find out product prices when choosing their own products to order. In addition, 100% of consumers felt that they did not need to go to a catering place to place an order and 97% of consumers felt that they did not need to often ask about product details to the catering owner, 95% of consumers found it easy to use the web and 5% found it difficult to use it.

F. Discussion

The Table 3 is an analysis of the results of the discussion with previous similar studies:

TABLE III. ANALYSIS OF RESEARCH RESULTS WITH PREVIOUS RESEARCH

No	Research Title	Research Scope	Discrepancy
1	Sistem Informasi Pemesanan Catering Pada Maya Berbasis Java	This study discusses the flow of catering orders at Cimahi Catering.	The method used in previous research is a quantitative approach with the Java programming language, while the current research uses a web engineering method using the PHP programming language.
2	Pengolahan Bisnis Catering Ummi Nisa Medan Berbasis Web	This study discusses Ummi Nisa's catering business in Medan	The method used in previous research is data analysis and system design analysis, while the current research uses web engineering methods
3	Perancangan Aplikasi Pemesanan Catering Berbasis Mobile Android	This research focuses on designing a catering ordering application	The method used in the previous research was the extreme programming method, while the current research uses the web engineering method
4	Sistem Aplikasi Pemesanan Catering Pada Ibu Holipah	This research only focuses on Mrs. Holipah's catering business	The method used in previous research is the grounded method (grounded research) using the Java programming language, while the current research uses the web engineering method with the PHP programming language programming language.
5	Sistem Informasi Catering di Ratu Catering Berbasis Web	This study discusses the flow of catering orders at Ratu Catering	The method used in previous research is the SDLC waterfall method, while the current research is a web engineering method
6	Penggunaan Metode Web Enginering Dalam Aplikasi Penjualan Kain Khas Palembang	This study discusses the sale of typical Palembang fabrics	The application that was built in previous research is a cloth sales application, while the current research is building a catering sales application
7	Rancang Bangun Aplikasi Web Katalog Produk Wisata Kuliner Berbasis	This research focuses on mapping culinary tourism	Previous research discussed culinary tourism mapping, while research

IV. CONCLUSIONS

Based on the results of the research, it can be concluded that with a website that is designed and built, consumers can find out information about the products provided by the catering party, including prices, descriptions, and pictures. This certainly overcomes the problems related to the limitations mentioned in the introduction to this study. Not only that, consumers can order catering products for certain activities or events, and this website can manage incoming products and orders.

REFERENCES

[1] M. N. Anugrah, B. Dwi Hatmoko, and V. Ramdhan, "SISTEM INFORMASI PEMESANAN CATERING PADA MAYA CATERING BERBASIS JAVA," Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI), vol. 03, 2022.

[2] Khairunnisa and F. Damayanti, "Jurnal Pengolahan Bisnis Catering Ummi Nisa Medan Berbasis Web," Jurnal Sistem Informasi, vol. 02, pp. 63–71, 2018.

[3] M. Syani and N. Werstantia, "PERANCANGAN APLIKASI PEMESANAN CATERING BERBASIS MOBILE ANDROID," Jurnal Ilmiah Ilmu dan Teknologi Rekayasa |, vol. 1, no. 2, 2018.

[4] N. Siregar, Zamzami, and Guntoro, "Sistem Informasi Catering di Ratu Catering Berbasis Web," Jurnal Sistem Informasi, vol. 1, no. 2, pp. 104–114, 2019.

[5] A. Kurniawan, R. Nuzulah, and A. Saputra, "SISTEM APLIKASI PEMESANAN CATERING PADA IBU HOLIPAH," Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI), vol. 02, 2021.

[6] N. Sopiah, E. Puji Agustina, and U. Bina Darma, "PENGGUNAAN METODE WEB ENGINEERING DALAM APLIKASI PENJUALAN KAIN KHAS PALEMBANG."

[7] Y. Septiana, R. Erwin, G. Rahayu, and M. S. Aghna, "Rancang Bangun Aplikasi Web Katalog Produk Wisata Kuliner Berbasis Geographic Information System", [Online]. Available: http://jurnal.sttgarut.ac.id/

[8] A. N. Nurhayati, A. Josi, and N. A. Hutagalung, "RANCANG BANGUN APLIKASI PENJUALAN DAN PEMBELIAN BARANG PADA KOPERASI KARTIKA SAMARA GRAWIRA PRABUMULIH".

[9] A. Susanto and Asmira, "Perancangan Website Sebagai Media Promosi dan Informasi Menggunakan Metode Web Engineering," SIMKOM, vol. 2, no. 3, 2017, [Online]. Available: http://e-jurnal.stmikbinsa.ac.id/simkom

[10] R. S. Pressman and D. B. Lowe, Web engineering : a practitioner's approach. McGraw-Hill Higher Education, 2009.

[11] S. Mulyati and M. Hisyam, "RANCANG BANGUN SISTEM INFORMASI PENYEWAAN WEDDING ORGANIZER BERBASIS WEB DENGAN PHP DAN MYSQL PADA KIKI RIAS," Jurnal Teknik: Universitas Muhammadiyah Tangerang, vol. Vol. 7, No. 2, pp. 29–35, 2018.

[12] L. Setiyani, "Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi PENGUJIAN SISTEM INFORMASI INVENTORY PADA PERUSAHAAN DISTRIBUTOR FARMASI MENGGUNAKAN METODE BLACK BOX TESTING," 2019.

[13] A. F. Sallaby and I. Kanedi, "Perancangan Sistem Informasi Jadwal Dokter Menggunakan Framework Codeigniter."

[14] A. Christian, S. Hesinto, J. Patra No, K. Sukaraja Kecamatan Prabumulih Selatan, and S. Selatan STMIK Prabumulih, "Rancang Bangun Website Sekolah Dengan Menggunakan Framework Bootstrap (Studi Kasus SMP Negeri 6 Prabumulih)," 2018.

[15] D. Pradiatiningtyas and Suparwanto, "E-Learning Sebagai Media Pembelajaran Berbasis Web Pada Smk N 4 Purworejo," Ijns.org Indonesian Journal on Networking and Security, vol. 7, no. 2, pp. 2302–5700, 2017.

[16] Rizqulloh, S. R., Hadikusuma, R. S., Aulia, N., & Hidayat, R. (2022). Automatic Sluice Monitoring Based On The Water Ph In a Brackish Water Pool Using a Web Server. Jurnal Elektro Teknik, 1(1), 22-29. 2022.

ISSN 2085-4552

Automatic Portal Access Application Using Static Qr Code Reading

Suhartono[1], Nurhani[2], Satria Gunawan Zain[3]
[1,2,3] Informatics & Computer Engineering, Makassar State University, Makassar, Indonesia
[1]suhartono@unm.ac.id, [2]nurhaniana99@gmail.com, [3]sg.zain@gmail.com

Accepted 02 October 2022
Approved 31 October 2022

***Abstract*— This study is a study that aims to determine the results of the automatic portal access application using a static QR Code reading based on the results of the ISO 25010 testing standard and to find out the results of using an automatic portal access application using a QR Code reading. The development model used in this research is the Rapid Application Development (RAD) model and the application quality test is based on the ISO 25010 quality standard. Data were collected using documentation and observation techniques. The instrument and system were validated by two system experts. Data were analyzed using descriptive statistical analysis techniques. Based on the test results using the ISO 25010 quality standard, the automatic parking application using QR Code reading has met the standards where: a) testing the functionality suitability aspect Testing the application is carried out by involving 2 (two) expert validators who are experienced in the field of application programming, the percentage is obtained in the category Very Good and has met the functionality suitability aspect; b) testing the performance efficiency aspect of the application can run well without experiencing a memory leak which results in a force close and launch fail so that the application is said to be successful and has met the performance efficiency aspect; c) Testing the maintainability aspect From the test results on the maintainability aspect, it shows that the maintainability aspect test has met three aspects of the test, including instrumentation, consistency, and simplicity, so the results of the maintainability aspect test can be said to meet the maintainability standard.**

***Index Terms*— Application; Development; QR Code**.

I. INTRODUCTION

The development of technology is currently growing very rapidly, so technology has become a necessity at this time. Many needs in today's era can be met through mobile phones. One of the requirements needed is access to the parking area for motorists. Often the need for a parking area for motorists is still lacking in service.

These are some of the things that encourage thinking about building an automatic portal access application using a mobile, fast, effective and efficient QR Code reading. Android and QR Code are used in this study because Android is a very popular and widely used mobile OS (Operating System) and QR Code is a medium used to deliver information quickly and get a fast Response without inputting manually by typing.

Android is an operating system for Linux-based mobile devices that includes an operating system, middleware and applications. Zafaat (2015).

According to Supardi (2017) Android is a Linux-based mobile device operating system that includes an operating system, middleware, and applications.

Hilmawan et al (2019) Parking is a temporary state of immobility of a vehicle because it is abandoned by the driver. By law it is forbidden to park the vehicle in the middle of the highway, but parking on the side of the road is generally allowed. Parking facilities are built together with most buildings, to facilitate vehicles using the building.

Quick Response Code or commonly referred to as QR Code is a two-dimensional barcode introduced by the Japanese Company Denso Wave in 1994. This type of barcode was originally used for data collection of vehicle parts production inventory and is now used in various fields of business services and services for marketing and promotional activities. Basically, the QR Code was developed as a code that allows its contents to be translated at high speed Rouillard (2008).

The use of QR Codes is currently widely used in the business and industrial world, such as in Japan, China and Korea. QR Code is an evolution of barcodes from 1 dimension to 2 dimensions developed by Denso Wave. Along with the development of information technology, parking security structuring activities are also increasing according to the complexity of the existing work, the presence of android technology that can be used as a data processor is quite interesting for several conditions to use it, one of which is the problem of parking. The connection in this research is how to design a parking system concept that provides security and comfort for Android-based motorists that can be accessed directly through an Android smartphone. The design of an android-based application that is installed on the driver's smartphone contains user data which is input when the driver parks.

QR code is a type of matrix code or two-dimensional barcode developed by Denso Wave, a

division of Denso Corporation which is a Japanese company and published in 1994 with the main functionality that it can be easily read by a scanner QR Code stands for quick Response or rapid Response, which is in line with the goal is to convey information quickly and get a fast Response as well.

II. Research Method

This type of research is a Research & Development (R&D) research which aims to develop automatic portal access software using a static QR Code reading, which is an application that can be used to manage automatic parking in the Department of Informatics and Computer Engineering. To produce results that are in accordance with user needs, in its development using the Rapid Application Development (RAD) software development model. Suhartono & Susilawatizahraen (2022).

The model used in this development is the Rapid Application Development (RAD) development method which has 3 phases, namely Requirements Planning, RAD Design Workshop, and Implementation. Rapid Application Development (RAD) is a strategy aimed at providing much faster development and obtaining better quality results than results achieved through traditional cycles McLeod (2002).

The instrument in this study is used to assist in analyzing the quality of the software developed using ISO 25010 which is a development of the previous model, namely ISO 9126. In this ISO 25010 model there are eight criteria.

There are 3 aspects of testing mobile devices in ISO 25010 covering Functional suitability, performance efficiency, and maintainability. Therefore, in this research the software testing phase uses 4 of the 8 characteristics in ISO 25010.

Data analysis techniques used in this study are as follows: 1). Functional Suitability, Analysis Functionality testing by experts is determined from the results of calculating the percentage score for each instrument. On the answer sheet for each question item using the Guttman scale. This study uses an instrument in the form of a checklist and with yes-no answer choices. The Respondent's answer with the highest score (yes) is worth 1 and the lowest score (no) is 0. 2). Performance Efficiency Analysis testing, the performance efficiency aspect of the ISO/IEC 25010 standard using GT Metrix software/tools, which aims to measure and determine the performance efficiency aspect. 3). Usability Analysis, the usability aspect analysis test is done by analyzing user responses. Usability testing was carried out on 30 respondents, using a Likert scale with 5 scale options. Likert scale is a type of scale used to measure attitudes, opinions, and perceptions of a person or group. The reason for using respondents is that at this time the portal access application and system are still in the trial phase so 30 respondents are deemed sufficient, after the trial is declared successful, the application and system will be fully implemented.

III. Results and Discussion

A. Research Results

1) Application Development Results

The results of this study resulted in an automatic portal access application using a static QR Code reading. This application was created using the Android Studio software. Application database management using Firebase Realtime Database.

This study aims to produce an automatic portal access information system using a static QR Code reading. This application is made with features that make it easier for vehicle users when they want to enter the parking area and have a high capacity in coding data or in information storage. The following is a description of the stages of making an automatic portal access application using a static QR Code reading based on the selected development model.

A) System Evaluation Results

At this stage it is carried out by the user to find out whether the application that has been built is in accordance with the needs or not. If it is appropriate, it will proceed to the application coding stage, if not, the application will be repeated from Steps 1, 2, then 3. Based on the evaluation of the application by the user, the application design can be continued to the application coding stage.

B) System Coding Results

At the coding stage, the approved application is made using the appropriate programming language. The programming language used is the Kotlin programming language, with Arduino Studio IDE software. The database management system uses the Firebase Real time Database.

Fig. 1. Initial Appearance of the Application

The Initial screen when the automatic portal access application is opened for the first time, in this initial view the application opens by displaying the application icon and the loading bar of the application at the bottom of the device screen.

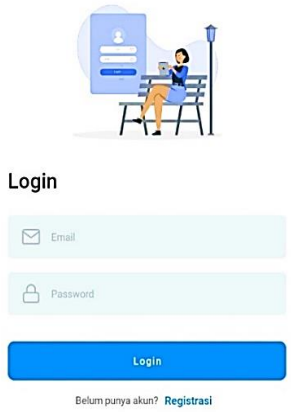


Fig. 2. Login Form

Login form is the display of the main page of the application that was built. This form contains the login and registration buttons. The login button is used to access the next page by entering the email and password data entered by the user, while the registration button will direct the user to a page containing the data entry form to register as a parking member.

Fig. 3. Registration Form

Registration Form is a registration form display on the application that was built. In this form there is a register button and a login list. The register button is used to register an account as a parking member by entering the name first then the NIP/NIM for the Lecturer or Student then entering the email then the mobile number then the password then re-entering the password. After completing the registration form, you will be directed return to the login page. The login button is used to access the next page according to the email and password data entered by the user.

Fig. 4. User Dashboard Form

The User Dashboard Form is a display from the user dashboard, this page contains a profile photo editing menu that contains data for the account user name, as well as the number registered in the application, also contains the mobile number of the account owner, as well as data on entry and exit times recorded in the system. In this display there is also a QR Code display that will be scanned by the reader system on the portal used to enter and exit.

C) System Test Results

1) QR Code System Response and Entry Time Barrier

Testing the Response of the QR Code and Barrier system which was carried out to determine the duration of time required when scanning to open the entrance portal. In the first experiment, the required scan time is 5.35 seconds, while the required Barrier Response time is 1.05 seconds, for more details, see the table below.

TABLE I. QR CODE RESPONSE AND ENTRY BARRIER TIME

No.	Scan Time	Barrier Response Time
1	5.35 Second	1.05 Second
2	6.07 Second	1.25 Second
3	3.04 Second	1. 42 Second
4	3.37 Second	1.12 Second
5	7.87 Second	1.90 Second
6	5.65 Second	2.00 Second
7	5.78 Second	1.34 Second
8	4.32 Second	2.00 Second
9	8.14 Second	1.95 Second
10	10.01 Second	1.30 Second
11	7.14 Second	1.83 Second
12	5.31 Second	1.74 Second
13	5.73 Second	1.35 Second
14	6.52 Second	1.92 Second
15	8.31 Second	1.74 Second
16	4.58 Second	1.56 Second
17	3.57 Second	1.42 Second
18	7.13 Second	1.67 Second
19	4.42 Second	1.81 Second
20	5.41 Second	1.48 Second
Average Scan Time 5.89 Second		Average Barrier Response Time 1.59 Second
Minimum Scan Time 3.04 Second		Minimum Barrier Response Time 1.05 Second

Maximum Scan Time 10.01 Second	Maximum Barrier Response Time1.95 Second

2) QR Code System Response and Barrier Time Out

Testing the Response of the QR Code and Barrier system which is carried out to determine the duration of the exit time required when scanning to open an exit portal. In the first experiment, the required scan time is 2.90 seconds, while the Barrier Response time is 1.98 seconds. For more details, see the table below.

TABLE II. QR CODE SYSTEM RESPONSE AND BARRIER TIME OUT

No.	Scan Time	Barrier Response Time
1	2.90 Second	1.98 Second
2	5.96 Second	1.45 Second
3	4.78 Second	1.23 Second
4	3.98 Second	2.00 Second
5	3.00 Second	1.76 Second
6	5.97 Second	1.87 Second
7	6.67 Second	1.24 Second
8	3.65 Second	1.47 Second
9	6.34 Second	1.68 Second
10	2.08 Second	1.47 Second
11	5.95 Second	1.54 Second
12	4.10 Second	1.59 Second
13	6.74 Second	1.56 Second
14	7.93 Second	1.47 Second
15	5.32 Second	1.40 Second
16	3.87 Second	1.37 Second
17	2.56 Second	1.52 Second
18	7.92 Second	1.57 Second
19	5.82 Second	1.35 Second
20	2.78 Second	1.09 Second
Average Scan Time 4.92 Second		Barrier Average Response Time 1.53 Second
Minimum Time 2.08 Second		Barrier Minimum Response Time 1.09 Second
Maximum Time 7.92 Second		Barrier Maximum Response Time 2.00 Second

3) Distance Scan QR Code and Barrier

The test is carried out to determine the distance from the QR Code scan process on the portal, and the time required by the system and the barrier in responding to the application. In the tests that have been carried out, it was found that various responses were found, as in the first experiment at a distance of 10 cm with a system response of 0.68 seconds and a response barrier of 1.00 seconds, for more details, see the table below.

TABLE III. DISTANCE TO SCAN QR CODE AND BARRIER

No	Scan Distance	System Response (Second)	Barrier Response (Second)
1	10 cm	0.68	1.00
2	10 cm	0.92	1.00
3	10 cm	0.74	1.00
4	10 cm	0.93	1.00
5	10 cm	0.78	1.00
6	15 cm	1.25	1.00
7	15 cm	1.02	1.00
8	15 cm	0.76	1.00
9	15 cm	0.98	1.00
10	15 cm	1.05	1.00
11	20 cm	1.75	1.00
12	20 cm	1.92	1.00
13	20 cm	1.68	1.00
14	20 cm	1.79	1.00
15	25 cm	1.90	1.00
16	25 cm	1.87	1.00
17	25 cm	1.96	1.00
18	25 cm	1.89	1.00
19	25 cm	1.92	1.00
20	25 cm	1.72	1.00

4) Test the Access Speed of Providers and Various Network Services

This test is carried out to test the speed of provider access from various network services, the effect of connectivity on the speed at which the scan data is sent to the system. In the first experiment, Telkomsel network service required 1.25 Second Response Time, for more details, see the table below.

TABLE IV. ACCESS SPEED TEST

No	System Response Time			
	TELKOMSEL	XL	SMART FREN	INDOSAT /TRI 3
1	1.25 Second	0.56 Second	1.27 Second	0.80 Second
2	0.87 Second	0.82 Second	1.31 Second	1.02 Second
3	1.76 Second	1.21 Second	0.47 Second	1.48 Second
4	1.45 Second	1.46 Second	0.85 Second	1.20 Second
5	1.37 Second	1.25 Second	0.92 Second	1.08 Second
Average	1.34 Second	1.06 Second	0.96 Second	1.12 Second

5) Test Results From Making Applications Based on Software Testing Standards ISO/IEC 25010

a) Application Test Results (Functional Suitability)

Application testing is carried out by involving 2 (two) expert validators who are experienced in the field of application programming. The validator tests the system directly by trying all the functions contained in the system, then fills out the test scores in the questionnaire table which has previously been validated by 2 (two) instrument validator experts by checking the options that are considered correct and providing input related to the development of the system being tested. . The results of the examiner's questionnaire from the 2 (two) experts are listed in the following table:

TABLE V. APPLICATION TESTING DATA PROCESSING RESULTS

Answer	Validator Score	
	Validator 1	Validator 2
Yes	7	7
No	-	-

The score result will then be calculated by the formula of the feature completeness matrix where in this formula the functional suitability result is the result

of the number of features that have been successfully implemented divided by the total designed features. Based on the formula, it is obtained.

$$Functional\ suitability\ (X) = \frac{7}{7} = 1 \qquad (1)$$

Based on the above, which is then calculated using the feature completeness formula, a Score of 1 is obtained, from the score it can be concluded that the quality of the software in the form of applications in this study is stated to be good and fulfills the functional suitability aspect. The description of the functional suitability aspect test data can be seen in the appendix.

b) Application Test Results (Functional Suitability)

Aspects of testing include the use of the Central Processing Unit (CPU), memory, memory, thread counts, and Frames Per Second (FPS). Testing is done using a desktop application called Apptim. The following are the results of testing the performance efficiency aspect with Apptim in Figure 5.

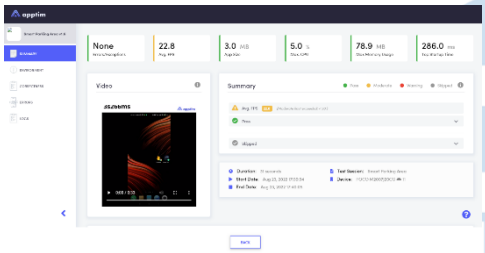


Fig. 5. Performance Efficiency Testing

Performance efficiency testing that has been carried out shows that the automatic parking application using a static QR code reading can be described in Table 6 as follows.

TABLE VI. PERFORMANCE EFFICIENCY ASPECT TEST RESULTS

No.	Category	Average	Maximum Score	Result
1.	CPU Use	0,9%	5%	Low Limit
2.	Memory Use	82,1 MB	114,2 MB	Low Limit

Based on the table above, the test results on the performance efficiency aspect of the Automatic Portal Access Application Using Static QR Code Reading using Apptim software show Average CPU usage is 0.9% with Maximum 5 percent (low limit), and Average memory usage is 82.1 MB with Maximum 114.2 (low limit). During ongoing testing the application runs well without experiencing a memory leak or application damage that results in a force close or launch fail, the test scores are also at a low limit, no test is above the maximum user limit, therefore it can be concluded that the test results on Performance efficiency aspects of Automatic Portal Access applications Using Static QR Code Reading have met ISO 25010 software quality testing standards on performance efficiency aspects.

c) Application Test Results (Functional Suitability)

Maintainability testing is carried out operationally in the field using three aspects of testing, including instrumentation, consistency, and simplicity by involving 2 (two) experienced experts in this field. The expert tested the system directly by trying 3 (three) aspects of testing in the system, then filling in the test scores in the questionnaire table. The results of the tester's questionnaire from the 2 (two) experts are listed in the following table:

TABLE VII. MAINTANABILITY DATA PROCESSING RESULTS

Answer	Validator Score	
	Validator 1	Validator 2
Yes	3	3
No	-	-

The quality of maintainability testing on a software is carried out according to the Land testing instrument contained in Table 7. From the test results of the maintainability aspect as in Table 7 above, it shows that the maintainability aspect test has met three aspects of the test, including instrumentation, consistency, and simplicity, so the results of the maintainability aspect test can be said to meet the maintainability standard.

B. Discussion

The automatic portal access application using a static QR Code reading is an application designed to provide convenience in parking services. Development of automatic portal access application using static QR Code reading using Research & Development (R&D) research type. The stages of application planning are starting from needs analysis, this stage uses data collection techniques in the form of interviews using irregular question instruments. The results of the interview then proceed to the application creation stage, in the design process a system design is made which consists of database design, business process design, user interface design, and system architecture design. The application that has been designed is then evaluated by the user to determine whether the system design can be continued or not, if it is appropriate then the process can be continued to the system coding process, if not, the application will be designed from scratch. The results of the evaluation of applications that have been carried out by developers with users, namely development can proceed to the system coding stage. The coding of the system uses the Kotlin programming language with the Arduino Studio IDE software. The database is managed using the Firebase Realtime Database. Based on the results of the research, an automatic portal access application was produced using a valid, practical, and effective static QR Code reading for use in parking managers.

Testing automatic parking applications using static QR Code readings is carried out through several stages, namely testing using expert validation. Validation of

experts (expert) that is asking for feedback and suggestions from 2 system experts, the validator is in charge of providing an assessment of whether the developed application is good and feasible to be implemented or not. The system expert provides an assessment of the system by filling out a questionnaire sheet that was made and has been previously validated by 2 instrument experts. Testing the system using the ISO/IEC 25010 testing standard which takes 3 aspects of testing, namely functional suitability is a test carried out to test all functional requirements of the system and then calculates the percentage of system success, performance efficiency is a test carried out to measure how fast in accessing applications, requests and responses. data, maintainability is a test conducted to determine the ability of application maintenance.

The results of application testing from the functional suitability aspect based on the test table can be concluded that the quality of the software in the form of applications in this study is stated to be good and fulfills the functional suitability aspect. The results of performance efficiency testing that have been carried out show that the automatic portal access application using a static QR Code reading can run well without experiencing a memory leak or application damage resulting in a force close or launch fail, the automatic portal access application using a static QR Code reading has met the testing standards. software quality in terms of performance efficiency. The results of testing the maintainability aspect based on the test table, showed that it had met three aspects of the test, including instrumentation, consistency, and simplicity, so the results of testing the maintainability aspect could be said to meet the maintainability standard.

Based on the results of the research on designing an automatic portal access application using a static QR Code reading after being validated and tested using the ISO/IEC 25010 software testing standard using aspects of functional suitability, performance efficiency, maintainability, it can be concluded that this application is suitable for use by end users.

IV. CONCLUSIONS

Based on the results of the research conducted, the following conclusions were obtained:

1. This study resulted in an automatic portal access application using a static QR Code reading, in testing the application it has a Time Response scan and a Time Response barrier with an Average Time Response scan of 5.89 Second, while the Average Time Response barrier of 1.59 Second is quite practical in using the application and has been feasible to use based on the results of the application's feasibility test.
2. The automatic portal access application using a static QR Code reading shows that this application has met the ISO/25010 standard in 3 aspects of testing, namely the functional suitability aspect with the test results being well received, performance efficiency with the test results running very well without experiencing memory leak or application damage, and maintainability can be said to meet the standard.

