

## EXPLORING THE EFFECTIVENESS OF DEEP LEARNING IN ANALYZING REVIEW SENTIMENT

Mariyanto <sup>1\*</sup>, Hilman F. Pardede <sup>2</sup>

<sup>1,2</sup> Ilmu Komputer, Universitas Nusa Mandiri, Jakarta, Indonesia  
\*Email: <sup>1</sup>14210250@nusamandiri.ac.id, <sup>2</sup>hilm003@brin.go.id

(Received: 14 July 2023, Revised: 21 July 2023, Accepted: 31 July 2023)

### Abstract

This study aimed to analyze sentiment in office product reviews by using word embedding with three neural network modeling approaches: Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU). Office product review data is taken from Amazon's reviews of office products covering a wide range of sentiments. Word embedding converts text into a numerical vector representation for neural network processing. Experimental comparison of this model reveals that CNN achieves the highest accuracy, 77.99%. The CNN model effectively extracts significant features from review text, improving sentiment classification performance. Although the LSTM and GRU models show satisfactory results, they do not match CNN performance. These findings demonstrate the effectiveness of word embedding and neural networks for sentiment analysis in office product reviews. This provides valuable insights for companies to improve their products based on user feedback from online reviews. Additionally, this research serves as a foundation for further advances in sentiment analysis across a wide range of other products and services.

**Keywords:** *word embedding, CNN, LSTM, GRU, sentiment analysis*

This is an open access article under the [CC BY](#) license.



\*Corresponding Author: Mariyanto

## 1. INTRODUCTION

Studying product reviews from customers holds immense importance in the business world. Customer reviews provide valuable insights into the perceptions, opinions, and experiences of consumers with specific products or services. This information is highly beneficial for companies in several ways.

Firstly, analyzing product reviews allows businesses to understand the strengths and weaknesses of their offerings. By studying customer feedback, companies can identify the aspects of their products that are well-received and appreciated, enabling them to further enhance and highlight these positive features. On the other hand, negative reviews provide valuable feedback on areas that need improvement, allowing companies to address any issues and make necessary adjustments to enhance customer satisfaction.

Studies have demonstrated the importance of analyzing customer reviews in order to improve the quality of products [1]. They conducted a study on

sentiment analysis of customer reviews for a range of products and found that companies can identify specific areas of improvement based on customer feedback. This approach enables companies to make data-driven decisions and enhance product development strategies.

Secondly, customer reviews serve as a powerful tool for reputation management. Positive reviews act as testimonials, showcasing the satisfaction and positive experiences of customers, which can significantly influence the purchasing decisions of potential buyers. Conversely, negative reviews highlight areas of concern and give companies an opportunity to respond and rectify any issues, demonstrating their commitment to customer service and resolving problems.

Customer reviews have been shown to play an important role in managing online reputations [2]. They found that companies that actively engage with customers and respond to their feedback, both positive and negative [16], can improve their online reputation and build trust with consumers.

Furthermore, studying product reviews enables businesses to gain a competitive edge in the market. By analyzing customer feedback, companies can identify emerging trends, preferences, and demands, allowing them to tailor their products and services to meet the evolving needs of their target audience. This customer-centric approach can help companies stay ahead of the competition, attract new customers, and retain existing ones.

In conclusion, studying product reviews from customers is crucial for businesses to gain insights, improve their offerings, manage their reputation, and stay competitive. By actively analyzing and leveraging customer feedback, companies can make informed decisions, enhance customer satisfaction, and ultimately drive business growth.

This study focuses on the analysis of a dataset consisting of reviews of office products from Amazon annotated with sentiment labels. The dataset undergoes pre-processing steps, such as text cleansing, tokenization, and stopword removal. To convert textual data into numeric vectors, word embedding, which captures the semantic meaning of the word, is used.

This study focuses on CNN and LSTM models in office product reviews. In addition to the CNN and LSTM models, the author introduces the GRU+Multihead method. By combining a Gated Recurrent Unit (GRU) architecture with Multihead Attention, this novel method enhances the model's ability to capture sequential information and pay attention to relevant features in office product review text.