REFERENCES

[1] Siregar, H. F., & Siregar, Y. H. (2018). Perancangan aplikasi komik hadist berbasis multimedia. 2(2), 113–121.

[2] Supardi, I. Y. (2017). Semua bisa menjadi programer android. Elex Media Komputindo.

[3] Hilmawan, N., Akhmadali, & Sumiyattinah. (2019). Analisa kebutuhan parkir pada pasar flamboyan sekadau. 1–4.

[4] Suhartono, & Susilawatizahraen, R. (2022). Development of e-document information system in information and computer engineering department using quick Response code. 1–9.

[5] M. I. Perkasa and E. B. Setiawan, "Pembangunan Web Service Data Masyarakat Menggunakan REST API dengan Access Token,"J. Ultim. Comput., vol. 10, no. 1, app. 19-26, 2018, doi: 10.319337/sk.v10i1.838.

[6] A. Pulungan and A. Saleh, "Perancangan Aplikasi Absensi Menggunakan QR Code Berbasis Android," J. Mhs. Fak. Tek. Dan Ilmu Komput., vol. 1, no. 1, pp. 1063-1074.

[7] R. H. D. A. M. N. H. Khosa, Aplikasi Inventory barang menggunakan QR code. Bandung: Kreatif Industri Nusantara, 2020.

[8] Sugiyono, Metode Penelitian kuantitatif, Kualitatif dan R&D. B: Alfabeta Cv, 2017.

[9] K. B. Dsouza, S. Mohammed and Y. Hussain, "Smart parking-An integrated solution for an urban setting," 2017 2nd Internasional Conference for Convergence in Technology (12CT), 2017, pp. 174-177, doi: 10.1109/12CT.2017.8226115.

[10] K. Jing Yong and Muataz H. Salih. "Design and Implementation of Embedded Auto Car Parking System Using FPGA for Emergency Conditions". Indonesia Journal of Electrical Engineering and Computer Science, 2019, Vol. 13, No. 3, pp. 876-883.

[11] Huang, P. C., Chang, C. C., Li, Y. H., and Liu, Y., "Efficient access control system based on aesthetic QR code," Personal and Ubiquitous Computing, vol. 22, pp. 81-91, 2018.

[12] K. Peng, H. Sanabria, D. Wu, and C. Zhu, "Security overview of QR codes," Student project in the MIT course, vol. 6, pp. 1–20, 2014.

[13] S. Tiwari, "An introduction to QR code technology," in 2016 International Conference on Information Technology (ICIT), Dec. 2016, pp. 39-44, doi: 10.1109/ICIT.2016.021

[14] Denso Wave Inc. "QR code essentials." Denso Wave, vol. 900, 2011.

[15] A. Somani, S. Periwal, K. Patel, and P. Gaikwad, "Cross Platform Smart Reservation Based Parking System," 2018 International Conference on Smart City and Emerging Technology(ICSCET), pp. 1-5, 2018.

[16] T. S. Gunawan, A. Mutholib, and M. Kartiwi, "Design of automatic number plate recognition on android smart-phone platform," Indonesia Journal of Electrical Engineering and Computer Science, vol. 5, no. 1,, pp. 99-108, 2017.

[17] Sri Hastuti, Raditya A.R, S.T., M.T, Slamet Indriyanto, S.T.,M.T. Rancang Bangun Sistem Informasi Lokasi Meja Pada Food Court Menggunakan Rfid (Radio Frequency Identification) Berbasis Aplikasi Android". Jurnal EECCIS Vol 14, No 3 (2020) pp. 101-107.

[18] Noor H, dkk. Sistem Informasi Perpustakaan. Berbasis Web Menggunakan Barcode Pada SMP Negeri 11 Banjarmasin. Technologia. 2020.

[19] Tri Haryanti. "Sistem Informasi Perpustakaan Menggunakan Barcode". Smart Comp Vol. 11 No. 1, 2022.

[20] Abdul B, dkk. Smart Door Lock Berbasis QR Code. Smart Comp Vol. 11 No. 1, 2022.

ISSN 2085-4552

Use of eGovernment in Public Services Area

Case Study: Bojonegoro Regency

Hastie Audytra[1], Rahmat Irsyada[2]
[1]Information System Department, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
[2]Computer System Department, Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
[1]hastie@unugiri.ac.id, [2]irsyada.rahmat@unugiri.ac.id

Accepted 18 October 2022
Approved 22 November 2022

***Abstract*— The objective of e-Government is the availability of access to information by the citizen rather than financial benefits. e-Government system is built based on web technology to provide better public services to the citizen. Public services are services provided by the government to the citizen. Bojonegoro itself has several times received awards in the field of Smart City or e-Government. In this study, we want to know how are the results of using the eGovernment application in the Public Service Area. From the results of the literature study and data collection from 250 public service users, it appears that the applications that are widely used or that attract public interest are national applications that have been around for a long time, such as the Lapor! And online tax service application. Local applications that are being actively socialized actually get the smallest percentage. This study provides recommendations for solutions that can be applied for the implementation of better public services.**

***Index Terms*— eGovernment; Public Services; Smart City; Smart Government; Smart Governance**.

I. INTRODUCTION

The rapid development of technology, information and communication (ICT) will open opportunities and challenges to create, access, process, and utilize accurate information. ICT promises efficiency, speed of information delivery, global reach and transparency. Therefore, to realize good governance, one of the efforts is to utilize ICT or better known as Digital Government (e-Government) [1].

e-Government is not changing the way the government communicate with citizen. In the concept of e-Government, people can still contact service posts, talk by telephone to get government services, or send letters. So, e-Government is the use of ICT that can improve relations between government and other parties. e-Government is an effort to develop electronic-based government administration in order to improve the quality of public services effectively and efficiently [2]

Public services are services provided by the government to the community. However, in its implementation there are still discrepancies with what is expected by the community as recipients of public services. This is evidenced by the number of complaints submitted by the public [3]. Good quality of public services can be achieved if the government and other public institutions as a whole are able to be open to ideas, new ideas and criticism based on the interests of the community. Community responsiveness will increase if the community has access to information about the development process, government services, and involvement in the policy-making process. To strive for good governance, it will be more effective if the government also opens access to information to the public [4] e-Government provides new opportunities to improve the quality of government, by provide new services, increasing efficiency, increasing global information infrastructure and increasing citizen participation. Thus e-Government will improve the quality of public information services as a way to realize good governance [5]

East Java is one of the provinces that is starting to improve towards e-Government. There have been several studies in the East Java area that lead to e-Government to improve the quality of public services, including: making E-Health which is applied in the development of the health service sector of the Surabaya city government to improve public services [6], analysis of the application of E-Government and Good Governance towards E-SAMSAT services in East Java, precisely at the Mojokerto SAMSAT Joint Office in the service operations section [7], analysis of the effectiveness and efficiency of implementing public service transformation through an electronic-based government system in the city of Malang [8], analysis of the application of e-Government in the Regency Sidoarjo in improving the performance of public services [5]. For Bojonegoro Regency, there has been research on open government, namely SIAP LAPOR application management [4] and Open Government Partnership (OGP) implementation [9].

Bojonegoro has several times received awards in the field of Smart City or e-Government. This shows that Bojonegoro has started to move towards e-Government. Therefore, with the existence of several previous studies in the Bojonegoro area, and with the Covid19 pandemic as an accelerator of digital

transformation in the provision of public services [10], [11] it is hoped that this research can produce an in-depth analysis of the condition of e-Government, especially in the scope of public services in Bojonegoro Regency to be taken into consideration for the development of public services in Bojonegoro Regency and increase knowledge and provide information about the factors that affect good governance in terms of public services so that it can be used as a basis for consideration by the Bojonegoro Regency Government in making policies to develop strategic steps to improve public services to achieve a better quality of government.

Problem Statement/Problem Formulation

The ideal of a good e-Government is to produce good services for the community and government employees themselves. As one of the cities that has received several awards in the field of Smart City, Bojonegoro Regency must be good at eGovernment. This research wants to find out about eGovernment applications for public services area in Bojonegoro Regency. Are these applications used properly?

Scope of problem

This research is part of the research roadmap that the research team aspires to, namely the Digital Government and Smart City of Bojonegoro Regency as a whole. However, this research will only focus on public service applications because one of the goals of a smart city or digital government is the availability of good public services.

Research purposes

This study aims to produce an in-depth analysis report on the implementation of e-Government and its impact on public services. From the results of the literature study, interviews on the application of e-Government in Bojonegoro Regency and data collection from public service users, it is expected to produce an analysis document of the condition of e-Government in Bojonegoro Regency and a framework that can help Digital Government in the Bojonegoro Regency Government in a more structured and more efficient manner. especially in terms of public services.

Benefits of research

This research is expected to have benefits both theoretically and practically. From the research results, it is expected to produce a framework or recommendations that are useful for the implementation of e-Government based on public services. For the researchers themselves, it is hoped that this research will produce new knowledge, especially to support Smart Government for Smart City. For users of e-Government public services, it is hoped that the results of this study can be a reference for how to create and implement good public services.

II. Literature Review

A. E-Government

The vast geographical conditions are a big challenge for the government and the people of Indonesia in the era of globalization and information, especially how to connect the scattered areas of Indonesia with a large population through ICT [12]. The use of internet technology is expected to be a vehicle for accelerating the exchange of information, providing service facilities and transaction activities [13]. One of the important roles of e-Government which is very much expected is to start an efficient work culture that is free from non-transparent and corrupt behavior, collusion, and nepotism (KKN) in public services [14].

e-Government leads to three main classifications, namely government-to-business (G2B), government-to-citizen (G2C) and government-to-government (G2G) [15].

1. G2B e-government is about the interaction between government and business (G2B). This helps businesses deal with government more effectively. The most common example is e-procurement services. This service enables a transparent bidding process to all businesses, where previously these businesses needed to constantly contact the government for updates during the bidding process.
2. G2C e-government focuses on the interaction between government and citizens. It is about using e-government to get better public services. For example, G2C e-Government allows citizens to apply for passports online, whereas previously this service could only be done by physically visiting the immigration office and residents had to wait in long queues.
3. G2G e-government enables information sharing among public organizations. It deals with the exchange of information between government agencies and employees at the national, provincial and local levels. The goals of G2G e-government include better coordination between government agencies, standardized procedures, and greater efficiency for government.

Challenges of e-Government Implementation

Factors that can hinder and become challenges for the implementation of e-Government are as follows: [16]

1. There is no clear standardization regarding the implementation of e-Government and how to implement e-Government correctly
2. The information network infrastructure is not evenly distributed to the regions.
3. The lack of good human resources (HR) to manage e-Government.
4. Low leadership commitment to support e-Government
5. Low public literacy to use e-Government

6. Organizational culture that is less supportive of change and low culture of sharing information, especially between government institutions
7. Some new areas understand that the e-Government is only to building a website from manual to digital.

In addition, the application of e-government is a form of implementing the use of information technology for government services to the public. That is how the government provides information to stakeholders through a web portal. Differences in understanding, perspective and application of e-government have led to distortions and deviations from the purpose of making e-government itself. The weak use of e-government in a bureaucratic environment that is interrelated with the still limited application in the business world has caused the slow implementation of e-government programs [13].

B. Related Researches

Many researches on e-Government have been carried out, including:

1. Research on the existence of document tracking technology [17], [18] This research was conducted to simplify how to manage, track, and handle electronic documents in government agencies, which can also support frequent modifications of business processes. This study proposes to integrate document tracking technology into e-Government business processes to improve the efficiency and effectiveness of e-Government applications in Indonesia.
2. Development of an e-Government software requirements specification model [19], this research was conducted to produce a software requirements specification model for e-Government in order to obtain a good e-Government SKPL document.
3. Development of an e-Government software quality model [20], this research was conducted to produce a model of e-Government software quality in order to find out how good the software quality is.
4. e-Government Maturity Model [21], [22], to achieve this goal, the researcher considers five determinants, namely: trust and awareness; simplified services; use of the most advanced technology; detailed processes and. agile accessibility.
5. The Garuda Smart City Framework [23], this research, uses the tools of the Garuda Smart City Framework 2 (GSCF 2) whose indicators include digital government as part of the measurement in addition to sustainable indicators such as economic, social and environmental as well as enabler indicators such as technology. infrastructure, people and government. Through this measurement, it is hoped that cities will know their position and are expected to understand how to move towards a smarter and more sustainable city.
6. Development of an assessment model to assess e-Government software [24], this study develops an e-Government software asset valuation model for recommendations for its maintenance at the Regional Personnel Agency (BKD) Bandung.
7. Measuring the quality of e-Government services with eGovQual [25], [26], This study uses the e-GovQual model by adding one dimension, the purpose of this study is to obtain user ratings of the quality of public service applications. that there are 24 attributes from five dimensions: trust reliability, efficiency, user satisfaction and community support.
8. Challenges and barriers to eGovernment in Indonesia [15], this study analyzes the challenges for the development of e-Government in Indonesia.
9. E-Government resource management with Big Data [27], the purpose of this study is to explore how to manage the resource from eGovernment using big data.
10. The development of community-oriented public services [28] to improve the efficiency and effectiveness of government services this research focused on the transition from analog to (parallel) digital services and digital transformation to reengineer and redesign government services from scratch to meet the needs users who continue to change.

III. RESEARCH METHODOLOGY

A. Data Source

The data used in this study is secondary data. The data collection process was carried out through a search of the results of previous research, as well as data from the Bojonegoro Regency Government related to e-Government public services. Secondary data collected in the form of 250 questionnaires from the public about the assessment for the quality of public e-government services in Bojonegoro Regency.

The lead researcher is responsible for data collection, data analysis, literature study, analysis of the Bojonegoro eGovernment condition and the obstacles found.

Research members are tasked with collecting data from the community, analyzing existing public service systems/applications, and analyzing infrastructure that supports the implementation of Digital Government in Bojonegoro.

B. Methodology

Phase I: Understanding Existing Condition
Phase Ib: Identify Failure and Success Factors
Phase II: Pilot Study for Solutions and Development
Phase III: Further Action

IV. Result and Discussion

A. *Phase 1 : Understanding Existing Condition*

In order to understanding existing condition, this research have to collect the data first. The data collection of this paper was performed questionnaire or interview to the stakeholders and based on a literature review. But first we focused on the systematic literature review. A systematic literature review is defined through the usage of a comprehensive search that scans the relevant body of literature with clearly stated and comprehensible search choices and selection criteria [29].

Adoption from [30], the systematic literature review consisted of open searches using the following words: "smart cities" & "governance", "smart governance", "e-Government" and "smart government" by means of browsing through titles, author supplied keywords and abstracts. Research data is taken from several sources: Link Springer, IEEE Xplore, Research Gate, Science Direct, Google Scholar, ACM and others.

There are around 40 journals articles for systematic literature review about smart city, e-Government, smart governance, including how it is applied in Indonesia and Bojonegoro.

There are 17 priority programs of the Bojonegoro Regency Government:

1. Repair of roads, bridges and infrastructure and PJU (public street lighting);
2. KPM (Independent Farmer Card) Plus;
3. Supporting the regulation of Madrasah Diniyah through the initiation of the Regional Regulation on Madin;
4. The best education services for elementary to high school schools;
5. New job opportunities;
6. Additional incentives for GTT and PTT;
7. ALADIN Program (Repair of roofs, floors and walls of underprivileged houses);
8. MSME capital for BUMDES based on tourism and Agribis;
9. A friendly environment for women, children, persons with disabilities as well as the poor and child care parks;
10. Health includes 24-hour Puskesmas services, strengthening Polindes, additional nutrition for pregnant and lactating women;
11. Revise Perbup No. 35 of 2015 to improve the welfare of the Village Head and Village apparatus;
12. RT/RW Remuneration;
13. Compensation for grief;
14. Bojonegoro Green and Smart City include green open spaces, child-friendly open houses, as well as access to information and the internet;
15. Ease of licensing and One-Stop Investment;
16. Annual Bojonegoro Cultural Festival;
17. Arrangement of traditional markets towards modern markets

The existence of 17 excellent programs is supported by the application provided by the Government of Bojonegoro Regency. One of the targets of the Ministry of Communication and Informatics as a pilot project for the eGovernment of the Bojonegoro Regency Government is to find out how to use the application and how deep the socialization of the 17 flagship programs reaches the community.

1. LAPOR!!
2. Online Single Submission (OSS)
3. Online Licensing Service (SIPPADU Application)
4. Online Tax Service e-Tax Application
5. Online MSME Services
6. IT Helpdesk Service
7. Assistance Submission Service (SiBantu Application)
8. Grief Compensation Service (SANDUK Application)

The current study has completed taking data in the form of a questionnaire on the Use of Public Service Applications by the Bojonegoro Regency Government. Questionnaires were distributed to 250 respondents with various backgrounds.

a) Respondent Gender:

Based on Figure 1, we know that 53% respondents are women and 47% respondents are men.

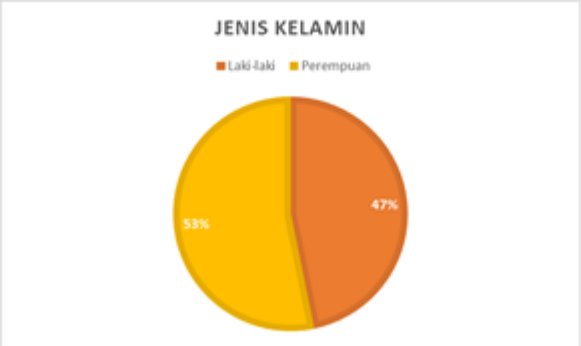


Fig. 1. Respondent Gender

b) Respondent Age:

As seen in Figure 2, most of respondents are in the productive ages. Most of them are 18-25 years old and 26-40 years old.

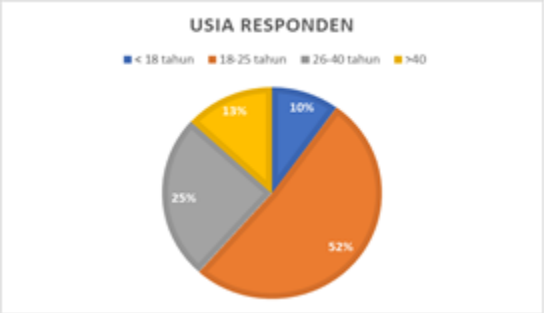


Fig. 2. Respondent Age

c) *Respondent Education Level:*
The education level is good enough. As seen in Figure 3, most of them have finished high school degree.

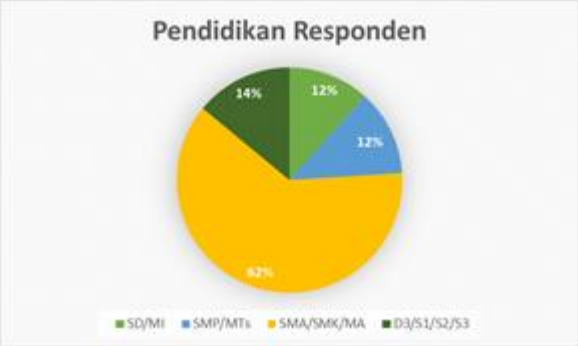


Fig. 3. Respondent Education Level

d) *Respondent Occupation:*
Respondents came from many field of work. It makes diversity about occupation.

Fig. 4.Respondent Occupation

e) *Respondent Website Usage:*

Fig. 5. Respondent Website Usage

f) *Respondent Access to the Government Website:*

Fig. 6. Respondent Access to the Government Website

From the results of the questionnaire, the following are the results of the public's understanding of the application of these public services:

1. Public service applications that are known to the public:

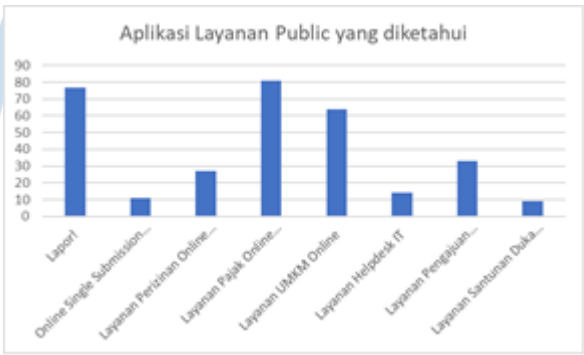


Fig. 7. Public service applications that are known to the public

2. Public service applications that have been used by the public:

Fig. 8. Public service applications that have been used by the public

3. Public service applications that may later be used:

Fig. 9. Public service applications that may later be used

From the results of the questionnaire, it can be seen that the applications that are widely used or that attract people's interest to use them are national applications that have been around for a long time, such as the Lapor! And an online tax service application. Local applications that are being actively socialized (SANDUK Applications) actually get the smallest percentage.

B. Phase Ib: Identify Failure and Success Factors

Failure Factors:

1. Lack of human resources (sufficient to manage e Government).
2. Network infrastructure and information dissemination that has not been evenly distributed to the regions.
3. Low public literacy to use e-Government.
4. Lack of information about e-Government applications or website in public services area.

Success Factors:

1. Leadership in terms of its commitment to support e-Government in Bojonegoro Regency is already good.

C. Phase II: Pilot Study for Solutions and Development

1. Discussion of Questionnaire Results.
   From 250 questionnaire data, it turns out that the results obtained are not satisfactory. Therefore, further discussion is needed to convey the results of this questionnaire to the Department of Information and Communication Technology of Bojonegoro Regency so that they can take the next step.
2. HR Workshop and Training
   Train human resources managing public service applications so that they can use the application properly
3. Application training and socialization
   Conduct training and socialization of public service applications to the community
4. Collaboration Initiation
5. Collaborate with students and students to assist the socialization process in each village so that the reach of information can be wider.

D. Phase III: Further Action

Most of the obstacles faced are bureaucratic constraints to the Bojonegoro Regency Government. The process of taking care of permits and correspondence is quite time-consuming. It is hoped that for future research this can be reduced so that the research carried out can run better.

V. CONCLUSIONS

There is still a lot of work to be done in the Bojonegoro Regency e-government, especially in terms of public services. This research is not intended to vilify or bring down any party. This research is purely done to provide solutions so that e-government in Bojonegoro Regency can be even better.

REFERENCES

[1] D. A. D. Putra et al., "Tactical steps for e-government development," Int. J. Pure Appl. Math., vol. 119, no. 15, 2018.

[2] E. A. Sosiawan, "Tantangan Dan Hambatan Dalam Implementasi E-Government Di Indonesia," Semin. Nas. Inform., vol. 2008, no. semnasIF, pp. 99–108, 2008.

[3] H. Azkiya, "Penerapan E-Government dalam Peningkatan Pelayanan Publik (Studi pada Pelaksanaan Program e-Tax Pajak Restoran di Kota Malang)," Kebijak. dan Manaj. Publik, vol. 6, no. 2, pp. 1–10, 2018.

[4] R. R. Girianom, "Pengelolaan Sistem Integrasi Aspirasi Pengaduan – Layanan Aspirasi dan Pengaduan Online Rakyat (SIAP-LAPOR!) dalam Mewujudkan Open government di Kabupaten Bojonegoro," pp. 1–21, 2018.

[5] S. Mariono, "Penerapan E-Government Dalam Pelayanan Publik Di Kabupaten Sidoarjo," J. Chem. Inf. Model., vol. 53, no. 9, pp. 1689–1699, 2019.

[6] I. Putri, K. Rachmawati, and F. Mardianto, "Implementation of Surabaya Digital Government Service (DGS) With E-Health Services To Support Sustainable Development Goals (Sdgs) in Indonesia Based on Chi-Square Method," PalArch's J. Archaeol. Egypt, vol. 18, no. 1, pp. 886–895, 2021.

[7] W. Fannisya Surya and S. Ngumar, "Penerapan E-Government dan Good Governance Terhadap Pelayanan E-Samsat Jawa Timur," J. Ilmu dan Ris. Akunt., vol. 7, no. 2, pp. 1–15, 2018.

[8] B. K. Putra and C. B. Dhanuarta, "Transformation of Public Services Through Electronic-Based Government System in Malang City," Int. J. Educ. Inf. Technol. Others, vol. 4, no. 3, pp. 452–459, 2021, doi: 10.5281/zenodo.5205295.

[9] M. M. Huda, "Implementasi Open Government Partnership (Ogp) Di Kabupaten Bojonegoro," J. Ilm. Manaj. Publik dan

Kebijak. Sos., vol. 3, no. 2, pp. 374–388, 2020, doi: 10.25139/jmnegara.v3i2.2138.
[10] S. Ibad and Y. W. Lolita, "Development and Urgency of Public Services through E-Government in the Middle of Pandemi Covid-19," J. Public Adm. Gov., vol. 10, no. 4, p. 273, 2020, doi: 10.5296/jpag.v10i4.17867.
[11] D. Agostino, M. Arnaboldi, and M. D. Lema, "New development: COVID-19 as an accelerator of digital transformation in public service delivery," Public Money Manag., pp. 1–4, 2020, doi: 10.1080/09540962.2020.1764206.
[12] B. Hermana and W. Silfianti, "Evaluating e-government implementation by local government: digital divide in internet based public services in Indonesia," Int. J. Bus. Soc. …, vol. 2, no. 3, pp. 156–164, 2011, [Online]. Available: http://ijbssnet.com/journals/Vol._2_No._3_[Special_Issue_-_January_2011]/18.pdf.
[13] S. Jaya, "Implementasi Dan Perkembangan E-Government di Indonesia," J. Inform. Multimed. STIMED NUSA PALAPA, vol. 2, no. 1, pp. 37–52, 2001.
[14] E. Satriya, "PENTINGNYA REVITALISASI E-GOVERNMENT DI INDONESIA," pp. 38–43, 2006.
[15] A. Sabani, H. Deng, and V. Thai, "Evaluating the development of E-government in Indonesia," PervasiveHealth Pervasive Comput. Technol. Healthc., pp. 254–258, 2019, doi: 10.1145/3305160.3305191.
[16] M. Silalahi, D. Napitupulu, and G. Patria, "Kajian Konsep dan Kondisi E-Government di Indonesia," Teknol. Inf. dan Komun., vol. 1, no. 1, pp. 10–16, 2015, [Online]. Available: http://id.portalgaruda.org/?ref=browse&mod=viewarticle&article=356188.
[17] B. Hendradjaya and W. D. Sunindyo, "Towards Document Tracking Measurement Model to Support e-Government Business Processes," 4th Int. Conf. Inf. Commun. Technol. (ICoICT), Bandung, 2016.
[18] W. Sunindyo, B. Hendradjaya, G. A. P. Saptawati, and T. E. Widagdo, "Document tracking technology to support Indonesian local E-governments," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8407 LNCS, no. 1, pp. 338–347, 2014, doi: 10.1007/978-3-642-55032-4_33.
[19] H. Audytra, B. Hendradjaya, and W. D. Sunindyo, "A proposal for quality assessment model for software requirements specification in Indonesian language for e-Government," Proc. 2016 Int. Conf. Data Softw. Eng. ICoDSE 2016, 2016, doi: 10.1109/ICODSE.2016.7936125.
[20] R. Andrian, B. Hendradjaya, and W. D. Sunindyo, "Software Assessment Model Using Metrics Products for e-Government In The G2B Model," 4th Int. Conf. Inf. Commun. Technol. Bandung, 2016.
[21] I Made Sukarsa et al., "Evaluation of E-Government Maturity Models in Sub-District Public Services in Indonesia Using the SPBE Framework," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 4, no. 2, pp. 243–253, 2020, doi: 10.29207/resti.v4i2.1825.
[22] P. R. Joshi and S. Islam, "E-government maturity model for sustainable E-government services from the perspective of developing countries," Sustain., vol. 10, no. 6, 2018, doi: 10.3390/su10061882.
[23] S. H. Supangkat, A. A. Arman, R. A. Nugraha, and Y. A. Fatimah, "The Implementation of Garuda Smart City Framework for Smart City Readiness Mapping in Indonesia," J. Asia-Pacific Stud., vol. 32, no. 4, pp. 169–176, 2018, [Online]. Available: https://core.ac.uk/download/pdf/159504667.pdf.
[24] R. Atimi, B. Hendradjaya, and W. Sunindyo, "Developing an Assessment Model of E-Government Software Assets for Maintenance Recommendations: A Case Study in BKD Bandung," 2020, doi: 10.4108/eai.12-10-2019.2296517.
[25] N. A. Hidayah, M. C. Utami, and N. Fajrisani, "Measurement of Public Service Applications Quality Using the Electronic Government Quality (E-GovQual) Framework," vol. 408, no. November 2018, pp. 106–109, 2020, doi: 10.2991/assehr.k.200220.019.
[26] T. E. Wijatmoko, "E-Government Service Quality Using E-GovQual Dimensions Case Study Ministry of Law and Human Rights DIY," Proceeding Int. Conf. Sci. Eng., vol. 3, no. April, pp. 213–219, 2020, doi: 10.14421/icse.v3.500.
[27] M. N. I. Sarker, M. A. Hossin, A. N. K. Frimpong, and Y. Xiaohua, "Promoting information resource management for E-government through big data approach," ACM Int. Conf. Proceeding Ser., pp. 99–104, 2018, doi: 10.1145/3277139.3277155.
[28] I. Mergel, R. Kattel, V. Lember, and K. McBride, "Citizen-oriented digital transformation in the public sector," *ACM Int. Conf. Proceeding Ser.*, pp. 5–7, 2018, doi: 10.1145/3209281.3209294.
[29] R. W. S. Ruhlandt, "The governance of smart cities: A systematic literature review," *Cities*, vol. 81, no. October 2017, pp. 1–23, 2018, doi: 10.1016/j.cities.2018.02.014.
[30] M. P. R. Bolívar, "Mapping dimensions of governance in smart cities. Practitioners versus Prior Research," *ACM Int. Conf. Proceeding Ser.*, vol. 08-10-June, pp. 312–324, 2016, doi: 10.1145/2912160.2912176.

ISSN 2085-4552

Implementing the Chaotic Permutation Multicircular Cryptography Technique using Asymmetric Key

Aria Lesmana[1], Ruki Harwahyu[2], Yohan Suryanto[3]
[1,2,3] Department of Electrical Engineering, Universitas Indonesia, Depok, Indonesia
[1]aria.lesmana01@ui.ac.id, [2]ruki.h@ui.ac.id, [3]yohan.suryanto@ui.ac.id

Accepted 18 October 2022
Approved 30 November 2022

***Abstract*— In digital computing, cryptographic methods consider performance in both speed and security. This study aims to explore and improve a permutation-based symmetric chaotic cryptography technique called Chaotic Permutation Multicircular (CPMC). In this study, a method is proposed to implement asymmetric key system from CPMC technique by generating a reverser key for reverting the permutation result of CPMC encryption back to its original arrangement using the same function as the encryption process. The reverser key alongside the CPMC key act as encryption and decryption key pair. The pair key generation and cryptographic function utilizes the encryption function of CPMC technique dubbed CPMC Shrinking algorithm. Asymmetric implementation can simplify CPMC technique by also using CPMC Shrinking algorithm for decryption, therefore enabling it as a single function for encryption and decryption. The asymmetric implementation test showed improvement in total speed compared to initial implementation by average of 75.87% from tests using different block sizes.**

***Index Terms*—asymmetric cryptography; chaos cryptography; chaotic permutation; cryptography; permutation-based cryptography**.

I. INTRODUCTION

Digital communications are still vulnerable to security risks, such as eavesdropping or data leakage. Digital systems generally secure data by implementing a cryptographic system. Cryptographic systems in addition to requiring key confidentiality guarantees, also require high randomization for the confidentiality of the results and algorithms that are not heavy for encryption and decryption.

The application of Chaos Theory in mathematics for cryptographic methods gave rise Chaos Cryptography, through making use of mathematics functions with chaotic properties to be used as a pseudorandom number generator and can be combined with other functions or algorithms to become a cryptographic method. The widely used chaos mathematical function is in the form of a map, or dubbed chaotic map. Chaos cryptography methods is widely applied as an image encryption method [1-6], such algorithms also been developed using public key or as an asymmetric cryptography, especially the ones using Chebyshev Polynomial [15], [16], another method proposed by Silva-Garcia et. al. uses Elliptic Curve function alongside S-box Permutation generated by chaos resulting from Logistic Map function [17].

Permutation method in cryptography is widely applied alongside the substitution cipher method [7-14]. In this study, we explore the application of a permutation-based cryptography with a chaotic characteristics dubbed Chaotic Permutation Multicircular (CPMC). CPMC is a chaotic cryptographic technique based on permutations by Suryanto et al. [18], CPMC algorithm is suitable for data encryption due to its large keyspace of $(2^N)!$ [19]. CPMC algorithm applies rotational shift of a set elements, using a key generated with special modulus and LCM rules towards a set of natural number sequences to produce chaotic properties. CPMC algorithm has been implemented for image encryption [20-22] and audio [23], however both implementations are limited to symmetric block cipher algorithm.

This study proposes an asymmetric key cryptography implementation of the CPMC technique, by designing an algorithm that can generate a key pair from the CPMC encryption key. The current implementation of CPMC is a symmetric cryptography with a single key value, but the permutation functions are feasible to be utilized for an asymmetrical key values, so the proposed design expands the implementation to enable the usage of asymmetrical cryptography. For the implementation of cryptography, the encryption algorithm function of the CPMC technique, namely CPMC Shrinking, can be used as an encryption algorithm as well as decryption using the generated key pair. This works aims to:

- Explore an alternative implementation for the CPMC encryption technique as an asymmetric cryptography and its feasibility.
- Simplify the encryption-decryption processes by using a single algorithm function for running

both processes and analyzing its cryptographic performance improvements.

- Discover a different approach in implementing key generation and management for CPMC, by using different values of keys that can only encrypt and decrypt each other's messages, as opposed to the basic CPMC implementation as symmetric key algorithm.

II. PRELIMINARIES

A. *Asymmetric Key*

Asymmetric key cryptography is a cryptography algorithm in which the algorithms use a pair of different keys and use a different component of the pair for conducting encryption and decryption operations separately. Also widely known as Public Key Encryption [24]. For CPMC impelementation, both the encryption and decryption keys are presented as set of values, and both uses the same single function to conduct their respective operations. Here the usage of asymmetric pair of keys in CPMC algorithm is analyzed for feasibility of public key cryptography.

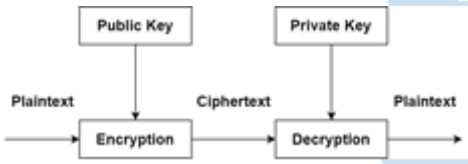


Fig. 1. Asymmetric Key Cryptography

B. *Chaotic Permutation Multicircular*

Chaotic Permutation Multicircular (CPMC) is a symmetric cryptographic technique utilizing permutation function with chaotic mathematical properties. The algorithm performs rotational permutations to the arrangement of array elements, accompanied by changes in the permutation space each iteration of the permutation process is complete. The CPMC technique consists of two cryptographic functions, CPMC Shrinking as the encryption function and CPMC Expanding as the decryption function. CPMC technique utilize its own key generator function dubbed Expanded Key Generator [18] [19], while the key value used by CPMC is an array of natural numbers with a length of N – 1 where N is the length of the plaintext. The value of the key element determines the shift distance of the permutation on the plaintext element, while the position of the key element value in the array determines which order the elements is permuted.

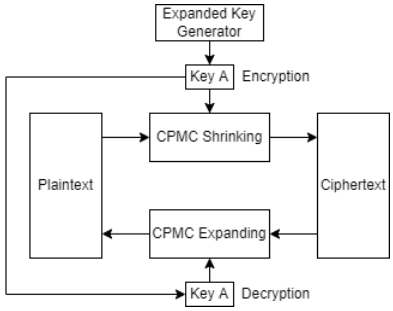


Fig. 2. Basic CPMC Encryption and Decryption Process

1) *CPMC Shrinking*

CPMC Shrinking (CPMCS) is a permutation function that acts as the encryption algorithm for CPMC. The permutation begins on all elements of the array, then the permutation is continued with the previous result with a range starting at the next index to the last index. The CPMCS permutation algorithm can be summarized in several steps [19]:

1. Determine the input set of plaintext X of size N, and the key array set Key of size N – 1.
2. Initiate the index loop iterator value n = 1.
3. Start of the (n)th Permutation by clockwise array rotation within range of (N – n + 1) on array X from its (n)th element to the last (X[n:N]), elements in the range are shifted as far as the (n)th element value of the key set array (Key[n]).
4. The first element of the result set of X is stored in Y[n], the stored element is not included in the next permutation.
5. Increment the loop iterator n = n + 1.
6. Check whether the value of n == N. If not, repeat step 4. If yes, continue to the next step.
7. The results of the remaining X permutations are stored at the last index Y (Y[N] = X[N]). Array Y is output as the result of CPMCS permutations.

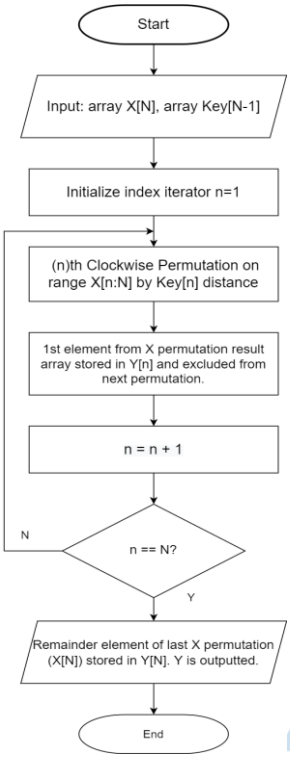


Fig. 3. CPMC Shrinking

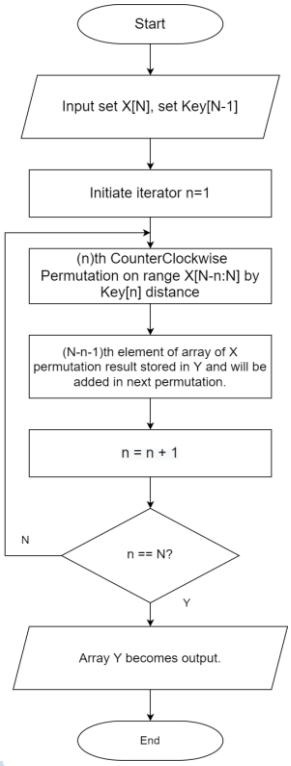


Fig. 4. CPMC Expanding

2) CPMC Expanding

CPMC Expanding (CPMCE) acts as a decryption function of the CPMC technique, the permutation process applied is the opposite of the CPMCS permutation. The CPMCE permutation algorithm can be summarized in several steps [19]:

1. Determine the input set of ciphertext X of size N, and the key array set Key of size N−1.
2. Initiate the index loop iterator value n = 1, with the last two elements X (X[N] and X[N – 1]).
3. Permuting the (n)th counterclockwise rotation with a range of elements of n + 1 on X from index N−n−1 to the last as far as the value of Key[n].
4. The (N − n − 1)th element of the result of the permutation X is stored in the Y set and becomes an additional element in the input of the next permutation.
5. Increment the loop iterator n = n + 1.
6. Check whether the value of n == N. If not, repeat step 4. If yes, continue to the next step.
7. Array Y results from the last iteration is output as the result of CPMCE permutation.

3) CPMC Expanded Key Generator

The expanded key generator algorithm for the CPMC technique serves as a generator for the permutation key value set for the CPMCS and CPMCE algorithms to determine the element shift in the permutation's multi-rotational movement process. This function uses three parameters of integer inputs, the first two are the initial key and sequence key as the main seed value parameters that determines how high the values of the key elements which determine the permutation shift distances. The third parameter is block size N to determine the key array size and the size limit of input that is eligible for the encryption and decryption process. The output of the function is an array of integers with the size of N−1 elements as the key. The keys generated from this function has a key space of (N!). The CPMC Expanded Key Generator algorithm can be summarized in several steps [19]:

1. Input the parameters values of 'Initial Key' and 'Sequence key' as the key seed, and N as the block size and the element permutation length.
2. Initiate iterator n = 1, and first element of array 'Temp' as Temp[n] = Initial Key.
3. Determine the value of the base modulus of the initial key 'Bi' as Bi[n] = (N – n + 1), and the base modulus of the sequence key 'Bs' as Bs[n] = (N – n + 1).
4. If initially Bi[n] is a prime number, then Bi[n] = 1, or if all prime factors of Bi[n] are the same value, then Bi[n]=Bi[n]/prime factor. If neither then the value of Bi[n] remains.

5. Calculate KeyI[n] as the modulus of Temp[n] with base Bi[n] (KeyI[n] = mod(Temp[n], Bi[n]))), and KeyS[n] as the modulus of the sequence key with base Bs[n] (KeyS[n] = mod(Sequence, Bi[n])).
6. Calculate Key[n] as the modulus of the sum KeyI[n] + KeyS[n] with base Bs[n].
7. Calculating the rounding down of the quotient Temp[n]/Bi[n], the result is added with KeyI[n] and stored as Temp[n].
8. Increment the iterator n = n + 1.
9. Check whether n == N, if not then the process returns to stage 3, if yes then the process continues to stage 10.
10. Array Key is outputted as the CPMC cryptographic key set.

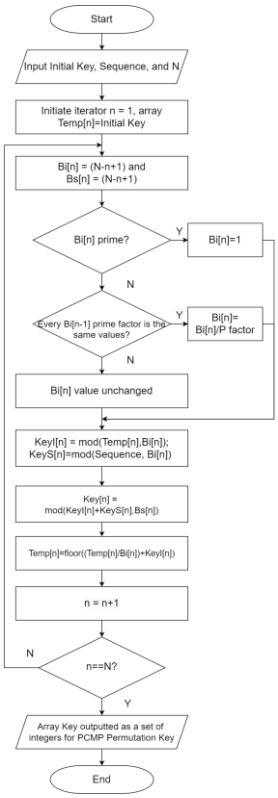


Fig. 5. CPMC Expanded Key Generator Function

III. Proposed Method

A. CPMC Asymmetric Implementation

The asymmetric implementation of CPMC can be described as such, if a CPMC function permute a set X using Key A into different arrangement resulting in set Y (Y=f(X,A)), then there is another set of values denoted as Key B that can revert Y to X using the same function. The encryption and decryption functions are implemented using the CPMCS function. From its permutative nature, the two keys in the pair can decrypt each other's permutations. The results of the encryption from the base key can be decrypted using the pair key, as well as the result of the pair key permutation can be decrypted using the base key (Figure 6).

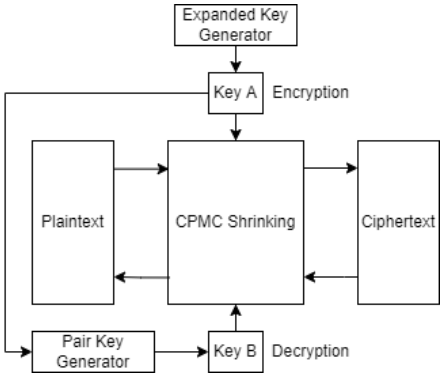


Fig. 6. CPMC Cryptography with Asymmetric Key

B. CPMC Asymmetric Pair Key Generation

The asymmetric key implementation of the CPMC technique uses a CPMC Permutation Key Finder algorithm as the pair key generator. The CPMC key finder algorithm is able to generate a CPMC key based on existing permutation patterns. If the CPMCS function permutes setA to setB using key KeyA, then the Key Finder function will return the reverser key dubbed KeyB that can permute setB back to setA using the CPMCS function as well. The CPMC Key Finder Algorithm generates an asymmetric pair key by using a set filled with unique values. The algorithm steps for the CPMC Key Pair Finder function, dubbed as "kFind", is described as such:

1. Get input setA as the initial set containing unique values of numbers up to the N length of block and setB as the permutation result of setA using KeyA. Declare iterator i as 0, and KeyB as an empty array initially for storing resulting key value.
2. Element of KeyB on KeyB[i] is incremented. Then KeyB is used to permute setB and the result is stored in Temp. (Temp=CPMCS(setB,KeyB)).
3. Check if the value of the Temp[i] element is equal to the setA[i] element. If not, repeat previous step until the element Temp[i] is equal to the value in setA[i].
4. If the element value Temp[i] == setA[i], then the iterator i is incremented to change the position index check of KeyB and setA to the next index. Then repeat step 3 until i reaches the last index of the key (i == N−1).
5. KeyB is outputted as the key pair of KeyA, which can permute setB back to setA.

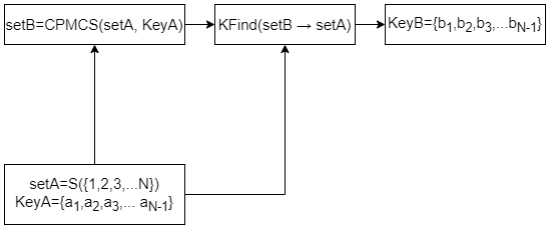


Fig. 7. CPMC Asymmetric Key Generation

IV. RESULTS

The algorithm design and test are applied using Matlab 2018b platform. The test is conducted in a computer with the specifications of i7-10750H 2.60GHz CPU, 16 GB of RAM and a 64-bit Windows 10 OS. The test consists of Pair Key Generation function, encryption and decryption test, and its runtime measurement to compare with the basic implementation of source method.

A. Pair Key Generation

The key generation algorithm is tested for an array length of 100 elements (N). For the test, the base key is generated using the PCMP Expanded Key Generator function with the input parameters of N=100, Initial Key = 12345678 and Sequence Key = 1234. The resulting base key contains the set of values: {12, 28, 14, 70, 53, 66, 79, 92, 13, 27, 41, 77, 69, 83, 11, 26, 41, 72, 71, 32, 49, 49, 1, 17, 33, 49, 65, 66, 25, 27, 59, 7, 25, 28, 61, 14, 33, 52, 9, 14, 49, 54, 31, 52, 17, 39, 7, 15, 1, 25, 49, 10, 37, 12, 41, 22, 5, 30, 19, 4, 37, 28, 21, 13, 13, 12, 13, 16, 21, 25, 7, 16, 5, 22, 15, 12, 13, 15, 5, 19, 17, 18, 13, 10, 5, 7, 5, 12, 1, 2, 7, 1, 3, 2, 5, 4, 3, 1, 0}.

Using Pair Key Finder, the base key is inputted and the resulting pair key is: { 19, 23, 36, 62, 67, 19, 26, 88, 67, 9, 23, 55, 61, 84, 0, 11, 34, 21, 53, 43, 41, 22, 74, 36, 23, 52, 45, 20, 65, 45, 54, 11, 9, 11, 2, 11, 32, 26, 61, 12, 6, 37, 47, 8, 55, 26, 46, 50, 15, 33, 11, 21, 26, 29, 45, 37, 5, 15, 41, 24, 11, 36, 20, 17, 20, 21, 17, 24, 27, 11, 27, 24, 15, 18, 22, 12, 7, 2, 16, 6, 13, 10, 17, 1, 15, 2, 11, 12, 4, 5, 8, 3, 6, 1, 3, 1, 2, 2, 1}.

B. Encryption & Decryption Function Test

The key pair generated is used for permutation testing of the encryption function against a text input string of 93 characters long including spaces, which has been padded with additional spaces to achieve a match with a key block size of 100 characters. The input text is stored in the 'plaintext' variable. Since CPMC uses permutation as its method for encryption and decryption, the encryption and decryption test using the asymmetric key pair shows the interchangeability of each key usage for either encryption or decryption function, with one key encryption can be decrypted using the other in the pair.

Encryption is applied using both keys in the pair, the encryption result of the base key (key1) is stored in the variable 'ciphertext1' (Figure 10), while the encryption result of the pair key (key2) is stored in the variable 'ciphertext2' (Figure 11). The decryption function test is carried out using the ciphertext generated from the encryption experiment, namely 'ciphertext1' and 'ciphertext2'. Decryption is performed using the PCMP key as opposed to the key used for encryption. 'ciphertext1' is decrypted using the key pair 'key2' (Figure 12), while 'ciphertext2' is decrypted using the base key 'key1'.

```
>> plaintext=pad(plaintext,length(key1)+1)
plaintext =
    'Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut.      '
>> plaintext-0
ans =
  Columns 1 through 17
    76   111   114   101   109    32   105   112   115   117   109    32   100   111   108   111   114
  Columns 18 through 34
    32   115   105   116    32    97   109   101   116    44    32    99   111   110   115   101    99
  Columns 35 through 51
   116   101   116   117   114    32    97   100   105   112   105   115    99   105   110   103    32
  Columns 52 through 68
   101   108   105   116    44    32   115   101   100    32   100   111    32   101   105   117   115
  Columns 69 through 85
   109   111   100    32   116   101   109   112   111   114    32   105   110    99   105   100   105
  Columns 86 through 100
   100   117   110   116    32   117   116    46    32    32    32    32    32    32    32
```

Fig. 8. Plaintext Input as String and ASCII Number

```
>> ciphertext2=PCMPK(plaintext,key2)
ciphertext2 =
  Columns 1 through 17
   105   112   110   115   108   101   101   100   115   105   100   111    76    32    32    32   116
  Columns 18 through 34
   105    99    32   101    32   109    32   105    97   100   109   101   115   105   111   111    46
  Columns 35 through 51
    32   111   105   112   105    32   117   111   100    32    32   105   101    44   115   109   110
  Columns 52 through 68
   110   100   115    32   117   114   114   116   117   116   115    32   116   101    97   105    32
  Columns 69 through 85
    44   105   112   100    32   116   111    99    32   117    32   116   116    32   116   117   109
  Columns 86 through 100
   114    32    99    99   109   101   111   114   101   108   100    32   103   110    32
>> sprintf(char(ciphertext2))
ans =
  'ipnsleedsidoL   tic e m iadmesioo. oipi uod  ie,smnnds urrtuts teai ,ipd toc u tt tumr ccmeoreld gn '
```

Fig. 9. Pair Key Encryption

```
>> decrypted2=PCMPK(ciphertext2,key1)
decrypted2 =
  Columns 1 through 17
    76   111   114   101   109    32   105   112   115   117   109    32   100   111   108   111   114
  Columns 18 through 34
    32   115   105   116    32    97   109   101   116    44    32    99   111   110   115   101    99
  Columns 35 through 51
   116   101   116   117   114    32    97   100   105   112   105   115    99   105   110   103    32
  Columns 52 through 68
   101   108   105   116    44    32   115   101   100    32   100   111    32   101   105   117   115
  Columns 69 through 85
   109   111   100    32   116   101   109   112   111   114    32   105   110    99   105   100   105
  Columns 86 through 100
   100   117   110   116    32   117   116    46    32    32    32    32    32    32    32
>> sprintf(char(decrypted2))
ans =
  'Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut.      '
```

Fig. 10. Base Key as Description Key

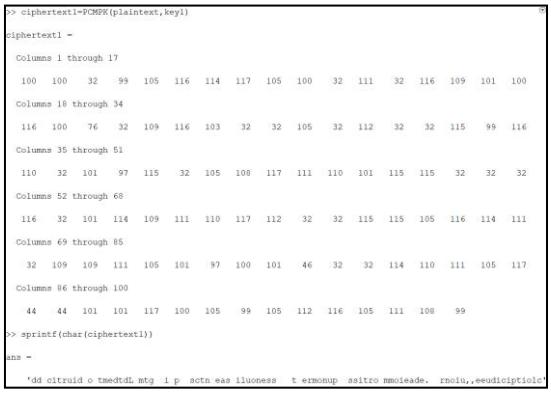


Fig. 11. Base Key Encryption

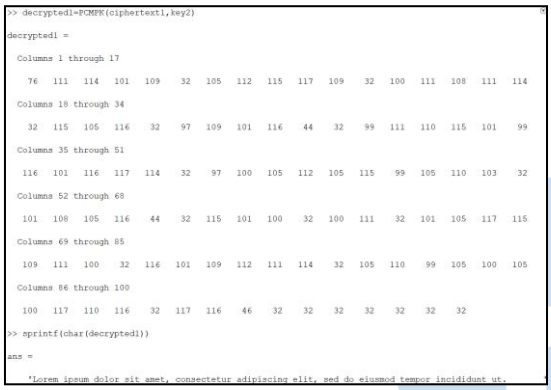


Fig. 12. Pair Key Description

C. Encryption & Decryption Runtime

In testing the CPMC cryptographic algorithm, runtime measurements of the encryption and decryption functions were also carried out in Matlab. Measurements were made during the running of the proposed asymmetric implementation algorithm program and the basic implementation as symmetric cryptography, both of which were also run with different input lengths (N) starting from 100 to 1000 elements in size. The basic implementation process uses the CPMC Shrinking and Expanding permutation algorithm separately as encryption and decryption functions, respectively, using the key generated by the Expanded Key Generator function, while the asymmetric implementation uses CPMC Shrinking as the encryption and decryption function, where the encryption process uses the same key (Expanded Key) as in the basic implementation, and the decryption process uses the key (Pair Key) generated by the KFind function, so that the runtime measured is one time encryption process using CPMCS with two decryption processes using the CPMCE and CPMCS functions respectively along with their difference ratio percentage.

TABLE I. RUNTIME OF ENCRYPTION AND DECRYPTION TEST BY PERMUTATION LENGTH (N)

N	Runtime (s)			*Runtime Difference (%)*
	CPMCS Encryption	***CPMCE***	***CPMCS Decryption***	
100	1.11E-04	5.13E-04	1.12E-04	78.18
200	2.22E-04	9.76E-04	2.13E-04	78.2
300	2.87E-04	1.28E-03	2.69E-04	76.98
400	4.44E-04	1.78E-03	4.41E-04	75.21
500	4.91E-04	2.28E-03	4.95E-04	78.29
600	6.41E-04	2.56E-03	6.42E-04	74.94
700	7.84E-04	3.02E-03	7.57E-04	74.94
800	9.72E-04	3.48E-03	8.75E-04	74.85
900	1.13E-03	3.94E-03	1.01E-03	74.32
1000	1.14E-03	4.19E-03	1.14E-03	72.81
Average				75.87

TABLE II. RUNTIME COMPARISON WITH SOURCE DATA AT N=300

Test	Runtime (s)		
	Encryption	***Decryption***	***Total***
CPMCS + CPMCE Test	2.87E-04	1.28E-03	1.57E-03
Asymmetric Key CPMCS	2.87E-04	2.69E-04	5.57E-04
CPMC Source [19]	3.08E-03	3.22E-03	6.30E-03

The runtime measurement results in Table I show that the speed of the CPMCS function both as an encryption and decryption function runs faster than CPMCE, where the use of CPMCS as a decryption function increases the decryption runtime by an average of 75.87%. Then the measured runtimes were compared with data from literature sources that recorded measurements at N=300 (Table 2). This comparison is used to show the difference between the results of the basic implementation test and the proposed method with the source data, in which the Matlab test version runtime are found to be faster for both basic and proposed implementation compared to the source data. The basic implementation CPMC encryption and decryption runtime on the running test in Matlab give results of a total of 1.57E-03 seconds, while the source data gives a total of 6.30E-03 seconds. The proposed asymmetric implementation results in a total runtime of 5.57E-04 seconds, which is an increase of 75.16% compared to basic CPMC from the Matlab test and by 91.17% from the source data.

D. Public Key Feasibility

The permutation key and CPMC Shrinking algorithm properties made it so that a certain key set values are only pairable with a specific key set value so that one key's encryption result can only be decrypted using the other key using the same permutation

function. However, since the base permutation algorithms were meant as a symmetric cryptographic function, the leakage of one pair of keys can compromise the secrecy of the encrypted data content. A modification on the permutation step and encryption process while without requiring to modify the key value can add better secrecy.

V. CONCLUSIONS

CPMC technique can apply cryptography asymmetrically using key pairs that can perform decryption using the same algorithm for encryption, namely CPMC Expanding, where the results of one key encryption permutation can be decrypted using another key in the pair. The implementation of asymmetric keys in the CPMC technique can simplify the work of the algorithm by using the same function to perform encryption and decryption, and produce better performance at the decryption speed than the basic CPMC technique. This means the speed of the CPMC Shrinking algorithm is found to be faster than the CPMC Expanding, so the implementation of asymmetric key cryptography for CPMC can be used as a more efficient alternative for fast encryption and decryption method using a pre-generated key. The use of CPMC Shrinking as a decryption function results in an average decryption speed increase of 75.87% from testing with different iterations of block size.

REFERENCES

[1] M. Kumar, A. Saxena, and S. S. Vuppala, "A Survey on Chaos Based Image Encryption Techniques," in Multimedia Security Using Chaotic Maps: Principles and Methodologies, K. M. Hosny Ed. Cham: Springer International Publishing, pp. 1-26, 2020.

[2] M. T. Suryadi, Y. Satria, and M. Fauzi, "Implementation of digital image encryption algorithm using logistic function and DNA encoding," Journal of Physics: Conference Series, vol. 974, p. 012028, 2018/03 2018, doi: 10.1088/1742-6596/974/1/012028.

[3] M. Wang, X. Wang, Y. Zhang, S. Zhou, T. Zhao, and N. Yao, "A novel chaotic system and its application in a color image cryptosystem," Optics and Lasers in Engineering, vol. 121, pp. 479-494, 2019, doi: https://doi.org/10.1016/j.optlaseng.2019.05.013

[4] A. M. Elshamy, A. I. Hussein, H. F. A. Hamed, M. A. Abdelghany, and H. M. Kelash, "Color Image Encryption Technique Based on Chaos," Procedia Computer Science, vol. 163, pp. 49-53, 2019/01/01/ 2019, doi: https://doi.org/10.1016/j.procs.2019.12.085.

[5] Y. Liu, J. Zhang, D. Han, P. Wu, Y. Sun, and Y. S. Moon, "A multidimensional chaotic image encryption algorithm based on the region of interest," Multimedia Tools and Applications, vol. 79, no. 25, pp. 17669-17705, 2020/07/01 2020, doi: 10.1007/s11042-020-08645-8.

[6] S. Zhou, X. Wang, M. Wang, and Y. Zhang, "Simple colour image cryptosystem with very high level of security," Chaos, Solitons & Fractals, vol. 141, p. 110225, 2020/12/01/ 2020, doi: https://doi.org/10.1016/j.chaos.2020.110225.

[7] Y. Naseer, T. Shah, and D. Shah, "A novel hybrid permutation substitution base colored image encryption scheme for multimedia data," Journal of Information Security and Applications, vol. 59, p. 102829, 2021/06/01/ 2021, doi: https://doi.org/10.1016/j.jisa.2021.102829.

[8] L. Teng, X. Wang, and Y. Xian, "Image encryption algorithm based on a 2D-CLSS hyperchaotic map using simultaneous permutation and diffusion," Information Sciences, vol. 605, pp. 71-85, 2022/08/01/ 2022, doi: https://doi.org/10.1016/j.ins.2022.05.032.

[9] E. A. Albahrani and T. K. Alshekly, "New Chaotic Substation and Permutation Method for Image Encryption," International Journal of Applied Information Systems, vol. 12, pp. 33-39, 2017.

[10] J. I. Moreira Bezerra, V. Valduga de Almeida Camargo, and A. Molter, "A new efficient permutation-diffusion encryption algorithm based on a chaotic map," Chaos, Solitons & Fractals, vol. 151, p. 111235, 2021/10/01/ 2021, doi: https://doi.org/10.1016/j.chaos.2021.111235.

[11] R. S. Devi, K. Thenmozhi, R. Amirtharajan, and P. Padmapriya, "A Novel Multiple Segmented Image Encryption," in 2019 International Conference on Computer Communication and Informatics (ICCCI), 23-25 Jan. 2019 2019, pp. 1-5, doi: 10.1109/ICCCI.2019.8822155.

[12] M. Wang, X. Wang, Y. Zhang, and Z. Gao, "A novel chaotic encryption scheme based on image segmentation and multiple diffusion models," Optics & Laser Technology, vol. 108, pp. 558-573, 2018/12/01/ 2018, doi: https://doi.org/10.1016/j.optlastec.2018.07.052.

[13] S. Sun, Y. Guo, and R. Wu, "A Novel Image Encryption Scheme Based on 7D Hyperchaotic System and Row-column Simultaneous Swapping," IEEE Access, vol. 7, pp. 28539-28547, 2019, doi: 10.1109/ACCESS.2019.2901870.

[14] A. A. Karawia and Y. A. Elmasry, "New Encryption Algorithm Using Bit-Level Permutation and Non-Invertible Chaotic Map," IEEE Access, vol. 9, pp. 101357-101368, 2021, doi: 10.1109/ACCESS.2021.3096995.

[15] I. Mishkovski and L. Kocarev, "Chaos-Based Public-Key Cryptography," in Chaos-Based Cryptography: Theory, Algorithms and Applications, L. Kocarev and S. Lian Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 27-65, 2011.

[16] D. Yoshioka, "Security of Public-Key Cryptosystems Based on Chebyshev Polynomials Over Z/pk Z," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 67, no. 10, pp. 2204-2208, 2020, doi: 10.1109/TCSII.2019.2954855.

[17] V. M. Silva-García, R. Flores-Carapia, M. D. González-Ramírez, E. Vega-Alvarado, and M. G. Villarreal-Cervantes, "Cryptosystem Based on the Elliptic Curve With a High Degree of Resistance to Damage on the Encrypted Images," IEEE Access, vol. 8, pp. 218777-218792, 2020, doi: 10.1109/ACCESS.2020.3042475.

[18] Y. Suryanto, Suryadi and K. Ramli, "Chaos properties of the Chaotic Permutation generated by Multi Circular Shrinking and Expanding Movement," 2015 International Conference on Quality in Research (QiR), 2015, pp. 65-68, doi: 10.1109/QiR.2015.7374896.

[19] Y. Suryanto, "Pengembangan dan Analisis Metode Permutasi Chaotic Baru Berbasis Multiputaran Mengecil dan Membesar untuk Enkripsi Citra dengan Tingkat Keamanan Tinggi, Cepat dan Tahan Terhadap Gangguan," Program Doktor [Dissertation], Universitas Indonesia, 2016.

[20] Y. Suryanto, Suryadi, and K. Ramli, "A Secure and Robust Image Encryption Based on Chaotic Permutation Multiple Circular Shrinking and Expanding," J. Inf. Hiding Multim. Signal Process., vol. 7, pp. 697-713, 2016.

[21] Y. Suryanto, Suryadi, and K. Ramli, "A new image encryption using color scrambling based on chaotic permutation multiple circular shrinking and expanding," Multimedia Tools and Applications, vol. 76, no. 15, pp. 16831-16854, 2017, doi: 10.1007/s11042-016-3954-5.

[22] K. Ramli, Y. Suryanto, Magfirawaty, and N. Hayati, "Novel Image Encryption Using a Pseudoset Generated by Chaotic Permutation Multicircular Shrinking With a Gradual Deletion of the Input Set," IEEE Access, vol. 8, pp. 110351-110361, 2020, doi: 10.1109/ACCESS.2020.3001949.

[23] N. Hayati, Y. Suryanto, K. Ramli, and M. Suryanegara, "End-to-End Voice Encryption Based on Multiple Circular Chaotic Permutation," in 2019 2nd International Conference on Communication Engineering and Technology (ICCET), 12-15 April 2019 2019, pp. 101-106, doi: 10.1109/ICCET.2019.8726890.

[24] R. Shirey, "Internet Security Glossary, Version 2", FYI 36, RFC 4949, doi: 10.17487/RFC4949, August 2007. Network Working Group [Internet]. Available: https://www.rfc-editor.org/info/rfc4949.

ISSN 2085-4552

Topic Modelling Using VSM-LDA For Document Summarization

Luthfi Atikah[1], Novrindah Alvi Hasanah[2], Agus Zainal Arifin[2]
[1] Management of Informatic: Politeknik Astra, Kabupaten Bekasi, Indonesia
[2] Departement of Informatic: Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia
[3] Departement of Informatic: Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[1]luthfiatikah@polythecnic.astra.ac.id , [2]novrindah@uin-malang.ac.id, [3]agusza@cs.its.ac.id

Accepted 01 November 2022
Approved 24 December 2022

***Abstract*— Summarization is a process to simplify the contents of a document by eliminating elements that are considered unimportant but do not reduce the core meaning the document wants to convey. However, as is known, a document will contain more than one topic. So it is necessary to identify the topic so that the summarization process is more effective. Latent Dirichlet Allocation (LDA) is a commonly used method of identifying topics. However, when running a program on a different dataset, LDA experiences "order effects", that is, the resulting topic will be different if the train data sequence is changed. In the same document input, LDA will provide inconsistent topics resulting in low coherence values. Therefore, this paper proposes a topic modelling method using a combination of LDA and VSM (Vector Space Model) for automatic summarization. The proposed method can overcome order effects and identify document topics that are calculated based on the TF-IDF weight on VSM generated by LDA. The results of the proposed topic modeling method on the 1300 Twitter data resulted in the highest coherence value reaching 0.72. The summary results obtained Rouge 1 is 0.78, Rouge 2 is 0.67 dan Rouge L is 0.80.**

***Index Terms*— LDA; Order Effects; Summarization; Topic Modelling; VSM-LDA**.

I. INTRODUCTION

The rapid development of the internet has made the information circulating on the internet also increasing. This makes it difficult for information seekers to conclude what news happened. So, needs a way to find useful information efficiently. Summarization can help readers quickly understand the themes and concepts of the entire document, and effectively save time reading.

Summarization is a process to simplify the contents of a document or text by eliminating elements that are deemed unimportant but do not reduce the core meaning wants to convey. However, as it is known, a document will contain more than one topic. So it is necessary to search for topics so that the summarization process is more effective. To find out the topics contained in a document, there have been many methods used to formulate a document making it easier for readers to find out important information contained in a document based on that topic. One of the most commonly used methods is LDA. LDA is a model that calculates the probability of each word on a random topic [1]. LDA has two stages, namely modelling each word into the topic and calculating the probability of each topic repeatedly.

Some studies like [2][3] use the LDA method to determine the topic as the basis for automatic summarization. However, in the process of running a program on a different dataset, LDA experiences an "order effect", that is, the resulting topic will be different if the training data sequence is changed [4]. Such sequencing effects lead to systematic errors for any study in text mining. Then the resulting topic becomes inaccurate. To solve this problem, this paper proposes a Vector Space Model (VSM). This is based on several previous studies that used VSM as a method of retrieval of information on documents such as research conducted by [5][6]. In the VSM method, several online documents will be indexed and sorted based on the weight of search words contained in online documents using the TF-IDF algorithm. In the process, if there is other data that is executed in the LDA program it will still be processed properly because the data train is not randomized, but the sequence has been determined by word weight using the TF-IDF algorithm on VSM [5]. It aims to produce relevant topics based on modelling topics.

However, before doing topic modelling with VSM-LDA, in this study data clustering was carried out which aims to retrieve data in certain clusters so that it can be processed in the next process. Based on [7] using sentence clustering before doing topic modelling for automatic summarization of documents. This study determines the threshold of a particular cluster so that not all clusters are taken. The threshold used is a cluster with a minimum number of members of 110 data. Many clustering methods have been developed so far. This study uses the k-means clustering method. This is based on previous research which is used as a reference such as research conducted by [8] [9][10] which shows that effective clustering is used for data clustering techniques.

Furthermore, the summarization method used in this study is TextRank. The TextRank method is a graphically based method that sorts words in commonly

used text processing [11]. TextRank is used in several studies including [12] performing tweet text summarization to analyze user interest. This method generates a rank for each word which is then mapped to a data set of user interests. Several other studies previously used TextRank as a summarization method, such as in the study [11][13] which improved the TextRank method as a summary method.

This study focuses on the performance of LDA which has an order effect. In the same document input, LDA will provide inconsistent topics resulting in low coherence values. Therefore, this paper proposes a topic modelling method using a combination of LDA and VSM for automatic summarization. Our proposed method can overcome "order effects" so that it can identifies the document topics that are calculated based on the TF-IDF weights on the VSM generated by the LDA. Our topic modeling method can find relevant topics in documents that can produce a high rouge score on the summary results.

II. Related Work

Several previous works have referred us to using this method. In this study we used k-Means as a clustering method before determining topics using topic modelling with LDA. A study, [8] conducted k-means grouping for Al-Qur'an Verses Grouping using the K-Means Algorithm. Another study from [9] used k-Means grouping to apply to the results of modelling topics using LDA to summarize document grouping.

This study combines VSM and LDA. The LDA reference used is based on previous research [3]. Computation of text similarity based on the LDA topic model and coincidence Word is carried out to analyze the semantic correlation of text themes, the results of this study, indicate that LDA will cover text topics better if processed together with coincidence Word. Another study [2] used LDA to perform document summarization, they proposed punitive-based LDA to be summarized. In that study, LDA showed good results. Another study, the LDA method is also used in analyzing public sentiment in various industrial fields, including economics in terms of product marketing [14]. Meanwhile, VSM has shown good performance in word weighting using the TF-IDF algorithm based on previous research conducted by [5]. Then, [6] did study using VSM to query similarity indexing and the results show that each word in the document is weighted so that it can be sequential according to weight.

To summarize the document, this study uses the TextRank for summarization method. This is based on research [11] which applies Text Rank for Automatic Summarization, and the summarization results show that the TextRank method is better than BM25. Meanwhile, another study [13] compared the summary results using TF-IDF and TextRank which showed that the summary results with TextRank were better than TF-IDF for the precision, recall and f-measure values.

III. Proposed Method

To achieve the desired results, this research will go through several stages, namely data preparation, data clustering, topic modelling, and the last id summerizing the document as shown in Figure 1. The following subsections present the explanation of the stages in detail.

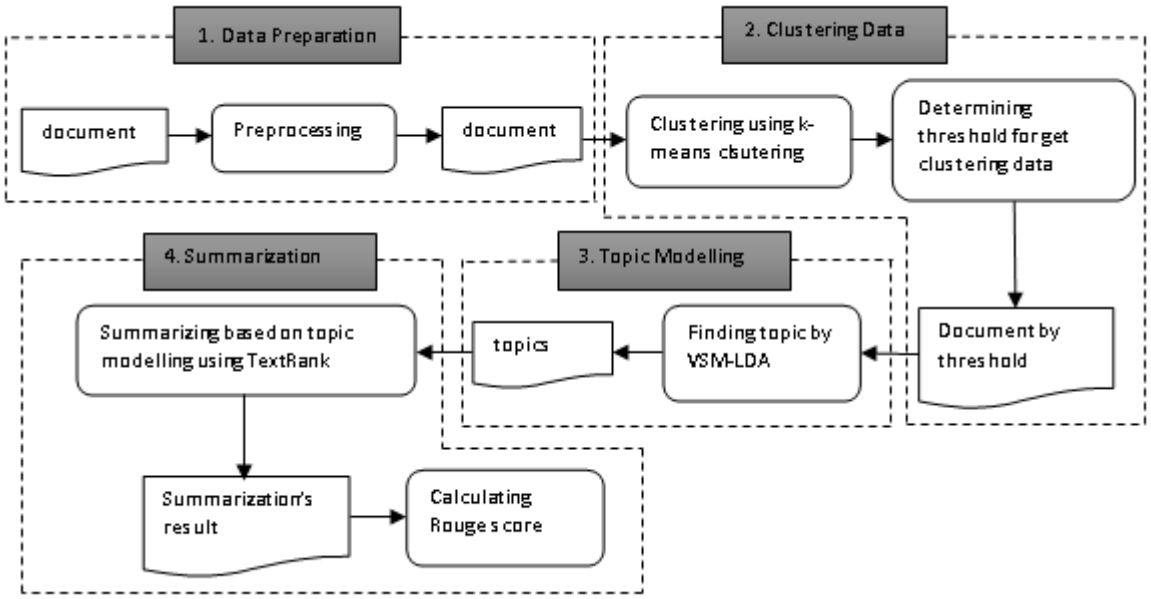


Fig. 1. Research Method

TABLE I. SAMPLE DATA

No	Original data	Clean data
1.	@MeltingIce Assuming max acceleration of 2 to 3 g's, but in a comfortable direction. Will feel like a mild to moder? https://t.co/fpjmEgrHfC	assuming max acceleration of to gs but in a comfortable direction will feel like a mild to moder
2.	Tesla Semi truck unveil & test ride tentatively scheduled for Oct 26th in Hawthorne. Worth seeing this beast in person. Unreal.	tesla semi-truck unveil test ride tentatively scheduled for oct th in hawthorne worth seeing this beast in person unreal
3.	@Nickg_uk @Model3Owners Feature coming soon	feature coming soon

A. *Data Preparation*

This study uses an experimental object in the form of a dataset available online from the Kaggle website. The dataset is a collection data from the Elonmusk Twitter timeline of around 1.300 data which contains automotive issues. The column 'Text' in the dataset becomes the main object that is summarized. The first step in experimenting is cleaning the dataset in the preprocessing process to remove elements that are not needed in the next process to get a summary. Noise in data such as punctuation marks, numbers, and upper letters are replaced with lowercase letters so that the data obtained is cleaned. Data that has been cleaned becomes data that is ready for use in the next process. Table 1 shows an example of original data from Elon Musk that has been preprocessed to obtain clean data.

B. *Clustering Data*

Data that had been cleaned from preprocessing were grouped using k-means clustering. In k-means clusters, the resulting word classes are grouped in semantic similarity under the Euclidean metric boundary. In this study, the threshold is used to take the results of the cluster clustering to be processed at the topic modelling and summarization stage. The k-means algorithm procedure is presented as follows:

1. Step 1. Determine the cluster *K* value. This study uses the value of $K = 7$.
2. Step 2. Allocate data into clusters randomly.
3. Step 3. Calculate the average centroid in each cluster.
4. Step 4. Allocate each data to the nearest centroid/average.
5. Step 5. Return to Step 3 if there are still changes to the centroid of the data cluster has been moved. And stop when nothing changes.
6. Step 6. Count the amount of data in each cluster.
7. Step 7. Determine the threshold value.
8. Step 8. Get the data according to the threshold value.

C. *Topic Modelling*

The vector space model (VSM) is a method used to represent documents as spatial vectors and calculate the similarity between vectors to measure the similarity between documents [3]. VSM is generally run with the TF-IDF (Term frequency-inverse document frequency) algorithm to weight words. This study uses VSM which is processed using the TF-IDF algorithm. We process VSM with Latent Dirichlet Allocation (LDA) as a method of modelling topics so that the topics contained in the document can be found with high coherence values. Meanwhile, LDA is a topic modelling method. Topic modelling aims to find topics automatically in the data or corpus. LDA is based on a Bayesian probabilistic model in which each topic has a separate word probability distribution, and each document consists of a mixture of topics [3]. The basic idea of LDA is that a document is represented as a random mix of latent (invisible) topics [15]. The VSM-LDA we use is shown in Figure 2 below. Cospus document are processed using VSM run with the TF-IDF algorithm. Each word in the document is weighted with TF-IDF weight in VSM before being processed to identify its topic.

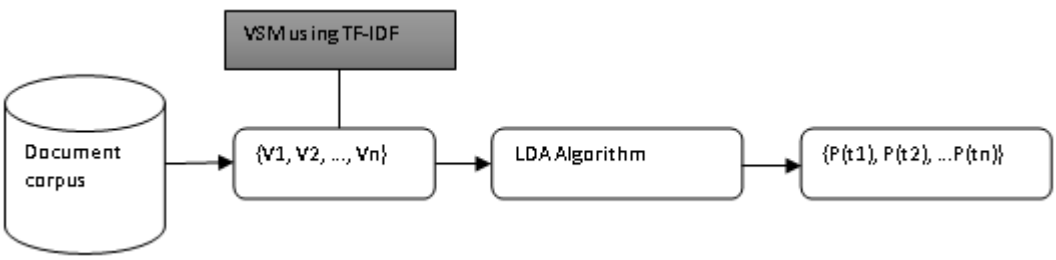


Fig. 2. VSM+LDA Algorithm

D. Summarization

The results of the modelling topic were summarized using the TextRank algorithm as the method used to summarize documents. Each sentence is ranked via the TextRank algorithm. Based on the similarity index, important sentences are selected. Term Frequency - Inverse Document Frequency is used to calculate the importance of terms in each sentence in the input document. According to [10] research, the TextRank algorithm will work in summarizing a document, that is, each sentence in the document represents a node and vertex that represents the similarity. After that, to get the sentence value calculation from a vertex, all the similarity values of the edges connected to each node are added. Sentence values of a primary node will be compared to all existing sentence nodes, sentence nodes that have the same sentence values as a primary node rank them highest because they have similarities.

IV. Result and Discussion

Automatic summarization is done based on the results of clustering using K-means clustering. In this experiment, data from the Elonmusk Twitter timeline, which is available online on the Kaggle website, becomes the object for system testing. The value of *K* used is *K* = 7. The value of *K* is determined based on fine tuning the best *k* on the dataset. This study uses a threshold value to retrieve data at a certain value for processing at the topic modelling stage. Cluster data to be processed on the modelling topic is a cluster with data more than 110 data. The clustering results are shown in Table 2. In the cluster results, cluster data from *n* = 1 and *n* = 3 are taken and used in the next process because the two clusters have data more than 110 data.

TABLE II. The Result Of Clustering Data Using K-Means Clustering

Cluster –n	Count
0	108
1	894
2	45
3	117
4	71
5	66
6	21

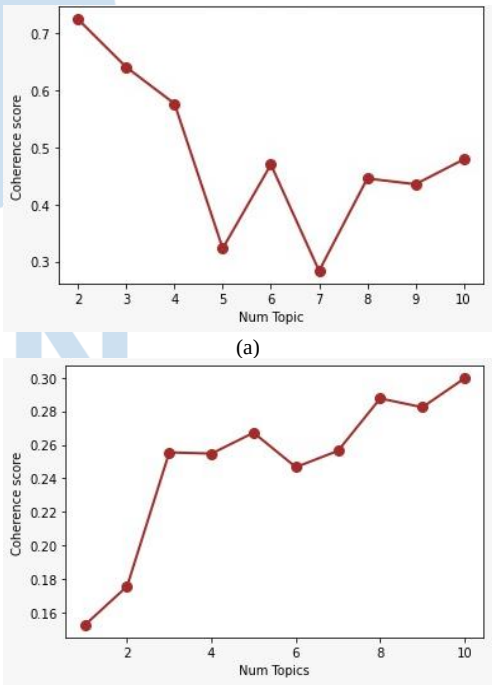
The runtime measurement results in Table I show After getting the cluster data that has been determined using the threshold value, the data is processed to search for topics using topic modelling. At this stage, the research conducted two experiments, namely by using VSM-LDA and LDA to compare how the results of the coherent value of the resulting topic. The coherent topics obtained by the VSM-LDA and LDA shown in Figure 2.

Figure 3 part (a) shows that the topics obtained from VSM-LDA and LDA both get 10 topics for the dataset used in the experiment but with different coherent values. The coherent value obtained by the VSM-LDA method is higher than the coherent value obtained by the LDA method. After getting the topic of the document, the next step is to summarize the document. The summarization trial was conducted twice. The first experiment was carried out using the proposed method, namely k-Means + VSM-LDA + textRank. The second experiment uses the k-means sampling method + LDA + textRank. The results of the two summaries are presented in Table 3 which shows the rouge of each summarized result.

TABLE III. Rouge Score

No	Method	Rouge-1	Rouge-2	Rouge-L
1	K-means+VSM-LDA+ textRank	0.78	**0.67**	**0.80**
2	K-means+LDA+ textRank	0.61	0.48	0.66

Table 3 shows that the scores for Rouge 1, Rouge 2, and Rouge 3 show that the proposed summarization method using k-means clustering + VSM-LDA + TextRank shows better results than the comparison method using k-means clustering + LDA + TextRank.



(a)

(b)

Fig. 3. Topic Coherence by (a) VSM-LDA (b) LDA

Topic modelling using LDA has become the most frequently used method of identifying topics. But in reality, as revealed in research from [5], when running a program on a different dataset, LDA will experience an "order effect", that is, the resulting topic will be different if the training data sequence is changed [4]. Such a sequence effect causes systematic errors for any study, for example text classification and grouping. So that the resulting topics are inaccurate and reduce the

effectiveness of text mining being carried out. This affects the coherent topic score. From the analysis, LDA is suitable for large-scale data because LDA can reduce dimensions. In addition, with LDA data is mapped into topics so that the relationship of each topic can be found.

Meanwhile, VSM is a method that is run based on the TF-IDF algorithm where each word is sorted based on the weight of search words contained in the document [5]. VSM can cover the shortcomings of LDA which often experience order effects so that the resulting topic has a low coherence value. In the process, if there is other data that will be executed in the LDA program, it will still be processed properly because the train data order will not be random, but the sequence has been determined by word weight which has previously been processed by VSM using the TF-IDF algorithm.

V. CONCLUSIONS

This is indicated by the coherence score obtained in the experiment which shows that the coherence score with VSM-LDA shows a good value when compared to the coherence score in the LDA method. The output of Latent Dirichlet Allocation (LDA) is a topic defined by K. Topic coherence is a measure used to evaluate modelling topics based on the top words, the higher the coherence score obtained, the higher the interpretation of the words in the modelling topic [16].

The results of modelling topics with good coherence values become the basis for summarization. This study uses the TextRank method. The results showed that the scores of Rouge 1, Rouge 2, and Rouge L which were calculated based on the summary results showed that the results of the summary based on the results of topic modelling using the proposed method, namely K-means clustering + VSM-LDA + TextRank had a higher value when compared to the summary results using the K-means clustering method + LDA + TextRank. This is influenced by the topics identified using modelling topics with good coherence. The better the coherence value, the better the summary results.

REFERENCES

[1] L. Hagen, "Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models" *Inf Process Manag*, vol. 54, no. 6, pp. 1292–1307, 2018, doi: 10.1016/j.ipm.2018.05.006.

[2] Y. L. Chang and J. T. Chien, "Latent Dirichlet learning for document summarization," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, no. April 2009, pp. 1689–1692, 2009, doi: 10.1109/ICASSP.2009.4959927.

[3] M. Shao and L. Qin, "Text Similarity Computing Based on LDA Topic Model and Word Co-occurrence," no. Sekeie, pp. 199–203, 2014, doi: 10.2991/sekeie-14.2014.47.

[4] A. Agrawal, W. Fu, and T. Menzies, "What is wrong with topic modeling? And how to fix it using search-based software engineering," *Inf Softw Technol*, vol. 98, pp. 74–88, 2018, doi: 10.1016/j.infsof.2018.02.005.

[5] Irmawati, "SISTEM TEMU KEMBALI INFORMASI PADA DOKUMEN DENGAN Irmawati," *Jurnal Ilmiah FIFO*, vol. IX, no. 1, pp. 74–80, 2017.

[6] P. K. Reshma, S. Rajagopal, and V. L. Lajish, "A Novel Document and Query Similarity Indexing using VSM for Unstructured Documents," *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, pp. 676–681, 2020.

[7] I. Lukmana, D. Swanjaya, A. Kurniawardhani, A. Z. Arifin, and D. Purwitasari, "Sentence Clustering Improved Using Topic Words," *Juti*, pp. 1–8, 2014.

[8] C. Slamet, A. Rahman, M. A. Ramdhani, and W. Dharmalaksana, "Clustering the Verses of the Holy Qur'an using K-means Algorithm," *Asian Journal of Information Technology*, vol. 15, no. 24, pp. 5159–5162, 2016.

[9] E. Y. Hidayat, F. Firdausillah, K. Hastuti, I. N. Dewi, and Azhari, "Automatic text summarization using latent drichlet allocation (LDA) for document clustering," *International Journal of Advances in Intelligent Informatics*, vol. 1, no. 3, pp. 132–139, 2015, doi: 10.26555/ijain.v1i3.43.

[10] M. Cha, Y. Gwon, and H. T. Kung, "Language modeling by clustering with word embeddings for text readability assessment," *International Conference on Information and Knowledge Management, Proceedings*, vol. Part F1318, pp. 2003–2006, 2017, doi: 10.1145/3132847.3133104.

[11] F. Barrios, F. López, L. Argerich, and R. Wachenchauzer, "Variations of the Similarity Function of TextRank for Automated Summarization," 2016.

[12] R. Niu and B. Shen, "Microblog User Interest Mining Based on Improved TextRank Model," *J Comput (Taipei)*, vol. 30, no. 1, pp. 42–51, 2019.

[13] N. Kumari and P. Singh, "Automated Hindi Text Summarization Using Tf-Idf and Textrank Algorithm," vol. 7, no. 17, pp. 2547–2555, 2020.

[14] X. Liu, A. C. Burns, and Y. Hou, "An Investigation of Brand-Related User-Generated Content on Twitter," *J Advert*, vol. 46, no. 2, pp. 236–247, 2017.

[15] Z. Tong and H. Zhang, "A Text Mining Research Based on LDA Topic Modelling," *Computer Science & Information Technology (CS & IT)*, pp. 201–210, 2016.

[16] F. Rosner, A. Hinneburg, M. Röder, M. Nettling, and A. Both, "Evaluating topic coherence measures," no. December, pp. 0–4, 2014.

ISSN 2085-4552

Crude Oil Price Forecasting Using Long Short-Term Memory and Support Vector Regression

Rifdah Amelia[1], Ahmad Zuhdi[2], Abdul Rochman[3]
[1, 2, 3] Informatics Engineering Department, Trisakti University, Jakarta, Indonesia
[1]rifdah064001900019@std.trisakti.ac.id, [2]zuhdi@trisakti.ac.id, [3]abdul.rochman@trisakti.ac.id

Accepted 02 November 2022
Approved 06 December 2022

***Abstract*— Crude oil or petroleum is a non-renewable energy source derived from organic materials whose formation process is lengthy. Crude oil is a commodity whose price often fluctuates. When there is a fluctuation, a nation's economy will be affected. The crude oil price datasets are categorized as non-linear. This research used two models to compare the performance of those two models to find the best model to predict Brent crude oil prices. The models used in this research are Long Short-Term Memory (LSTM) and Support Vector Regression (SVR). Those two methods are widely used for similar cases, such as forecasting the stock price. The dataset used in this study is the price of Brent crude oil from May 1987 to May 2022. The result of this study indicates that the deep learning algorithm, LSTM, performs better in forecasting the price of Brent crude oil, with a root mean squared error value of 1.543.**

***Index Terms*— Crude Oil; Deep Learning; Forecasting; Long Short-Term Memory (LSTM); Support Vector Regression (SVR)**.

I. INTRODUCTION

Crude oil is a non-renewable energy source. Crude oil is categorized as a non-renewable energy source because the formation process takes quite a long time [1]. Crude oil reached through the fractionation process will provide various final products, such as fuel, liquefied gas, diesel, and others [1].

Indonesia is a country rich in natural resources. As a country rich in natural resources, Indonesia is among the countries capable of producing crude oil. In the early 2022 period, Indonesia's crude oil consumption level fluctuated substantially. However, it tends to increase when the level of crude oil consumption in Indonesia reaches 1,710.52 thousand barrels per day [2].

From December 2021 to January 2022, there has been a decline in the level of crude oil production. Indonesia can only produce 651.68 thousand barrels per day, while in January 2022, Indonesia only produced 616.06 thousand barrels of crude oil per day [3].

The data presented in the previous paragraph shows that there has been a plunge in the level of crude oil production in Indonesia. On the contrary, an increase in crude oil consumption impacts the government because the government needs to import petroleum from other countries.

Petroleum is a mining product that has sensitive properties [4]. So, when fluctuations emerge, a country's economy (both micro and macro) can be affected. Consequently, we must pay close attention to crude oil prices to predict future prices.

The changes in oil prices can be recorded from time to time and aggregated to form a time series data. The time series data for crude oil prices are non-stationary because their average value, variance, and covariance change over time [5].

The comparison of the performance of the LSTM and SVR models in the case of stock prices of several companies, such as the NYSE, NSE, BSE, NASDAQ, and others, we have learned that the LSTM model provides the best forecasting results, with a MAPE value of 0.86. In comparison, the SVR model gives a MAPE error value of 1.44 [5].

In a similar case, the other research tries to forecast the closing price of the iShares MSCI United Kingdom fund, which is a non-linear and non-stationary dataset. The results of this study reveal that the LSTM model is the best-performing model. The MAE value given from this research is 0.210350. Then the SVR model followed, which has an MAE value of 0.24002, the random forest method, and the ANN method [6].

The Long Short-Term Memory (LSTM) model is an expansion model of the deep learning model, namely the Recurrent Neural Network (RNN) [7]. Hochrieter and Schmidhuber introduced this model in 1997. The RNN model has a long-term dependency, so the LSTM model builds to finish the problem that happens in the RNN model [7].

The Support Vector Regression (SVR) model is a model of the Support Vector Machine used to resolve regression cases [8]. The SVR model is used to solve regression cases by applying machine learning theory

to maximize the level of prediction accuracy while automatically avoiding over-fitting to data [8].

This research aims to compare the level of accuracy generated by the deep learning model, videlicet LSTM, and the machine learning model, SVR, in forecasting the price of crude oil, which is a non-linear time series data.

II. THEORY

A. *Long Short-Term Memory*

There is a problem that the Recurrent Neural Network (RNN) model needs to face, that is, a long-term dependency problem. So, to deal with the issues that ensued in the precursory model (RNN), the Long Short-Term Memory model was developed in 1997 [7]. In Figure 1, we can see the architecture of the LSTM model. The LSTM model has three types of gates: forget gates, input gates, and output gates.

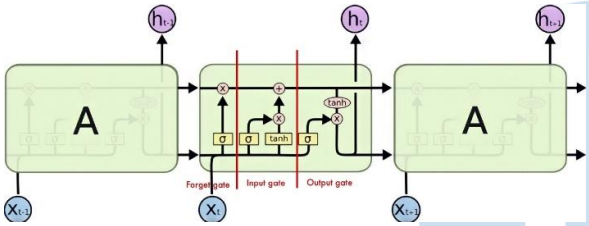


Fig. 1. The architecture of a cell in the LSTM unit

The charge of the forget gate component is to determine which data or information should be scraped or stored from the cell state [7]. From equation 1, we got the formulation of the forget gate.

$$f_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \quad (1)$$

Explanation :

- f_t = forget gate
- σ = sigmoid activation function
- W_f = forget gate weight
- h_{t-1} = previous layer's output
- x_t = new input vector at t time
- b_f = forget gate bias

We used the input gate to determine which new data would be stored in the cell state and determine the gate to write memory [7], [9]. To calculate the input gate, we can use equation (2), and to calculate the new candidate value to be added to the state, we use equation (3).

$$i_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \quad (2)$$

$$\tilde{c}_t = tanh\left(\sigma\left(W_c \times [h_{t-1}, x_t] + b_c\right)\right) \quad (3)$$

Explanation :

- i_t = input gate
- $\tilde{c}_t$ = new candidate values to be added to the state
- σ = sigmoid activation function
- tanh = tanh activation function
- W_f = forget gate weight
- W_c = cell state weight
- h_{t-1} = previous layer's output
- x_t = new input vector at t time
- b_f = forget gate bias
- b_c = cell state bias

Aside from various gates, an LSTM unit has a cell state. The task of the cell state is to remove information about the subject from the past and add the information formulated in equation (4) [7]. The operations performed on the cell state can be seen in Figure 2.

$$c_t = f_t \times c_{t-1} + \tilde{c}_t \times i_t \quad (4)$$

Explanation :

- c_t = cell state
- f_t = forget gate
- c_{t-1} = cell state value at time t - 1
- $\tilde{c}_t$ = new candidate value
- i_t = input gate

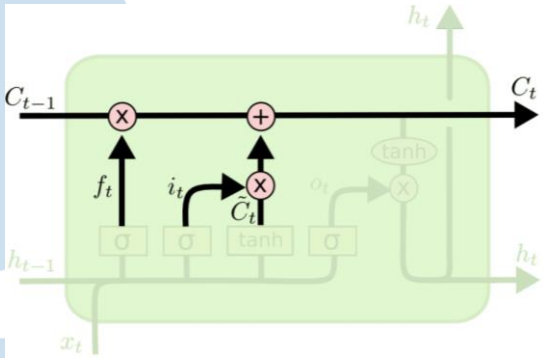


Fig. 2. LSTM cell state illustration [7]

The output gate is a gate that will decide what information should be issued and read from memory. We can use equation (5) to calculate the output gate, and we can use equation (6) to calculate the hidden state value [7].

$$o_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \quad (5)$$

$$h_t = o_t \times tanh\left(c_t\right) \quad (6)$$

Explanation :

- h_t = hidden state
- o_t = output gate
- σ = sigmoid activation function
- W_f = forget gate weight
- h_{t-1} = previous layer's output
- x_t = new input vector at t time
- b_f = forget gate vias
- c_t = cell state

B. Support Vector Regression

Support Vector Regression (SVR) is one of the machine learning models, scilicet Support Vector Machine (SVM), which is applied to solve regression problems. SVM is a model used to examine linear predictors in a high-dimensional feature space where the feature space's high dimensions can increase the sample's complexity and challenge the computational complexity [10].

This model was raised by Boser, Guyon, and Vapnik in 1992 [8]. SVR can be used to solve the case of the prediction of a value or price in the future [10]. The objective of the SVR model is to find a value that is close to the best within a certain margin (ε – tube). This model adjusts the error to a certain threshold. Moreover, this model also aims to insert as many data points as possible into the margins without breaking or crossing the margins. An illustration of how the SVR algorithm works and its components can be seen in Figure 3.

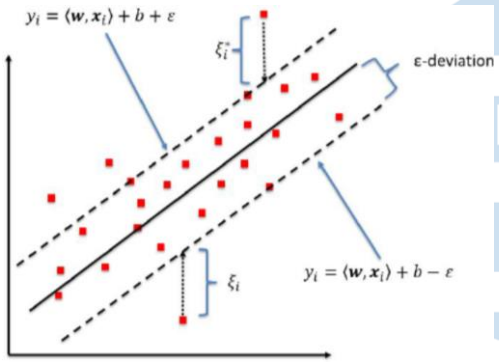


Fig. 3. Support Vector Regression component [11]

The basic concept of the SVR model is to map the x data into a feature space with high dimensions through non-linear mapping and execute linear regression in this space [12]. So, this model has the basic formula written in equation (7).

$$f(x) = \langle \omega, x \rangle + b \tag{7}$$

Explanation :

- $f(x)$ = the argument of the function
- ω = weight
- x = x value
- b = bias

Equation (7) shows that the training data were taken as $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \aleph \times \Re$, where $\aleph$ indicating the space of the input pattern. In equation (7), there is an operator $\langle ., . \rangle$. This operator indicates that the product dot operation ensues.

To minimize equation (7), we require the Euclidean norm using $\|\omega\|^2$ so that equation (7) turns into equation (8).

$$\text{Min} \quad \frac{1}{2}\|\omega\|^2 \tag{8}$$

$$\text{Subject to} \quad \begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \varepsilon \\ \langle \omega, x_i \rangle + b - y_i \leq \varepsilon \end{cases}$$

Explanation :

- y_i = dependent variable
- x_i = independent variable
- ω = weight
- b = bias
- ε = deviation or intensive loss function

We reconstruct equation (8) by introducing the slack variable (ξ_i, ξ_i^*) to solve the unfeasible optimization issue [12]. So, equation (9) is the result of the rebuild equation.

$$\text{Min} \quad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{9}$$

$$\text{Subject to} \quad \begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle \omega, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Explanation :

- C = Constant
- ξ_i = Slack variable

Linear and non-linear cases are the cases that the SVR model can handle. There are different formulas, whereas in the non-linear case, it is not explicitly given [12]. We can see the formula used for the non-linear case in equation (10) and Figure 4 to visualize the non-linear SVR case study.

$$f(x) = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)k(x_i, x) + b \tag{10}$$

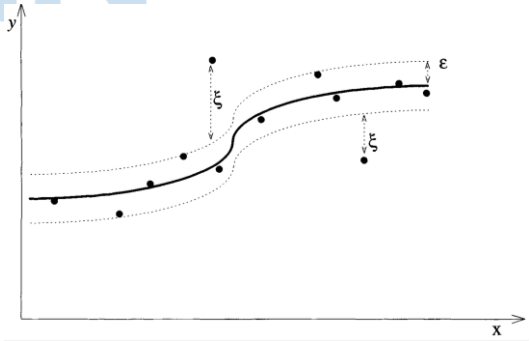


Fig. 4. SVR non-linear illustration [13]

From equation 8, we can see a kernel function $k(x_i, x)$, while α_i^* and α_i is a Lagrange multiplier and b is bias. The purpose of a kernel or kernel function is to overcome non-linear problems by involving linear classification [14].

The most common type of kernel used for the SVM case is the Radial Basis Function (RBF) kernel. RBF

evolves as the most common type of kernel used because the RBF kernel is dependent on the Gaussian distribution [15]. The RBF kernel intends to acquire an estimate of similarity or determine how close two points are, i.e., points x_1 and x_2 [15]. Equation (11) formulates the RBF kernel.

$$K(x_1, x_2) = \exp(-\gamma \|x_1 - x_2\|^2) \quad (11)$$

III. Research Method

This research consists of several stages that will be carried out, including data collection, pre-processing, parameter selection, training, testing, and predicting the future for 180 days. In this research, Nasdaq provides data for Brent crude oil so that we can collect the data from the website [16]. This dataset consists of 8887 rows of data and two columns.

The next step is pre-processing the data by dividing the dataset and normalizing the dataset.

A. Pre-Processing

The dataset utilized is Brent crude oil price data for May 1987 to May 2022. Overall, the dataset consists of 8887 rows and two columns, scilicet the Date column, and the Value column.

In this research, we saved 180 rows of data to predict the future after finishing the validating process. The testing data that will be forecasted is oil price data for the period of September 2021 to May 2022. This research will use the remaining 8707 rows of data for training and testing. This research will use 80 percent of the data for training, and we will use the rest of the dataset for validation.

TABLE I. Dataset Splitting

	Training Data	Validating Data
Amount and Percentage	6965 80%	1742 20%
Period	1987-05-20 to 2014-10-31	2014-11-03 to 2021-09-06

After splitting the data, we will normalize the data. The method we use to normalize the data is min-max scaling. With the normalization of min-max scaling, the actual price data will change in intervals between 0 and 1.

B. Parameter Selection

After going through the pre-processing process, the next step is specifying the parameters used in the experiments on the two models. The parameters used in each model will be explained in sections 1) to 2).

1) Long Short-Term Memory Model

In this research, we will combine several parameters, i.e., the number of LSTM units, batch sizes, and epochs. Table 2 contains various parameters to be combined and tested.

TABLE II. Estimation of LSTM Model Parameters

Parameter	Amount
Batch size	[32, 64, 128]
Layer	3 (LSTM, dropout, output)
Unit LSTM	[10, 30, 50]
Epoch	[5, 10, 15, 20, 25, 30]

2) Support Vector Regression Model

For the SVR model experiment, we will combine the parameters that we can see in table 3. The experiment uses a combination of these parameters to obtain the best results that the model can give in forecasting the price of Brent crude oil. The kernel used in this experiment is the Radial Basis Function (RBF) kernel.

TABLE III. Estimation of SVR Model Parameters

Parameter	Amount
C	[0.001, 0.01, 0.1, 1, 10, 100]
γ	[0.001, 0.01, 0.1, 1, 10, 100]

C. Training Model

This research used 80 percent of the dataset in the training process from May 20, 1987, to October 31, 2014. For the experiment, we used the various combinations of parameters listed in Table 2 and Table 3.

1) Long Short-Term Memory Model

The LSTM model uses three layers: an input layer, a dropout layer, and an output layer. In the input layer, we determine the number of LSTM units, then in the next layer, there is a dropout layer with a value of 0.05, and the output layer is dense, consisting of 1 unit neuron.

In the LSTM model, the activation function is tanh, and the optimizer is Adam optimizer. The epochs and batch size used in each experiment were also specified. To train the model, we need to execute the program by calling the fit function, which contains the parameters used.

To visualize the model's performance in each experiment, we create the MSE loss function graph. Moreover, using a loss function graph, we can use the RMSE value to determine how well the model performs. Also, we assemble a plot to make it easier to compare the actual and forecasted prices by the model.

2) Support Vector Regression Model

For this model, we trained the model using the various combination of C and γ. To perform the training of this model, we need to call the fit function after specifying the kernel type, C value, and γ value.

We will use the RMSE loss function to evaluate how accurate the model is. We will also visualize the plot by displaying the actual and forecasted prices.

D. *Validating Model*

At the validating step, we need to call the predict function and use the validation set that contains 20 percent of the total data using a model that has passed the training process. The period used for data validating is November 3, 2014, to September 6, 2021. After passing the forecasting process, we will calculate the RMSE value to determine the model's performance.

E. *Testing Model*

After passing the validation stage, the parameter that can provide the lowest error value from both models will be used at the testing stage. We use the testing data, which contains 180 days for the period from September 2021 to May 2022.

IV. RESULTS

In this research, we compare two models: scilicet LSTM, a deep learning model, and SVR, a machine learning model. From the Brent crude oil price data acquired from the Nasdaq website, we learned that the lowest price of crude oil was on December 10, 1998, at 9.1 USD per barrel. Meanwhile, the highest price was on July 3, 2008, at 143.950 USD per barrel. Figure 5 visualizes the crude oil price from May 20, 1987, to May 23, 2022.

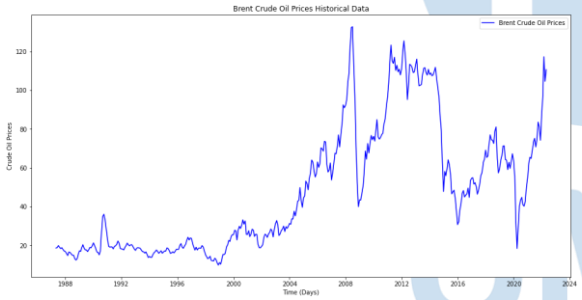


Fig. 5. Brent crude oil price chart for May 20, 1987, to May 23, 2022

A. *Long Short-Term Memory Model*

Using the combination of the parameters in Table 2, we conducted this research. The combination of these parameters gave 54 trials.

TABLE IV. LSTM TRAINING AND VALIDATION BEST RESULT

No	Unit LSTM	Epoch	Batch Size	RMSE	
				Train	Validation
1	10	20	64	1.968	1.543
2	30	20	64	2.046	1.588
3	50	20	64	1.948	1.742

Based on the results of the trials, we found that the numbers of LSTM units, the number of epochs, and batch sizes that give the best results are 20 and 64.

Table 4 shows that the LSTM model parameters that can give the most optimal results are the models with 10 LSTM units, 20 epochs, and a batch size of 64.

The root mean squared error value in the training process is 1.968. This RMSE value means that the model performs well at the training stage because the prediction values are close to the actual price.

To evaluate the model, aside from using the RMSE value, we can visualize the predicted value and the actual value from this model, which we can see in Figure 6. The blue line shows the actual price, while the red shows the forecast result. From Figure 6, it can be seen that the two lines have a small distance.

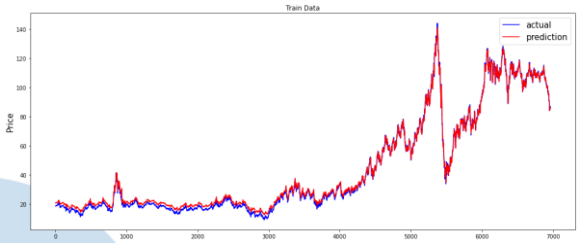


Fig. 6. LSTM training chart comparison of actual prices and predicted prices

In the validation process, the RMSE value was 1.543. Compared with the RMSE value at the training stage and the validation stage, we've got the RMSE value at the validation stage to have a smaller value; this means that the model's performance is getting better and has a good generalization ability. Figure 7 visualizes the results of the model validation, which show that the actual prices and forecasted prices have a small difference.

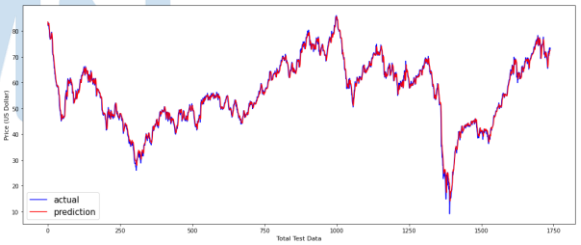


Fig. 7. LSTM validating chart comparison of actual prices and predicted results

From the validation outcomes above, the parameters that give the best results are used to forecast crude oil prices for the next 180 days. Figure 8 shows that the price for the next 180 days, visualized using a cyan-colored line, tends to fluctuate.

Figure 8 shows that the forecast results for the next 180 days have a relatively small difference from the actual price. The RMSE error value obtained is 3.628. Furthermore, we can see that the forecast price line almost covers the dark blue actual price line for the next 180 days.

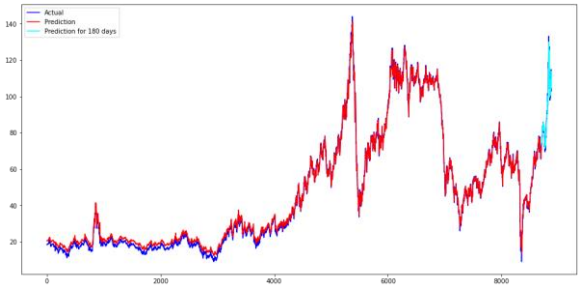


Fig. 8. Chart of LSTM model forecasting results for the next 180 days

B. *Support Vector Regression Model*

We trained the Support Vector Regression model using the parameters in Table 3. From the combination of various parameters, we did 36 training processes, so we have 36 SVR models.

TABLE V. SVR TRAINING AND VALIDATION BEST RESULT

No	C	Gamma	RMSE	
			Train	Validation
1	1	10	5.660	4.816
2	10	1	6.596	2.104
3	100	0.01	4.355	2.766

Table 5 displays that the model with the best performance is the model with a value of C 10 and Gamma 1. The training process indicates that the 3rd experiment's smallest root means squared error, where this model gives a 4.355 RMSE value. Furthermore, from Figure 9, we notice that the forecasting prices are close to the actual prices.

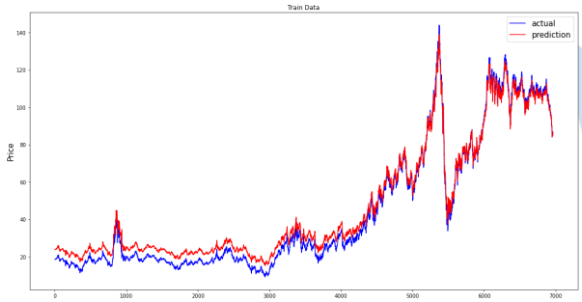


Fig. 9. SVR training chart comparison of actual prices and predicted prices

We got the second experiment from the training phase, which uses C 10 and Gamma 1 as the best model. So we used those parameters in the validation phase. Then, after we evaluated the model, we got the minimum error value RMSE is 2.104. Figure 10 visualizes the result of the SVR testing model.

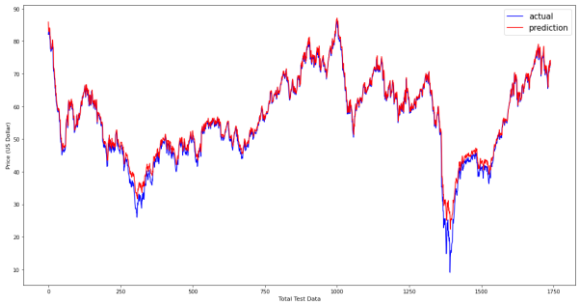


Fig. 10. SVR testing chart comparison of actual prices and predicted results

A model with a validation error value smaller than its training error value has a good generalization capability. So, depending on Table 5, we determine that the second experiment is the best model for forecasting the crude oil price.

After determining the best model, we use those parameters to predict the 180 days ahead. We can see the visualization of the future prediction in Figure 11, where the cyan-colored line represents the forecast value for the next 180 days.

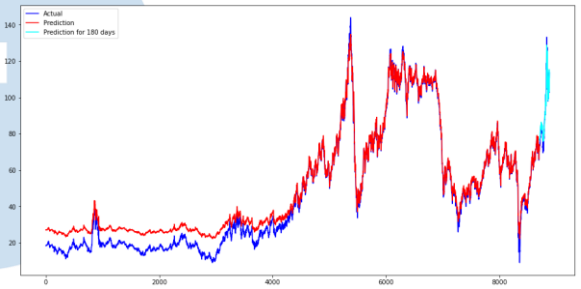


Fig. 11. Chart of SVR model forecasting results for the next 180 days

The RMSE value generated in the forecasting process for the next 180 days is 2.709. The visualization in Figure 11 shows that the price of Brent crude oil is volatile, and the actual price line and the forecasted price tend to have only a slight difference.

V. CONCLUSIONS

Brent crude oil price data recorded from time to time is time series data, and in this case, this dataset can be categorized as a non-linear dataset. In this research, we compare two models from deep learning and machine learning algorithms: Long Short-Term Memory and Support Vector Regression. We use the Keras library for building the Long Short-Term memory model and the Sklearn library for building the Support Vector Regression model.

The loss function Root Mean Squared Error (RMSE) is used to evaluate the model's performance. The time series data used is Brent crude oil price data collected from May 20, 1987, to February 28, 2022.

In this research, the LSTM model's evaluation results for the training process have an RMSE value of

1,968. While in the testing process, the RMSE value is 1.543. The parameters used are models with 10 LSTM units, 20 epochs, and 64 batch sizes.

In the Support Vector Regression model, the best RMSE value is 4.355 at the training stage. In comparison, the best RMSE value that can be generate at the validation stage is 2.104, where the parameters used in the SVR model that can give the best results are C 10 and γ 1.

The results of this study indicate that the Long Short-Term Memory (LSTM) model can predict the price of Brent crude oil, which is classified as data with non-linear characteristics better than the Support Vector Regression model. Moreover, from the experimental results, it is also known that the LSTM and SVR models have relatively small error value differences.

By doing this research, we hope that this research can be used as a reference for conducting research with other case studies or trying to combine the models that have been tested with other models.

REFERENCES

[1] G. A. Olah, Molnár Árpád, and P. G. K. Surya, "Hydrocarbon chemistry" 3rd ed. United States of America. 2018, ch. 1, pp 7-19.

[2] Knoema. Indonesia petroleum consumption, 2020-2022. Available: https://knoema.com/atlas/Indonesia/topics/Energy/Oil/Petroleum-consumption#:~:text=In%20January%202022%2C%20petroleum%20consumption,per%20day%20in%20January%202022.

[3] Knoema. Indonesia production of crude oil, 2020-2022. Available: https://knoema.com/atlas/Indonesia/topics/Energy/Oil/Production-of-crude-oil#:~:text=Indonesia%20production%20of%20crude%20oil%20was%20at%20level%20of%20616.06,is%20production%20of%20crude%20oil%3F

[4] Nizar, M. A., "Dampak fluktuasi harga minyak dunia terhadap perekonomian Indonesia". Buletin Ilmiah Litbang Perdagangan, 6(2), pp 189-210, 2012.

[5] G. Bathla, "Stock price prediction using LSTM and SVR", *Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2020, https://doi.org/10.1109/pdgc50313.2020.9315800.

[6] S. Carollo, *Understanding oil prices*. Cornwall, UK: John Wiley & Sons, 2012, pp. 89-101, https://doi.org/10.1002/9781118467251.

[7] C. Olah, "Understanding LSTM Networks", 27 August 2015, [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs.

[8] Jakkula, V.R., "Tutorial on support vector machine (SVM)", [Online]. Available: https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf.

[9] J. Firoujazee and P. Khaliliyan, "LSTM architecture for oil stocks prices prediction", *arXiv.org*, 2022. [Online]. Tersedia: https://arxiv.org/abs/2201.00350.

[10] M. Nikou, G. Mansourfar dan J. Bagherzadeh, "Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms", *Intelligent Systems in Accounting, Finance and Management*, vol. 26, no. 4, pp. 164-174, 2019. Available: https://doi.org/10.1002/isaf.1459.

[11] T. Kleynhans, M. Montanaro, A. Gerace, and C. Kanan, "Predicting top-of-atmosphere thermal radiance using MERRA-2 atmospheric data with Deep Learning," *Remote Sensing*, vol. 9, no. 11, p. 1133, 2017.

[12] Basak, D., Pal, S., & Patranabis, D.C., "Support vector regression", [Online]. Available: https://static.aminer.org/pdf/PDF/000/337/560/uncertainty_support_vector_method_for_ordinal_regression.pdf.

[13] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods". Cambridge University Press, Cambridge, 2000.

[14] A. Gupta. 1 June 2021. "Kernel tricks in support vector machine". [Online]. Available: https://medium.com/geekculture/kernel-methods-in-support-vector-machines-bb9409342c49.

[15] S. Sreenivasa. 12 October 2020. "Radial basis function (RBF) kernel: the go-to kernel". [Online]. https://towardsdatascience.com/radial-basis-function-rbf-kernel-the-go-to-kernel-acf0d22c798a.

[16] Nasdaq. 25 May 2022. "Crude Oil Prices: Brent - Europe". [Online]. Available: https://data.nasdaq.com/data/FRED/DCOILBRENTEU-crude-oil-prices-brent-europe.

ISSN 2085-4552

Air Temperature Sensor Estimation on Automatic Weather Station Using ARIMA and MLP

Haryas Subyantara Wicaksana[1], Naufal Ananda[2], Irvan Budiawan[3], Bayu Santoso[4], Roy Handoko[5], Asep Irwan Maulana[6], Suciarti[7], Ari Utoro[8]

[1,2,3,4,5] Faculty of Industrial Technology, Institut Teknologi Bandung, Bandung, Indonesia
[6,7,8] Center of Instrumentation, Calibration and Engineering, BMKG, Jakarta, Indonesia
[3] Faculty of Engineering, Jenderal Achmad Yani University, Cimahi, Indonesia
[1]haryas.wicaksana@bmkg.go.id, [2]naufal.ananda17@gmail.com, [3]budiawan.irvan@gmail.com, [4]bayusantos003@gmail.com, [5]rhandk@gmail.com, [6]aimmdo@gmail.com, [7]cafeinku@gmail.com, [8]utarari@gmail.com

Accepted 05 November 2022
Approved 26 November 2022

***Abstract*— Surface meteorological quantities are now measured by Automatic Weather Station (AWS). AWS Serang records weather parameters minutely in Banten Province of Indonesia. Air temperature sensor is one instrument of this system. This study aims to design an air temperature sensor estimator model using ARIMA and Artificial Neural Network (ANN) as solution for avoiding loss data. Air temperature sensor on AWS Serang data in August of 2022 period is segmented into training, validating and testing sections. Based on criterion calculation, ARIMA (1,1,5) is simulated. It obtains not more than 0.12 of RMSE, 0.052^{0}C of MAE, 0.193% of MAPE and 0.194% of SMAPE. Meanwhile, three different models of MLP ANN for air temperature estimator is also simulated. Input variables include air temperature, relative humidity and solar radiation intensity. Roy model has highest accuracy level for MLP ANN algorithm with 0.048 of RMSE, 0.026^{0}C for MAE, 5% of MAPE and 4.83% of SMAPE. Overall, ARIMA (1,1,5) is better than Roy MLP ANN model in estimating air temperature sensor data on AWS Serang. Nonetheless, both models are properly fulfilling WMO (World Meteorological Organization) accuracy requirements for air temperature measurement.**

***Index Terms*— air temperature sensor; ARIMA; Multi Layer Perceptron ANN**.

I. INTRODUCTION

Since 2013, Meteorology Climatology and Geophysics Agency of Indonesia has installed 367 Automatic Weather Stations (AWS). Automatic Weather Station is a digital meteorological measurement system. The measured parameters are consist of air temperature, relative humidity, air pressure, total daily rainfall, wind speed, wind direction and solar radiation intensity [1].

Air temperature has crucial roles on determining weather analysis and prediction. Air temperature is measured by outdoor thermometer inside a shield in the open air environment. Such shield covers the sensor material from solar direct radiation [2]. This parameter is measured by air temperature sensor of AWS minutely, hourly and daily. The sensor is installed inside an enclosure within relative humidity sensor 1,2 meters of height above ground.

Air temperature sensor of AWS is calibrated annually by field verification procedure. It is compared to reference portable AWS [3]. Nonetheless, the sensor has potential failure during operational time due to technical or non-technical factor. Sensor reparation or replacement needs certain time intervals, then it may causing loss data.

Air temperature sensor output can be estimated in order to minimize loss data. Smith et.al. (2007), have designed hourly temperature prediction system in Georgia using Artificial Neural Network (ANN) with air temperature, relative humidity, total rainfall and solar radiation intensity as model inputs [4]. Salcedo-Sanz et.al. (2016) forecast monthly temperature in New Zealand using Multi Layer Perceptron (MLP) with air temperature, Southern Oscillation Index (SOI), Indian Ocean Dipole (IOD) and Pacific Decadal Oscillation (PDO) as model inputs [5]. Then, Rahayu et.al. (2020) designed daily temperature estimation using Recurrent Neural Network (RNN) and Long-Short Term Memory (LSTM) with air temperature, relative humidity, total rainfall and wind speed as model inputs [6].

Tran et.al. (2021) has reviewed air temperature prediction models based on various machine learning algorithms. Mostly, ANN-based model such as MLP often provide more accurate air temperature estimation results. However, air temperature estimation based on ANN has not been sufficiently compared to another soft computing approach for time series data. It is recommended that air temperature estimation using ANN model should be compared to Auto Regressive

Integral Moving Average (ARIMA) model for future works [7].

This study intends to propose an air temperature sensor estimation model using ARIMA and ANN. At final steps, both model results will be analyzed and compared to each other. Besides, this study also utilizes direct measurement from ground weather station for improving input qualities of previous described researches.

II. DATA

This research uses AWS Serang data in August 2022 period. The parameter includes minutely air temperature, relative humidity, air pressure, wind speed and solar radiation intensity. Table I shows AWS Serang data.

TABLE I. RAW DATA OF AWS SERANG IN AUGUST 2022

Date	Time (UTC)	T (^{0}C)	RH %	P (mb)	WS m/s	SR (W/m^2)
1/8	2:50	30.2	57	1005.1	2.1	571.3
1/8	2:51	30.2	57	1005.1	1.5	643.3
1/8	2:52	30.3	56	1005.0	2	746.1
…	…	…	…	…	…	…
17/8	8:53	23.5	91	1005.0	1.5	10.9
17/8	8:55	23.5	91	1005.0	1.2	7.6
17/8	8:56	23.5	91	1005.0	1.4	28.9
…	…	…	…	…	…	…
31/8	3:52	30.4	56	1005.1	2.3	513.5
31/8	3:53	30.5	52	1005.1	1.6	921.8
31/8	3:54	30.6	52	1005.1	1.7	1028.6

AWS Serang is located on Serang Meteorological Station at 6,1111^0 S and 106,1218^0 E with 25 meter in elevation. The air temperature sensor is a Vaisala active mode HMP155A. Sensor data is recorded by using Campbell Scientific CR3000 logger.

III. METHODS

Air temperature sensor output estimation on AWS is designed based on ARIMA and MLP ANN models. ARIMA model only utilizes univariate temperature variable as input. While MLP ANN model utilizes multivariate variables as inputs. The multivariate variables are air temperature, relative humidity, air pressure, wind speed and solar radiation intensity.

Each model has its pros and cons. ARIMA is very effective and efficient for linear and stationary time series data. It also widely used for hydro-climatology parameters [8]. However, one parameter is often affected by other parameters in meteorology. ARIMA input is only consist of univariate variable. MLP is able to overcome such shortcomings of ARIMA weakness by using multivariate input variables. MLP also eases analysis for arbitrary nonlinear temperature data [7].

A. ARIMA Model

ARIMA model is combination of three statistical models: auto regression, integral and moving average. This model aims to analyze and to predict univariate variable in time series domain [8]. Mathematically, ARIMA model is stated as follows [9]:

$$W_t = \mu + \frac{\theta(B)}{\phi(B)} \quad (1)$$

$$(1-\mathrm{B})^{\mathrm{d}}\mathrm{Y_t} = \mu + \frac{(1-\theta_1\mathrm{B}-\ldots-\theta_\mathrm{p}\mathrm{B}^\mathrm{p})}{(1-\phi_1\mathrm{B}-\ldots-\phi_\mathrm{q}\mathrm{B}^\mathrm{q})}\mathrm{a_t} \quad (2)$$

Y_t as output variable, B as backshift operator, μ as mean value, θas auto regression operator, ϕas moving average operator, a_t as random errors. Value of p,d,q are ARIMA number models where p as auto regression number, d as differencing number and q as moving average number.

ARIMA model is arranged in some steps. These steps are consist of data identification, data preparation, data segmentation, model selection, and model evaluation [8]

1) Data Identification

First step of establishing ARIMA model is data stationary checking. Air temperature sensor data is plotted in time series domain. It is then plotted in auto correlation function graphic. Next, data stationary checking is held by Augmented Dicky-Fuller (ADF) Test, Phillip-Peron (PP) Test, Kowski-Phillips-Schmidt-Shin (KPSS) Test and Zivot-Andrews (ZA) Test [10].

2) Data Preparation

If air temperature sensor data is non-stationary, then it will be differenced. Differencing equation is stated as follow [10]:

$$X'_t = X_t - X_{t-1} \quad (3)$$

X_t is new input as result of subtraction of recent data to its previous data at t-1. Later, the differenced data will be identified again by data stationary checking methods. Data will be differenced more if it is still non-stationary.

3) Data Segmentation

After being prepared, air temperature sensor data is segmented into three parts: training data, validation data and testing data [11].

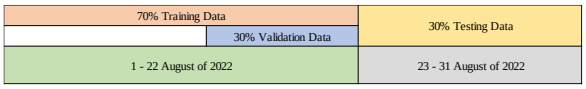


Fig. 1. First scenario of segmentation

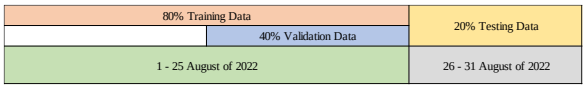


Fig. 2. Second scenario of segmentation

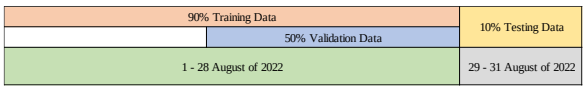


Fig. 3. Third scenario of segmentation

Figure 1 shows first scenario that divides data into 18.012 training data, 5.404 validation data which are overlapped to training data (13-22 August of 2022), and 7.719 testing data. Figure 2 shows second scenario that divides data into 20.585 training data, 8.234 validation data which are overlapped to training data (15-22 August of 2022), and 5.146 testing data. Figure 3 shows third scenario that divides data into 23.158 training data, 11.579 validation data which are overlapped to training data (14-28 August of 2022), and 2.573 testing data.

4) Model Selection

ARIMA model selection can be preceded by plotting auto correlation function (ACF) and partial auto correlation function (PACF) graph [12]. ACF is correlation between variable value in certain time with its value in whole previous times. PACF is partial correlation between variable value in certain time with its value in some of previous times. Equation (4) and (5) show ACF and PACF respectively [13].

$$\mathrm{ACF} = \mathrm{corr}(X_t, X_{t-k}) = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \quad (4)$$

$$\mathrm{PACF} = \mathrm{corr}(X_t, X_{t-k} \mid X_{t-1}, X_{t-2}, \ldots, X_{t-k+1}) \quad (5)$$

ACF plot is utilized in predetermining moving average number, while PACF plot is utilized in predetermining auto regression number [14]. Then, final selection will be assessed by using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [15][16]. Equation (6) and (7) state AIC and BIC respectively.

$$\mathrm{AIC} = -2\ln(L) + 2k \quad (6)$$

$$\mathrm{BIC} = -2\ln(L) + k\ln(n) \quad (7)$$

L is maximum value of likelihood function, k is amount of model parameters, and n is amount of model data. Less value of AIC and BIC give better ARIMA model selection [17].

5) Model Evaluation

ARIMA model estimation accuracy is evaluated by obtaining coefficient determination (R-squared), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE) [18]. Formulation of R-squared, RMSE, MAE, MAPE and SMAPE is respectively shown by equation (8), (9), (10), (11) and (12).

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y} - Y_i)^2} \quad (8)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \quad (9)$$

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i| \quad (10)$$

B. MLP ANN Model

ANN is one branch of artificial intelligence. ANN adopts human neural nervous system. Multi Layer Perceptron ANN (MLP ANN) is a type of ANN which is prominently applied for temperature forecasting [7]. It is formed by input, hidden and output layer. It processes input data through neural network operation with certain adaptive weighting. MLP ANN output is combination of activation function, input weighting, and bias. It is formulated as follow [19]:

$$Y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \quad (13)$$

Y is neuron output, x is input, w is weighting and b is bias. Equation (13) is called activation function. This research applies sigmoid, hyperbolic tangential and rectified linear unit function [20]. These functions are stated by equation (14), (15) and (16).

$$F_s = f(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

$$F_t = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (15)$$

$$F_r = f(x) = max(0, x) \quad (16)$$

Fs stands for sigmoid function, Ft stands for hyperbolic tangential function and Fr stands for rectified linear unit function [21]. MLP ANN model arrangement is consist of three main steps : data normalization, data selection, model training, validation and testing [22].

1) Data Normalization

Due to multivariate variables and quantities, input and output data should be normalized. Normalization aims to simplify correlation calculation [23]. All variables are transformed to be new contemporary variables based on their origin coordinates in range of 0-1. Data normalization is stated as follow [24]:

$$Z_{norm} = \frac{Z_i - \bar{Z}}{stdev(Z)} \quad (17)$$

Each variable is subtracted to its average and divided by its standard deviation. Data normalization result is then processed.

2) Data Selection

Input variables are adopted from recent air temperature forecasting ANN model [25]. It includes three previous delayed data for each variable [26]. Table II shows raw input variables.

TABLE II. RAW INPUT VARIABLES

Input	Unit	Delay
Air Temperature	^{0}C	(t-1),(t-2),(t-3)
Relative Humidity	% RH	(t-1),(t-2),(t-3)
Air Pressure	mbar	(t-1),(t-2),(t-3)
Wind Speed	m/s	(t-1),(t-2),(t-3)
Solar Radiation Intensity	W/m^2	(t-1),(t-2),(t-3)

Next, those variables are filtered by using Principal Component Analysis (PCA) algorithm. PCA is one of variability analysis method. It has purpose to reduce input dimension. PCA also determines the most significant input variable [27]. PCA runs based on matrix covariance calculation. After decomposing the eigen vectors, the biggest eigen values will be obtained from low dimensional subspace [28]. Eigen values correspond to variance of each input variable. It describes the input correlation strength to the output. Eigen vectors with biggest eigen value and largest variance will be detained as final input variables [29].

3) Support Vector Regression Model

Final input variables will be segmented in same scenarios as ARIMA model. It is also evaluated by using same accuracy parameters as ARIMA model, so that it will be fairly comparable.

IV. RESULT AND ANALYSIS

ARIMA and MLP model are tested based on comparison methods. The known actual tested value are compared simultaneously to the prediction results. Proportion of testing data has been explained in Figure 1, Figure 2 and Figure 3. Results of both models are checked by using five accuracy criteria: R-squared, RMSE, MAE, MAPE and SMAPE. R-squared shows correlation between actual and predicted value. RMSE and MAE evaluate the model based on same unit as air temperature in celsius degree. While MAPE and SMAPE assess the model in form of the error percentage.

A. ARIMA Model Result

Air temperature sensor data is plotted per minute. Figure 4 is an ACF graphic of air temperature original data in August of 2022 period. This figure shows a series of non-stationary data in variance, because its correlation value degradation moves down slowly from one lag to next lag.

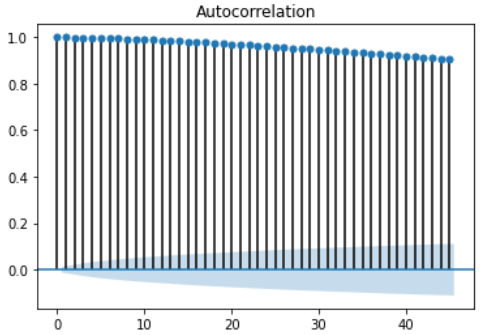


Fig. 4. ACF graphic of original air temperature sensor data

Next, air temperature sensor data is differenced once. Stationary characteristic of original data and differencing data are then tested. Table III shows stationary test result of both data.

TABLE III. STATIONARY DATA CHECKING

Test	Original Data		Differenced Data	
	p-value	characteristic	p-value	characteristic
ADF	0,000	weakly stationary	0,000	stationary
PP	0,000	weakly stationary	0,000	stationary
KPSS	0,117	weakly stationary	0,995	stationary
ZA	0,000	weakly stationary	0,000	stationary

The first order differenced data has became stationary data. Figure 5 shows ACF and PACF graphic of differenced data. ACF graphic describes significant correlation value degradation.

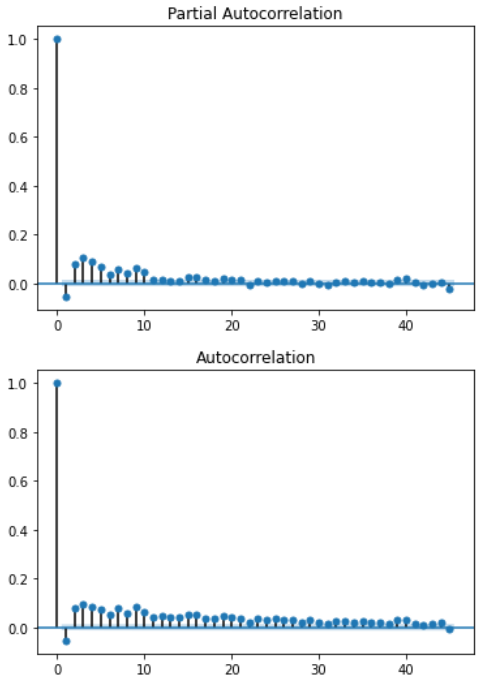


Fig. 5. PACF and ACF graphic of first differenced data

ACF graphic shows that fifth lag correlation is still significant, so moving average order is in range of 1-5. While PACF graphic shows that third lag partial correlation is significant, so auto regression order is in range of 1-3. These order estimations are then confirmed by calculating AIC and BIC values. Table IV describes AIC and BIC values for each ARIMA model simulation with d=1.

TABLE IV. AIC AND BIC VALUE FOR ARIMA MODEL

ARIMA Model	AIC	BIC
ARIMA (1,1,1)	-43,668.452	-43,643.985
ARIMA (1,1,2)	-44,173.694	-44,141.072
ARIMA (1,1,3)	-44,172.481	-44,131.704
ARIMA (1,1,4)	-44,177.763	-44,128.831
ARIMA (1,1,5)	-44,177.929	-44,120.841
ARIMA (2,1,1)	-44,155.510	-44,122.888
ARIMA (2,1,2)	-44,172.415	-44,131.638
ARIMA (2,1,3)	-44,172.704	-44,123.772
ARIMA (2,1,4)	-44,174.778	-44,117.691
ARIMA (2,1,5)	-44,174.467	-44,109.223
ARIMA (3,1,1)	-44,176.266	-44,135.489
ARIMA (3,1,2)	-44,175.169	-44,126.236
ARIMA (3,1,3)	-44,173.561	-44,116.473
ARIMA (3,1,4)	-44,171.465	-44,106.221
ARIMA (3,1,5)	-44,170.620	-44,097.221

Based on Table IV, ARIMA (1,1,5) has the smallest AIC value, and relatively smaller BIC value than the others. ARIMA (1,1,5) is then chosen as air temperature sensor data estimator model. ARIMA (1,1,5) equation for air temperature sensor estimator is stated as follow.

$$Y_t = 1{,}141 + 1{,}9575Y_{t-1} + 1{,}0470e_{t-1} - 0{,}1413e_{t-2} - 0{,}0194e_{t-3} + 0{,}0112e_{t-4} + 0{,}008e_{t-5} \quad (18)$$

ARIMA (1,1,5) model is trained, validated and tested based on three segmentation scenarios. Table V shows evaluation results of the model.

TABLE V. ARIMA (1,1,5) ACCURACY EVALUATION

Scenario	R^2	RMSE	MAE (^{0}C)	MAPE (%)	SMAPE (%)
1	0,998	0,119	0,052	0,193	0,194
2	0,998	0,105	0,048	0,181	0,181
3	0,998	0,114	0,047	0,172	0,172

Determination coefficient of ARIMA (1,1,5) is close to 1, so the estimator model is strongly correlated to the actual value. RMSE, MAE, MAPE and SMAPE values are very small, so the model is precisely accurate. An increase on training data percentage can increase the model's accuracy level.

B. MLP Model Result

MLP ANN input and output data are normalized. Figure 6 is variance ratio percentage graphic as PCA result. This figure states that only seven input variables are significant to be injected into the model.

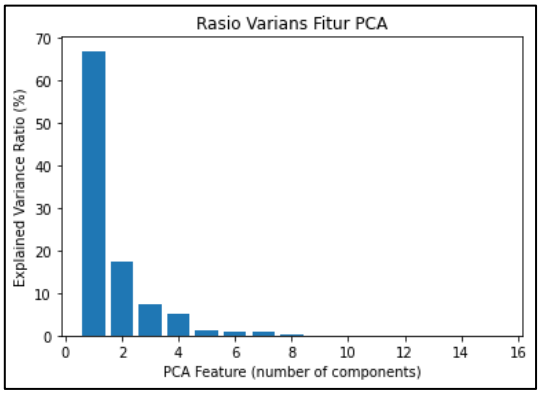


Fig. 6. Variance ratio percentage of PCA result

Table VI shows seven significant inputs for the MLP ANN model. These inputs has bigger eigen values than the rests. Such inputs are then detained. Significant inputs are dominated by the lagged air temperature data. Relative humidity is also strongly related to air temperature measurement dynamics, so it is significant too. However, lagged solar radiation intensity mainly influences all weather parameter measurement.

TABLE VI. SIGNIFICANT INPUT PARAMETERS

Significant Inputs	Eigen Value	Variance Ratio (%)
Air temperature (t-2)	0,30319	66,91
Air temperature (t-1)	0,30317	17,39
Air temperature (t-3)	0,30304	7,31
Relative humidity (t-1)	0,29850	4,95
Relative humidity (t-2)	0,29841	1,32
Relative humidity (t-3)	0,29815	0,90
Solar radiation intensity (t-3)	0,25150	0,84

MLP ANN model is adopted from previous existing air temperature forecasting model. Singh et.al. (2019), Roy (2020) and Lee et.al. (2020) have designed MLP ANN model with same inputs in Table VI partially [22], [31] and [32]. These three models are simulated for detained significant inputs. Table VII describes those models in detail.

TABLE VII. MLP ANN MODEL DETAIL

Model Version	Layer	Neuron	Activation Function
Singh et.al. (2019)	5	Layer 1 : 16 Layer 2 : 32 Layer 3 : 16 Layer 4 : 5 Layer 5 : 1	Layer 1 : relu Layer 2 : relu Layer 3 : relu Layer 4 : relu Layer 5 : linear
Roy (2020)	3	Layer 1 : 16 Layer 2 : 16 Layer 3 : 1	Layer 1 : relu Layer 2 : relu Layer 3 : linear
Lee et.al. (2020)	6	Layer 1 : 16 Layer 2 : 16	Layer 1 : tanh Layer 2 : tanh

		Layer 3 : 16 Layer 4 : 16 Layer 5 : 16 Layer 6 : 1	Layer 3 : tanh Layer 4 : sigmoid Layer 5 : sigmoid Layer 6 : linear

MLP ANN model is adopted from previous existing air temperature forecasting model. The models are trained and validated by using Adam optimizer with 32 batches and 100 epochs [25]. Figure 7 shows loss level against epochs for Singh et.al. model. Loss level is mean squared error for each epoch. Training and validating loss decrease fast at even less than 20 epochs. It proves that the increasing epochs should minimize losses until certain level. Loss value will be stagnant after reaching an effective total epochs.

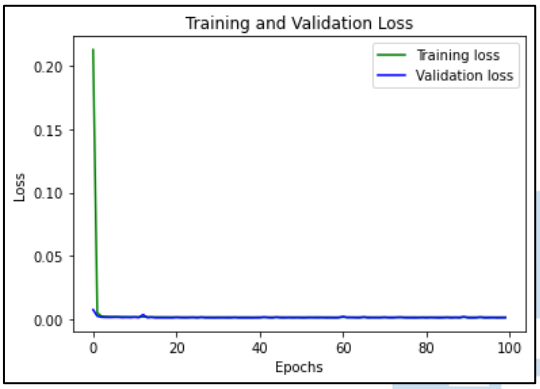


Fig. 7. Loss vs total epochs in Singh et.al. Model

Each model gives various timing process. Figure 8 shows training epochs duration for each model in three segmentation scenarios. Roy model has shortest timing process : 138.20 seconds for first scenario; 147.52 seconds for second scenario; and 161.69 seconds for third scenario. It is a fast computation process since it only has three layers of modelling.

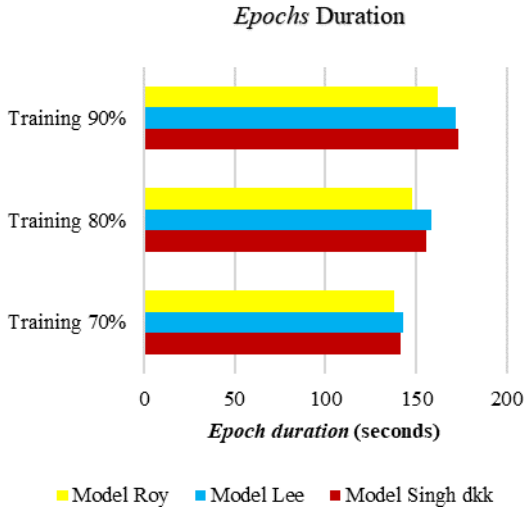


Fig. 8. Training epochs duration

Table VIII explains detailed training and validation loss for each model after 100 epochs. Singh et.al. and Roy model have same loss value due to similarity of activation function. Both are lesser than Lee et.al. model which is combined by sigmoid and tanh function.

TABLE VIII. MLP ANN MODEL DETAIL

Model Version	Scenario	Training Loss (MSE)	Validation Loss (MSE)
Singh et.al. (2019)	1	0,0015	0,0012
	2	0,0016	0,0012
	3	0,0013	0,0013
Roy (2020)	1	0,0018	0,0014
	2	0,0021	0,0014
	3	0,0018	0,0017
Lee et.al. (2020)	1	0,0015	0,0012
	2	0,0016	0,0012
	3	0,0013	0,0013

TABLE IX. MLP ANN MODEL DETAIL

Model	Scenario	R^2	RMSE	MAE (^{0}C)	MAPE (%)	SMAPE (%)
Singh et.al. (2019)	1	0,998	0,036	0,019	4,95	4,89
	2	0,998	0,038	0,022	5,30	5,09
	3	0,997	0,045	0,020	4,94	4,83
Lee et.al. (2020)	1	0,998	0,039	0,022	5,81	5,57
	2	0,998	0,041	0,023	5,32	5,04
	3	0,997	0,048	0,026	5,49	5,18
Roy (2020)	1	0,998	0,036	0,019	4,87	4,76
	2	0,998	0,038	0,020	5,00	4,83
	3	0,997	0,044	0,020	4,66	4,53

The models are then tested to estimate air temperature sensor data. Prediction result is evaluated against actual measurement value. Table IX shows model evaluation results.

All determination coefficients of MLP ANN model show strong correlation value. An increase of training data amount may decrease the correlation due to higher bias probability in small testing data amount.

Singh et.al. and Roy MLP ANN model have relatively same RMSE and MAE values for each segmentation scenario. RMSE, MAE, MAPE and SMAPE of Singh et.al. and Roy model are lower than Lee et.al. model. It proves that relu activation function is more suitable than sigmoid or tanh for air temperature sensor estimator model on AWS Serang. Roy model has smallest MAPE and SMAPE value at all segmentation scenarios.

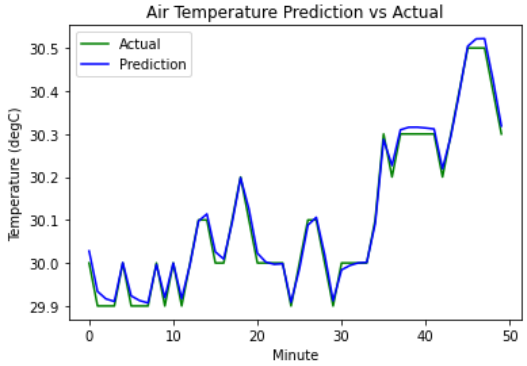
Therefore, Roy MLP ANN model has higher accuracy level than two other models. It indicates that an increase on layer or neuron amount of MLP ANN does not vouch the model accuracy. In addition, Roy model has fast computation time. It can be inferred that Roy model is quite effective and efficient for air temperature sensor data estimator on AWS Serang.

C. ARIMA vs MLP

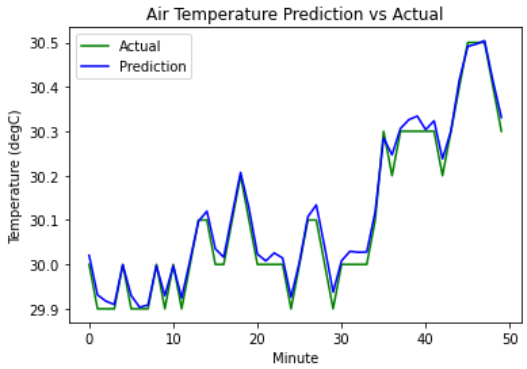
Based on Table V and Table IX, ARIMA (1,1,5) has smaller MAPE and SMAPE values than MLP ANN model significantly. Meanwhile, it has bigger RMSE and MAE values than MLP ANN model. Nevertheless, MAPE and SMAPE values are more sensitive on

explaining error parameters than RMSE and MSE, because both are proportionally compared against the actual values in percentage.

Figure 9 shows air temperature sensor output estimation plot of ARIMA (1,1,5) and Roy MLP ANN model in same certain periods. ARIMA (1,1,5) prediction plot is fitter than Roy MLP ANN model. Roy model has little ripples in stable air temperature condition.



(a)



(b)

Fig. 9. Comparison of (a) ARIMA(1,1,5) and (b) Roy MLP ANN prediction plot

However, MAE value of both models are still fulfilling WMO (World Meteorological Organization) accuracy requirements for air temperature measurement. Permitted maximum error in WMO No.8 document is below 0.2^0C. MAE value of both models are even less than 0.1^0C.

On the other aspect, ARIMA (1,1,5) needs lesser input variables than Roy MLP ANN model. Moreover, it has faster computation process and smaller memory capacities. By utilizing Python 3.6 programming, ARIMA (1,1,5) spends only 4 kilobytes, while MLP ANN model spends 7 kilobytes.

ARIMA (1,1,5) is potentially applied in AWS logger programming code or in server processing. New input can be maintained as new trained model routinely in order to enrich and improve air temperature estimation accuracy.

V. Conclusions

According to result analysis, it can be inferred that ARIMA and MLP ANN is able to estimate minutely air temperature sensor data on AWS Serang. ARIMA (1,1,5) has lowest AIC, so it is proper to estimate air temperature data with very small MAPE and SMAPE value. Roy model has better accuracy than Singh et.al. And Lee et.al. model as MLP ANN estimator for air temperature sensor data. Overall, ARIMA (1,1,5) is more accurate than Roy MLP ANN model. Besides, it also has simpler computation processing than MLP ANN model. Hybrid model of ARIMA and ANN is recommended for future works in order to improve air temperature estimator accuracy.

Acknowledgment

We would like to thank Mr. Ir. Endra Joelianto, Ph.D., SMIEEE and Mrs. Miranti Indar Mandasari, Ph.D for the guidance and constructive advices on regarding our studies.

References

[1] Meteorology Climatology and Geophysics Agency of Indonesia, Director of Indonesian Agency for Meteorology Climatology and Geophysics Regulation No. 7 2014 regarding Operational and Technical Standard for Observational Instruments Maintenance, Jakarta: BMKG, 2014.

[2] World Meteorological Organization, WMO No.8 Guide to Instruments and Methods of Observation, Geneva: WMO, 2018.

[3] Meteorology Climatology and Geophysics Agency of Indonesia, Director of Indonesian Agency for Meteorology Climatology and Geophysics Regulation No. 23 2015 regarding Procedure of Observational Instruments Calibration, Jakarta: BMKG, 2015.

[4] B.A. Smith, R.W. Mcclendon, G. Hoogenboom, “Improving air temperature prediction with artificial neural networks”, Int. J.Comput. Inf. Eng., 2007, 1, 3159.

[5] S. Salcedo-Sanz, R.C. Deo, L. Carro-Calvo, B. Saavedra-Moreno, “Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms”, Theor. Appl. Climatol, 2016, 125, pp. 13–25.

[6] I.S. Rahayu, E.C. Djamal, R. Ilyas, “Daily temperature prediction using recurrent neural networks and long-short term memory”, Proceedings of the 5th NA International Conference on Industrial Engineering and Operations Management, Detroit, Michigan, USA, 2020.

[7] T.T.K. Tran, S.M. Bateni, S.J. Ki, H. Vosoughifar, “A review of neural networks for air temperature forecasting”, Water, 2021, 13, 1294.

[8] E.G. Jain and B. Mallick, (2017) : “A study of time series models ARIMA and ETS”, I.J. Modern Education and Computer Science, 2017, 4, pp.57-63.

[9] SAS, SAS/ETS 13.2 User’s Guide : The ARIMA Procedure, NC : SAS Institute Inc., 2014.

[10] R.H. Shumway and D.S. Stoffer, Time Series Analysis and Its Application 3rd Edition, NY: Springer, 2011.

[11] S. Bandong, E. Joelianto, E. Leksono, A. Purwarianti, I.N. Haq, “One-step and multi-step performance ratio prediction of solar power plants using time series ARIMA”, Internetworking Indonesia Journal, vol.12, 2020, pp.39-45.

[12] B.A. Safitri, A. Iriany, N.W.S. Wardhani, “Accuracy comparison on rainfall forecasting using ARIMA, Hybrid ARIMA-NN, and FFNN in Malang Regency”, National Seminar Official Statistics, 2021, pp.245-253.

[13] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis, Forecasting and Control, NJ: Prentice-Hall, Inc., 1994.
[14] B.G. Prianda and E. Widodo, "Method comparison between seasonal ARIMA and extreme learning machine for foreigner visitor prediction in Bali", Jurnal Ilmu Matematika dan Terapan, vol.15, no.4, 2021, pp.639-650.
[15] H. Akaike, "Information theory and an extension of the maximum likelihood principle", in Petrov, B.N. and Csaki, F.Eds., International Symposium on Information Theory, 1973, pp.267-281.
[16] G. Schwartz, "Estimating the dimension of a model", Ann. Stat., vol. 6, no. 2, 1978, pp. 461–464.
[17] Y. Rahkmawati, I.M. Sumertajaya, M.N. Aidi, "Evaluation of accuracy in identification of ARIMA models based on model selection criteria for inflation forecasting with the TSClust approach", International Journal of Scientific and Research Publications, vol. 9, issue 9, 2019, pp.439-443.
[18] D. Chicco, M.J. Warrens, G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation", PeerJ Computer Science, 7, 2021, pp.1-24.
[19] M. Bilgili and B. Sahin, "Prediction of long-term monthly temperature and rainfall in Turkey", Energy Sources, Part A Recover. Util. Environ. Eff, 32, 2010, pp. 60–71.
[20] J. Feng and S. Lu, "Performance analysis of various activation functions in artificial neural networks", IOP Conf. Series: Journal of Physics, 1237, 2019.
[21] T. Szandala, Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks, SC: Springer Nature Singapore Pte Ltd., 2020.
[22] S. Singh, M. Kaushik, A. Gupta, A.K. Malviya, "Weather forecasting using machine learning techniques", Proceedings of 2nd International Conference on Advanced Computing and Software Engineering, 2019.
[23] S. Wang, and J. Cui, "Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method", Applied Energy 82, 2005, pp.197–213.
[24] F.L. Gewers, G.R. Ferreira, H.F. Arruda, F.N. Silva, C.H. Comin, D.R. Amancio, L.F. Costal, "Principal component analysis: a natural approach to data exploration", ACM Computing Surveys, 54, 2021, pp.1-34.
[25] M. Sundaram, M. Prakash, I. Surenther, N.V. Balaji, S. Kannimuthu, "Weather forecasting using machine learning techniques", Test Eng. Manag., 2020, pp.15264–15273.
[26] M.A. Jallal, S. Chabaa, A. El Yassini, A. Zeroual, S. Ibnyaich, "Air temperature forecasting using artificial neural networks with delayed exogenous input", Proceedings of the 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 2019.
[27] C.S. Kenfack, F.K. Mkankam, G. Alory, Y.Y. Penhoat, N.H. Mahouton, D.A. Vondou, B.G. Nfor, "Sea surface temperature patterns in the Tropical Atlantic: Principal component analysis and nonlinear principal component analysis", Terr. Atmos. Ocean. Sci., vol. 28, no. 3, 2017, pp.395-410.
[28] S. Liu, L. Feng, J. Wua, G. Houc, G. Hand, "Concept drift detection for data stream learning based on angle optimized global embedding and principal component analysis in sensor networks", Computers and Electrical Engineering, 58, 2017, pp.327-336.
[29] Y. Borgne, S. Raybaud, G. Bontempi, "Distributed principal component analysis for wireless sensor networks", Sensors, 8, 2008, pp.4821-4850.
[30] S. Lee, Y.S. Lee, Y. Son, "Forecasting daily temperatures with different time interval data using deep neural networks", Appl. Sci., 2020, 10, 1609.
[31] D.S. Roy, "Forecasting the air temperature at a weather station using deep neural networks", Procedia Comput. Sci., 178, 2020, pp.38–46.

ISSN 2085-4552

Analyzing Level of International Humanitarian Law Knowledge and its Compliance Through Military Simulation Game

William Hokianto [1], Melissa Indah Fianty[2]
[1,2] Department of Information Systems, Universitas Multimedia Nusantara, Tangerang, Indonesia
[1] william.hokianto@student.umn.ac.id, [2] melissa.indah@umn.ac.id

Accepted 22 November 2022
Approved 14 December 2022

***Abstract*— International humanitarian law in armed conflicts, the majority of whose contents originate from the results of the Geneva Convention, is the law used to protect human rights during an armed conflict. Based on research conducted by the International Red Cross Committee, video games that are now popular among the public, especially in the military genre, often ignore or contradict aspects of international law. This research was conducted to determine the knowledge of international law in armed conflict and its relationship to the level of compliance through military simulation video game scenarios. The scenario-based framework is a method used to develop game scenarios. The scenario was developed using an ARMA 3 video game editor and created to describe a situation of armed conflict where the respondent will take action whose results reflect the respondent's level of compliance with international law. Each respondent obtained two types of scores, the first is knowledge scores obtained from filling out questionnaires, and the second is compliance scores obtained from the results of running scenarios in the game. The score is divided into 12 parameters relating to aspects of international law in armed conflict. After the two types of scores are obtained, an analysis is carried out to determine whether the respondent's knowledge level influences the compliance level. The study's results showed that the respondent's knowledge and compliance levels were very high. The results of the comparative analysis show that only ten parameters have a level of knowledge that influences the level of compliance.**

***Index Terms*— International Law; Scenario; Video Game**.

I. INTRODUCTION

International Humanitarian Law has existed since the 19th century to protect non-combatants and minimize suffering and damage in a conflict. Three protocols in this law are written in the agreement at the Geneva convention, which has been continuously developed and updated until now. The first protocol is related to protecting victims of international armed conflicts. The second protocol is related to protecting victims of non-international armed conflicts. The third protocol is related to using additional special symbols [1]. This agreement requires that humanitarian law is part of the military and education program so that humanitarian law applies as widely as possible to the community.

Spreading knowledge of international humanitarian law is not enough to prevent violations of this law, so a complete integration of aspects and changes in the current civilization system is needed [2]. As many as 59% of teenagers who were survey respondents considered that torturing captives to obtain information was acceptable behaviour [3]. Knowledge and compliance with international humanitarian law are still low. However, the research does not mention video games, and the study provides insight into the attitudes of teenagers towards international humanitarian law for this research.

With hundreds of millions of active players worldwide, the video games industry has become a global phenomenon that transcends social, cultural, geographical, age and income brackets [4]. In contrast to passive media such as films, television and books, video games can significantly influence a person because they involve directly in the video game. One of the most popular video game genres is the military simulation genre. ARMA 3 is a realism-based, tactical military shooter video game developed and published by Bohemia Interactive. It was released for the Microsoft Windows PC platform in September 2013 and was later announced to be present on OS X and Linux in August 2015 [5].

With the increasing number of products and the number of military simulation video game players, this is certainly a problem because this can trigger and encourage players to carry out prohibited actions [6]. Military simulation video games offer players the to experience virtual warfare. However, many video game products of this genre are now lacking in providing an educational element of international humanitarian law and instead giving the wrong perspective on the actions its player must take to prioritize the entertainment element to increase sales.

Based on this problem, the research was conducted to determine how the level of players' knowledge of international humanitarian law affects the level of its compliance by using scenarios in military simulation video games as tools for this research [7].

II. METHOD

The research was carried out by developing scenarios first using the scenario-based framework methodology. The data obtained from the test results were analyzed using the Likert scale methodology.

A. Scenario Based Framework

A scenario can be understood as a means to change the game's initial state to its final state, following the main development trends, which are influenced by internal events and external activities as shown below:

- SI is the initial state of the game.
- T, predetermined development environment trend.
- E, predetermined internal event.
- A, external activities that have been carried out.
- SF, the game's final state after the scenario is terminated.

Scenario development often begins with an analysis of the present by determining the key factors influencing the considered system. Games and scenarios can build social structures or knowledge that, from a constructivist perspective, illustrate how future reality felt [8].

The scenario-based framework and development approach in Figure 2 and Figure 3 is a framework that can be used in the designing scenario of the simulation video game. Usually, the game manager or scenario designer chooses an activity before the game or scenario runs. And then, when the scenario starts, the player will decide based on the objective. If it reaches a specific result, the game manager can enter an event into the scenario to motivate the player to reach a goal. The following is a diagram of the framework of this method.

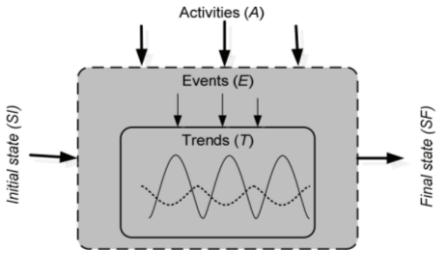


Fig. 1. Generic scenario structure

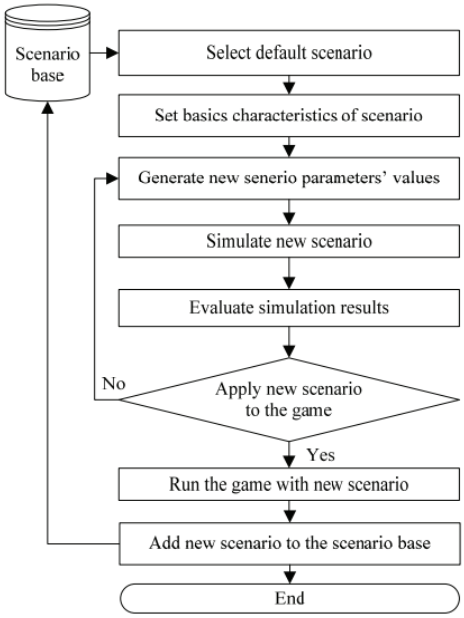


Fig. 2. Scenario development approach

The scenario-based framework method in video games is used to create games where players are placed in complex problem spaces that mimic real-world situations dan demonstrate how to design and develop a generic platform that enables rapid and flexible development and delivery of a wide variety of serious, scenario-based games that enable the acquisition of complex cognitive skills [9].

B. Scenario Development Methodology

The development of scenarios using the scenario-based framework method has nine steps consisting of [10]:

- System/object initial phase.
- Define goals.
- Define time steps.
- Estimate future states.
- Define rules for choosing future states.
- Choose the subsequent stage.
- Refine the initial state.
- Evaluate goal achievement degree.
- Check if the goal is achieved & terminate the procedure.

In this methodology, there are inputs and outputs for each stage, as shown in Figure 3.

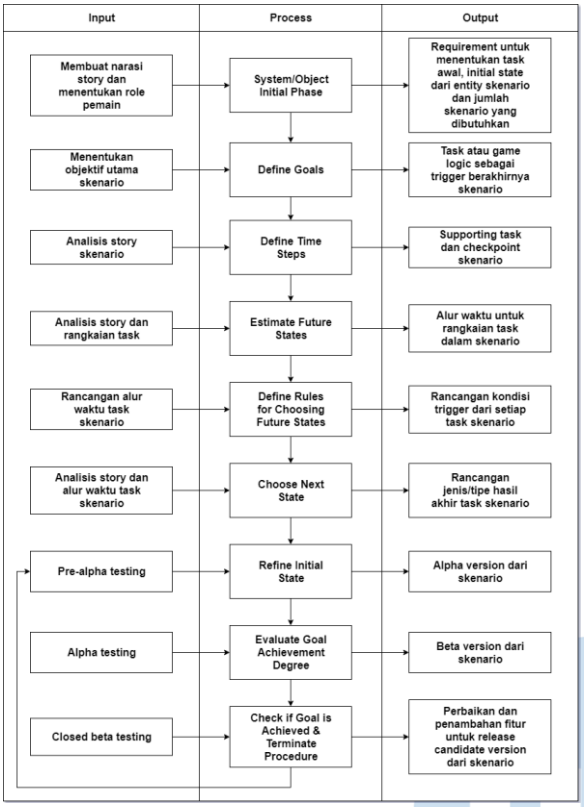


Fig. 3. Scenario development phases

C. *Data Analysis Methodology*

The data used for analysis were obtained from 2 sources: data from answering questionnaires and data from scenario completion. Data is divided into 3 test variables, each with four parameters [11]:

- Number of civilians killed by players directly.
- Number of prisoners of war that successfully secured by the player.
- Number of enemies the player identified before carrying out the attack.
- Number of civilians killed by players directly.

Variable 2 discusses post-conflict restoration and repair, measured by the following parameters:

- Number of scores the player gets in 1 specific scenario for this variable.
- Total score of successful humanitarian assistance given by players.
- Number of mine areas that the player has cleared.
- Total score of successful humanitarian assistance given by players.

Variable 3 discusses the protection and requirements of medical staff in conflict, measured by the following parameters:

- Total score of player's suitability in using medical equipment by international law.
- Total score for the level of use of weapons and ammunition by international law.
- Total score of the player participation rate in offensive and provocative attacks.
- Number of medical staff killed or medical equipment destroyed by players.

The score of each parameter in the questionnaire data is measured on a scale of 1 to 5, and the scenario data is measured on a scale of 1 to 10, which is then categorized into a scale of 1 to 5 according to Table 1 so that it can be compared with questionnaire data [12].

TABLE I. CATEGORY SCALE SCORE DATA SCENARIO

Score	0	1	2	3	4	5	6	7	8	9	10
Category Scale	1			2		3		4		5	

Data that has been wholly collected is analyzed through 7 calculation steps as follows:

- Determine max score value.

$$MS = RespondentCount \times 5 \quad (1)$$

- Calculate the percentage of the player's knowledge level of each parameter.

$$SP_n = \left(\sum P_n\right)/MS \quad (2)$$

- Calculate the percentage of the player's knowledge level from variables 1 to 3.

$$SVP1 = \left(\sum_{n=SP1}^{SP4} n\right)/4 \quad (3)$$

$$SVP2 = \left(\sum_{n=SP5}^{SP8} n\right)/4 \quad (4)$$

$$SVP3 = \left(\sum_{n=SP9}^{SP12} n\right)/4 \quad (5)$$

- Calculate the percentage of players' level of compliance from each parameter.

$$SK_n = \left(\sum P_n\right)/MS \quad (6)$$

- Calculate the percentage of the player's level of compliance from variables 1 to 3.

$$SVK1 = \left(\sum_{n=SK1}^{SK4} n\right)/4 \quad (7)$$

$$SVK2 = \left(\sum_{n=SK5}^{SK8} n\right)/4 \quad (8)$$

$$SVK3 = \left(\sum_{n=SK9}^{SK12} n \right) / 4 \quad (9)$$

- SPn, SKn, SVPn and SVKn values will be grouped according to Table 2, as shown below:

TABLE II. PERCENTAGE VALUE CATEGORIES

Percentage (%)	Category
0 - 20	Very Bad
21 - 40	Bad
41 - 60	Moderate
61 - 80	Good
81 - 100	Very Good

- Comparing the value of the player's knowledge level and the player's compliance level to determine whether the player's knowledge level has relevance to the level of compliance by finding the difference in the percentage.

$$D_n = SK_n - SP_n \quad (10)$$

The percentage relevance values are grouped by category in Table 3.

TABLE III. RELEVANCE VALUE CATEGORY

Percentage (%)	Category
0 - 20	Very Bad
21 - 40	Bad
41 - 60	Moderate
61 - 80	Good
81 - 100	Very Good

III. RESEARCH AND RESULTS

A. *Game Scenario*

The use case diagram in Figure 4 of the scenario developed for this research. Interaction in the scenario is only done by the player and the system to create automation so that researchers are not involved when players play video game scenarios.

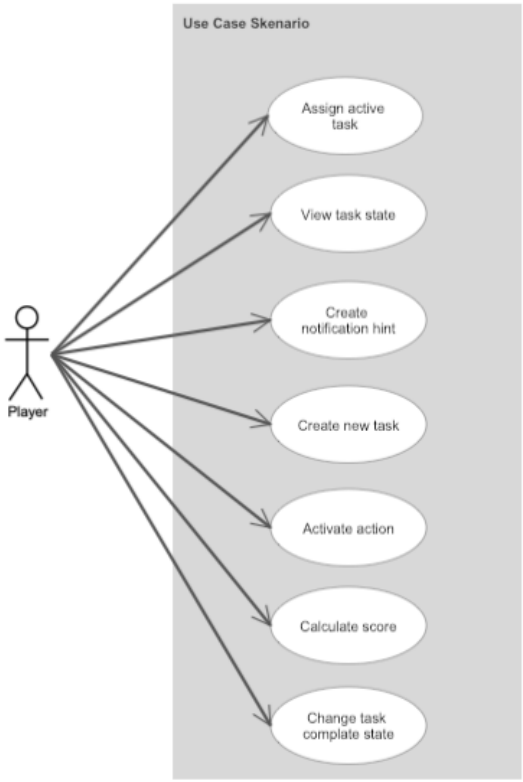


Fig. 4. Scenario use case diagram

Table 4 is an example of a task list for one of the developed scenarios. The left or centre column is the parent task, and the right column is given if the parent task has child tasks.

TABLE IV. TASKS LIST

Supplies delivery	Lift cargo (Building)
	Deliver cargo (Building)
	Lift cargo (Medical)
	Deliver cargo (Medical)
	Preserve cargo (O)
Demining	Connect to UAV
	Destroy mines
	Keep drones (O)
	End operation
Help IDAP	Talk to IDAP chief
	Transport patient
Help Altis PD	Talk to Commander
	Capture HVT
	Confirm & secure
Search & Rescue	Talk to team leader
	Find the convoy
	Talk to wounded police
	Find IDAP worker
	Rescue IDAP worker
Report	
Collateral damage	
Fire support ROE	

In the video game scenario, there is a scoring system to calculate player scores automatically. Each parameter in the research variable has a different scoring rule. Scoring system rules are created in the SQL script that is executed in the video game trigger.

The score of parameter number 3 in variable number 2 is used as an example, where the score for this parameter is obtained by the player when carrying out a demining task. This task instructs the player to clear the active mine area.

Fig. 5. Active mine area

The mine area in the scenario is built into five areas, as shown in Figure 5. Each area has a circular radius with dimensions of 50 meters x 50 meters, and 1 area has ten mines. Map markers are given to each mining area, with a red checkered line indicator if the area is still active or has not been cleared of mines.

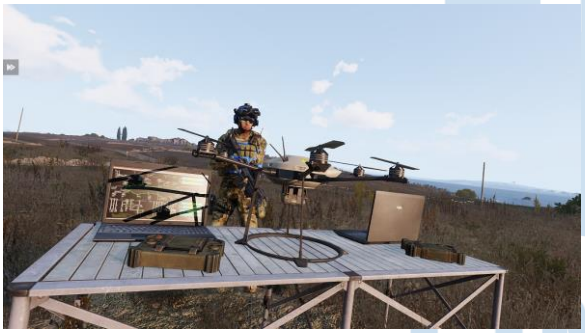


Fig. 6. Demining drone

The player's choice in choosing how to clear mines is not limited to scenarios such as approaching mines and turning them off manually or using demining drones, as shown in Figure 6.

```
Trigger: Expression
Condition
{["mine_", str(_x)] call BIS_fnc_inString} count (thisTrigger nearObjects
55) < 1;
On Activation
score_demining = score_demining + 2;
minefieldclear = minefieldclear + 1;
"marker_mine_area_1" setMarkerColor "ColorBlue";
hint "This minefield is clear";
```

Fig. 7. Mine area trigger

A trigger is used to determine if the area is clear of mines. Each area has a trigger, like in Figure 7, which is placed in the centre position of the area. The trigger will function as a counter to count the number of active mine objects in a circle radius of 55 meters x 55 meters from the trigger position.

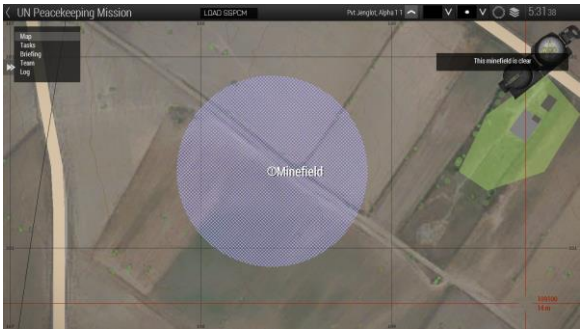


Fig. 8. Inactive mine area

If the number of mines in the trigger area is less than 1, then the mine area of the trigger is marked as clear of mines, the checkered line markers will turn blue as in Figure 8, and the player's points increase by 2. The variable in the script for this score is "score_demining". The score will be displayed to players when the scenario ends and exported into a log file for analysis in the next phase of this research.

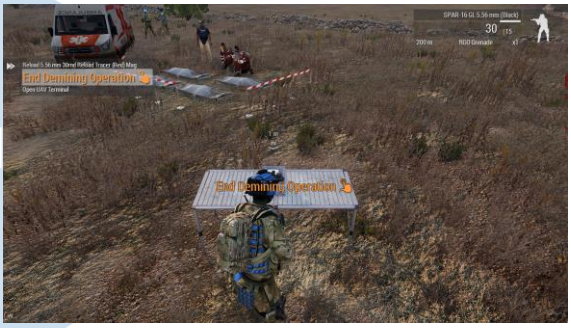


Fig. 9. Inactive mine area

Players must clear one area of the mine, while the other area is optional. To end the task, the player uses the interface in the laptop object on the scenario, as shown in Figure 9. If the player terminates this task using that interface, then the task cannot be resumed even if the player clears the remaining minefield area that has not been cleared.

B. Analysis Result

The steps are taken for testing the research subjects in figure 10. The subject must run the game tutorial first. If the subject does not experience difficulties and can pass the tutorial, then the subject can enter the testing phase. What the subject did at the beginning of the test was to fill in 12 points of a research questionnaire to measure the subject's level of knowledge of international law in armed conflict. After completing the questionnaire, the subject will run the whole scenario in 3 or more testing sessions.

After the test results in the form of scores from 20 subjects are obtained, the test results will be analyzed based on variables, parameters and methods that have been determined for this research. Before analyzing the variables and parameters of the test results, the max score is determined first. The Max score for analysis is

100, which is obtained from the multiplication between the number of respondents of 20 and the maximum score scale of 5.

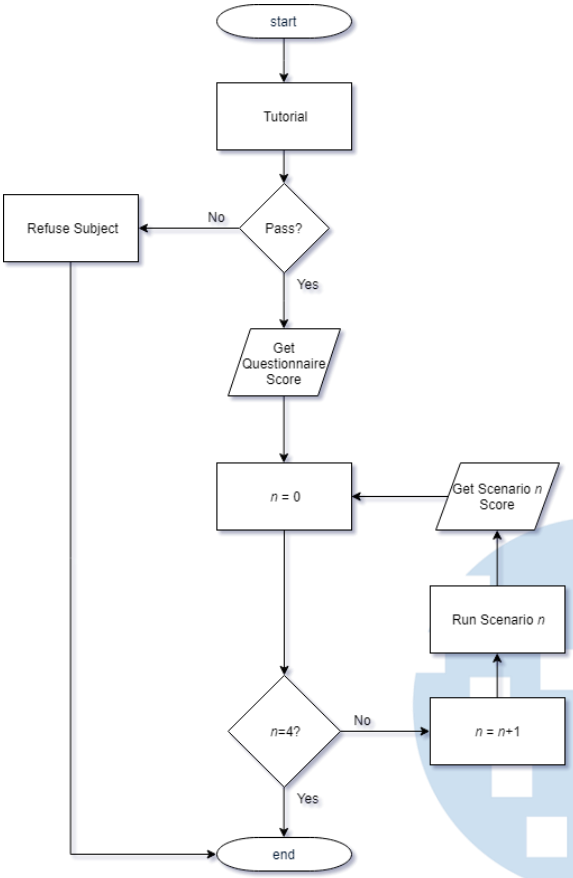


Fig. 10. Subject testing flowchart

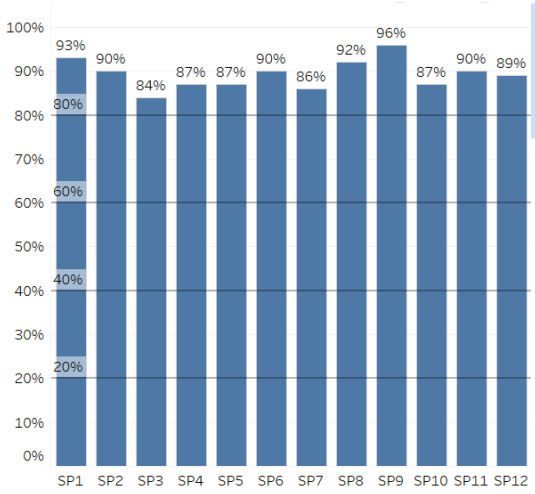


Fig. 11. Percentage of knowledge score of each parameter

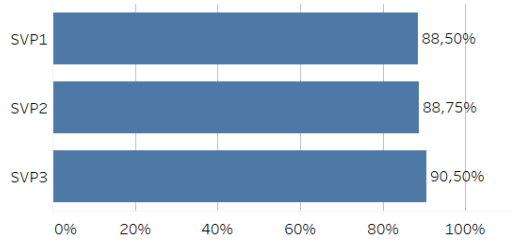


Fig. 12. Percentage of knowledge score of each variable

Based on the results taken from the questionnaire test, the subject met the excellent category for all parameters. In Figure 11, each parameter has an average score scale of 4.5. In figure 12, the percentage of subject knowledge level scores for variable 1 is 88.5%, variable 2 is 88.75%, and variable 3 is 90.5%. All variables fall into the excellent category. These results show that the subject has a very high knowledge of international law in armed conflict.

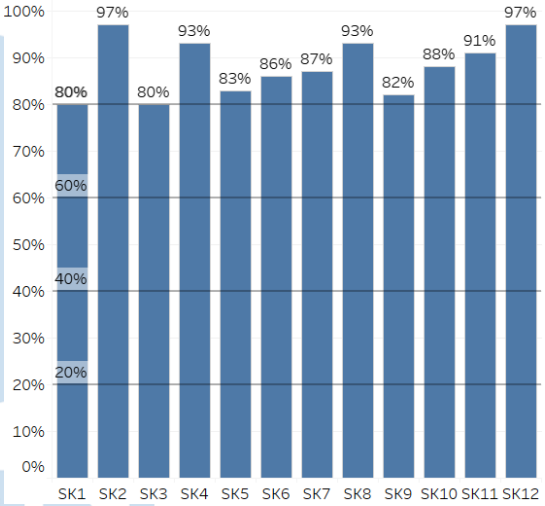


Fig. 13. Percentage of compliance score of each parameter

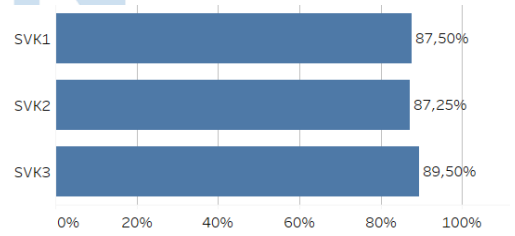


Fig. 14. Percentage of compliance score of each variable

Based on the results taken from test scenarios, subjects meet excellent categories in parameters 2, 4, 5, 6, 7, 8, 10, 11, and 12 and good categories in parameters 1 and 3. However, the percentage difference in parameters 1 and 3 is only 1% to reach the outstanding category. In Figure 13, each parameter has a score above 80%, with an average score scale of 4.4. In figure 14, the percentage of the subject's compliance level score for variable 1 is 87.5%, variable 2 is 87.25%, and variable 3 is 89.5%. All variables are included in the outstanding category. These results indicate that the

subjects have a very high level of compliance with international law in armed conflict.

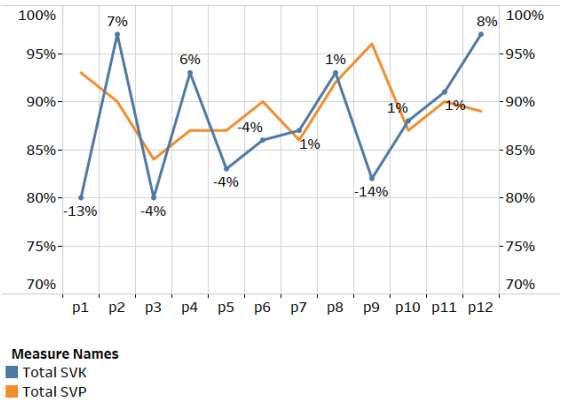


Fig. 15. Score comparison

The graph in Figure 15 and the results in Table 5 show the relevance between the knowledge level and subject compliance. From these graphs and tables, it can be concluded that the level of knowledge in parameters one and 9 are not more relevant to its level of compliance than the level of knowledge in parameters 2, 4, and 8 are still relevant to its level of compliance. The level of knowledge in parameters 3, 5, 6, 7, 8, 10, and 11 is very relevant to its level of compliance. The subject's level of compliance in parameters 1, 3, 5, 6, and 9 moves in a negative direction from its level of knowledge, while the subject's level of compliance in parameters 2, 4, 7, 8, 10, 11, and 12 moves in a positive direction from its level of knowledge.

TABLE V. RELEVANCE SCORE

Variable	Params	Difference Percentage	Relevance value category	Difference Bearing
V1	P1	13%	Irrelevant	Negative
	P2	7%	Relevant	Positive
	P3	4%	Strongly Relevant	Negative
	P4	6%	Relevant	Positive
V2	P5	4%	Strongly Relevant	Negative
	P6	4%	Strongly Relevant	Negative
	P7	1%	Strongly Relevant	Positive
	P8	1%	Strongly Relevant	Positive
V3	P9	14%	Irrelevant	Negative
	P10	1%	Strongly Relevant	Positive
	P11	1%	Strongly Relevant	Positive
	P12	8%	Relevant	Positive

C. *Analysis Result*

The test results show that respondents have an outstanding level of compliance with international humanitarian law, and most parameters have an outstanding level of compliance. This Discussion emphasizes the results of parameters with irrelevant criteria in Table 5. The results of the comparison between the level of knowledge and the level of compliance show that the level of knowledge of players in parameters 1 and 9 does not reflect its level of compliance.

The irrelevance of the respondent's level of compliance with the level of knowledge in parameter 1 is because the respondent tends to violate international humanitarian law by killing civilians. In contrast, previously, the respondent strongly agreed that civilians must be protected in the conflict zone so that the difference in knowledge and compliance moves negatively.

The irrelevance of the respondent's level of compliance with the level of knowledge in parameter 9 is because the respondent tends not to use equipment with protective symbols when acting as a combat medic in conflict. In contrast, previously, the respondent strongly agrees that medical personnel must use equipment with protective symbols that can be seen by the parties involved in the conflict so that the direction of differences in the level of knowledge and obedience moves in the negative direction.

For other parameters, respondents' knowledge level also reflects their level of compliance. It can be said to be relevant or very relevant so that the direction of difference does not have significant meaning.

IV. CONCLUSIONS

The application of the scenario-based framework method in developing scenarios for the video game ARMA 3 has successfully created scenarios that can measure the level of a player's compliance with international law in armed conflict.

From the analysis results, the subject has a very high level of knowledge and compliance with international law in armed conflict, with an average percentage level of knowledge and compliance level above 80%. Most parameters have a percentage difference below 5%, so the subject's level of knowledge is very relevant and influences the level of compliance as a whole. However, from the overall results, there are two exciting parameters, parameter 1 (variable 1) and parameter 9 (variable 3).

Both of these parameters result in subject knowledge that is irrelevant to their knowledge and moves in a negative direction. The value of parameter 1 is irrelevant because the subject has difficulty distinguishing between military personnel and civilians when engaging in armed contact, where the armed contact situation forces the subject to act quickly so that the subject does not have time to distinguish between military personnel and civilians. Then the value in parameter 9 is irrelevant because the subject does not

know in detail the form of equipment and vehicles that are mandatory and allowed to be used for combat medical personnel where combat medic personnel are medical personnel who are included in military organizations so that the forms of equipment and vehicles used are different from medical personnel civil that is known to most of the subjects.

ACKNOWLEDGMENT

Author would like to thank Universitas Multimedia Nusantara for supporting this research.

REFERENCES

[1] I. C. o. t. R. Cross, "The Geneva Conventions of 1949 and their Additional Protocols," 2019. [Online]. Available: www.icrc.org/en/document/geneva-conventions-1949-additional-protocols. [Accessed 3 July 2019].

[2] E. S. Bates, "Towards effective military training in international humanitarian law," *International Review of the Red Cross*, vol. 96, pp. 795-816, 2018.

[3] S. D. a. M. W. Brad A. Gutierrez, "What Americans think of international humanitarian law," *International Review of The Red Cross*, vol. 93, pp. 1009-1034, 2017.

[4] C. R. a. F. , S. . . Ben Clarke, "Beyond the Call of Duty: why shouldn't video game players face the same dilemmas as real soldiers?," *Internasional Review of the Red Cross*, vol. 94, pp. 711-737, 2017.

[5] E. Lahti, "Arma 3 announced," 2019. [Online]. Available: https://www.pcgamer.com/arma-3-announced-releasing/. [Accessed 6 February 2019].

[6] M. Sassoli, "International Humanitarian Law: Rules, Controversies, and Solutions to Problems Arising in Warfare," *International Review of the Red Cross*, vol. 101, pp. 951-958, 2019.

[7] J. Bikovska, "Scenario Development Approach to Management Simulation Games," *Information Technology and Management Science*, vol. 17, pp. 144-149, 2017.

[8] S. Dannenberg, "Gaming Scenarios: Making Sense of Diverging Developments," *Journal of Futures Studies*, vol. 22, pp. 15-26, 2017.

[9] A. Slootmaker, H. Kurvers, H. Hummel and R. Koper, "Developing Scenario-based Serious Games for Complex Cognitive Skills Acquisition: Design, Development and Evaluation of the EMERGO Platform," *Journal of Universal Computer Science*, vol. 20, pp. 561-582, 2019.

[10] J. Bikovska and G. Merkuryeva, "Scenario-Based Planning And Management Of Simulation Game," in *21st European Conference on Modelling and Simulation*, United Europe, 2017.

[11] R. V. Eck, Gaming and Cognition: Theories and Practice from the Learning Sciences, Hershey, New York: Information Science Publishing, 2017.

[12] S. E. Harpe, "How to analyze Likert and other rating scale data," *Currents in Pharmacy Teaching and Learning*, vol. 7, pp. 836-850, 2015.

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		

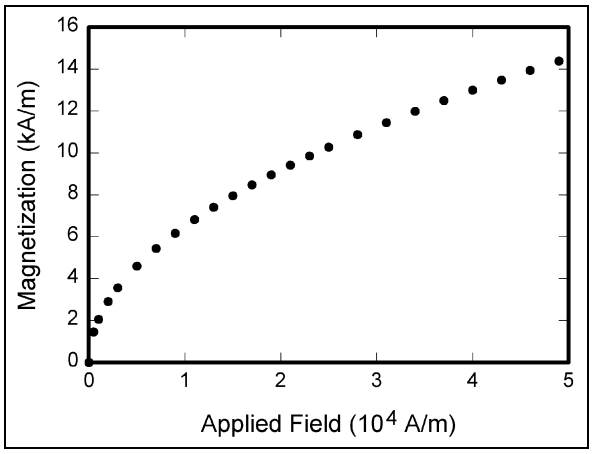


Fig. 1. Example of a figure caption

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