The experimental results show that the CNN model is most effective for sentiment analysis of office product reviews. In addition, the LSTM and GRU methods show definite performance, with performance almost on par with CNN in terms of accuracy and other evaluation metrics.

## 2. RELATED WORK

The study [3] presents a novel approach to movie review sentiment analysis using convolutional neural networks (CNN) and short-term memory networks (LSTM). The main objective of this study is to improve the accuracy of sentiment analysis in the context of film reviews. The study began by collecting a dataset of film reviews, including positive and negative sentiment labels. The weights and biases of the CNN-LSTM hybrid model were optimized using backpropagation and gradient descent to minimize the loss function and increase the accuracy of sentiment prediction [16]. The experimental results show that the CNN-LSTM hybrid model achieves higher accuracy and better sentiment analysis performance than the traditional approach. The integration of CNN and LSTM enabled the model to effectively capture local and sequential features, resulting in more accurate sentiment predictions for film reviews.

Contextual insights can be used to explore product reviews using LSTM networks and word embedding [4]. This research harnesses the power of deep learning techniques to gain a deeper understanding of product context and sentiments. Word embeddings such as Word2Vec or GloVe are used to represent words as numeric vectors that capture semantic information. The experimental results showed that the in-depth LSTM model, along with word embedding, effectively generated contextual insights from product reviews. The model achieves high accuracy in sentiment prediction and provides valuable contextual information that helps understand review nuances.

## 3. RESEARCH METHOD

By expressing words as dense numerical vectors and using machine learning, word embedding revolutionizes the analysis of product reviews. This is done by allowing computers to comprehend the context and sentiments represented in the text. Businesses can extract specific product features customers talk about and positive and negative feelings from customer feedback. This is done using techniques like sentiment analysis and aspect-based sentiment analysis. Companies may automate and expand their review research and make data-driven decisions to increase customer satisfaction and their products and services by utilizing word embedding and machine learning.

### 3.1 Data Source

The "Office product reviews" dataset, available at the Stanford Network Analysis Project (SNAP) website, contains 800,357 reviews of office products from Amazon. It includes information about the reviewers, products, ratings, helpfulness, and timestamps. This dataset is valuable for studying customer feedback and sentiment towards office products, and it can be used for tasks such as sentiment analysis and understanding customer preferences in the office product market.

The cleaning process was performed to remove incomplete or inaccurate data by removing null data. This is critical because incomplete or inaccurate data can affect model performance. After the cleaning process was completed, the dataset was divided into two parts: the training set and the test set. The training set consisted of 80% of the data, whereas the test data consisted of 20% data. The training set was used to train the model, and the test set was used to evaluate the model's performance.

### 3.2 Analysis of Variance (ANOVA)

ANOVA, or Analysis of Variance, is a statistical technique used to compare means across two or more groups and determine if there are significant differences between them. It is widely employed in data science and various research fields to analyze the impact of categorical independent variables on a continuous dependent variable.

In ANOVA [5][11], the total variation in the data is partitioned into two components: variation within each group and variation between the groups. The analysis aims to determine whether the observed between-group differences are statistically significant or simply due to random variation. ANOVA calculates an F statistic, which is used to assess the significance of these differences.

ANOVA offers several advantages in data science analysis. It allows for the simultaneous comparison of multiple groups, providing a more efficient and comprehensive analysis compared to conducting pairwise t-tests. ANOVA also provides insights into the magnitude of the observed differences and can account for potential interaction effects between independent variables.

### 3.2 Machine Learning

Deep learning is a potent branch of machine learning that focuses on building multi-layered artificial neural networks that analyze and comprehend complex data representations. Deep learning algorithms can extract high-level characteristics from enormous volumes of data and learn complex patterns by utilizing these powerful neural networks. This enables machines to execute tasks like speech recognition, natural language processing, and image recognition with remarkable accuracy. Deep learning has significantly advanced many disciplines, including computer vision, natural language processing, and data analytics. This has paved the way for creative applications and game-changing developments in artificial intelligence. Deep learning can automatically learn from data and adapt.

#### Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) [12] are a class of deep learning models widely used for various computer vision tasks, including image classification and object recognition. CNNs have revolutionized the field of computer vision due to their ability to automatically learn meaningful representations from raw pixel data. The core concept behind CNNs is the convolutional layer, which applies a set of learnable filters or kernels to the input image, extracting local features while preserving spatial relationships. The formula for the convolution operation in a CNN can be expressed as follows:

$$C(i, j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i + m, j + n) \cdot K(m, n) \quad (1)$$

where  $C(i, j)$  represents the output at position  $(i, j)$ ,  $I(i + m, j + n)$  is the input pixel value at position  $(i + m, j + n)$ ,  $K(m, n)$  denotes the kernel weight at position  $(m, n)$ , and  $M$  and  $N$  represent the size of the kernel.

The research of one of the researchers contributed to the development of CNN [6]. The authors introduced the LeNet-5 architecture, which featured a convolutional layer followed by pooling layers and

fully connected layers. This pioneering work demonstrated the effectiveness of CNNs in handwritten digit recognition and laid the foundation for subsequent advancements in the field.

Since then, numerous variations and improvements to CNN architectures have been proposed. In particular, the visual geometry group (VGG) [7] network achieves state-of-the-art performance on ImageNet datasets by leveraging a deeper architecture with smaller convolutional filters. This work highlighted the importance of depth in CNNs and inspired the development of deeper networks, such as the popular ResNet and Inception models.

In conclusion, CNNs have revolutionized computer vision by automatically learning meaningful features from raw pixel data. They have become fundamental tools for various tasks, and their effectiveness has been demonstrated through innovative works such as LeNet-5 [6] and VGG networks [7].

#### Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture that is particularly effective in modeling sequential data, such as natural language processing and time series analysis. LSTMs address the vanishing gradient problem associated with traditional RNNs by incorporating memory cells and gates that control the flow of information. This enables LSTMs to capture long-term dependencies and learn complex patterns over extended sequences.

The key components of an LSTM cell include an input gate ( $i_t$ ), a forget gate ( $f_t$ ), an output gate ( $o_t$ ), a memory cell ( $c_t$ ) and a hidden state ( $h_t$ ). The formulas governing the operations of an LSTM cell are as follows:

$$\begin{aligned} i_t &= \sigma(W_{ix} \cdot x_t + W_{ih} \cdot h_{(t-i)} + b_i) \\ f_t &= \sigma(W_{fx} \cdot x_t + W_{fh} \cdot h_{(t-i)} + b_f) \\ o_t &= \sigma(W_{ox} \cdot x_t + W_{oh} \cdot h_{(t-i)} + b_o) \\ g_t &= \tanh(W_{gx} \cdot x_t + W_{gh} \cdot h_{(t-i)} + b_g) \\ c_t &= f_x \cdot c_t + i_t \cdot g_t \\ h_t &= o_t \cdot \tanh(c_t) \end{aligned} \quad (2)$$

where  $x_t$  represents the input at time step  $t$ ,  $\sigma$  denotes the sigmoid activation function,  $W$  and  $b$  are the weight matrices and bias vectors, and  $\cdot$  indicates matrix multiplication.

The LSTM architecture was introduced in 1997 [8]. Their groundbreaking paper proposed the LSTM model as a solution to the vanishing gradient problem in RNNs. They demonstrated the effectiveness of LSTM networks in various tasks, including speech recognition and language modeling, highlighting their ability to capture long-term dependencies.

Since its inception, LSTM has become one of the most popular and widely adopted RNN architectures. Its effectiveness has been demonstrated in numerous applications, including machine translation, sentiment analysis, and speech recognition, among others.

In conclusion, LSTM networks have revolutionized the modeling of sequential data by addressing the vanishing gradient problem and capturing long-term dependencies [13]. The LSTM architecture [8], has become a fundamental tool in various fields and continues to drive advancements in sequence modeling.

**Gated Recurrent Unit (GRU)**

Gated Recurrent Unit (GRU) is a variant of the recurrent neural network (RNN) architecture that addresses the vanishing gradient problem and offers an efficient alternative to Long Short-Term Memory (LSTM) networks. GRUs utilize gating mechanisms to control the flow of information through the network, enabling them to capture and propagate relevant information over long sequences [14].

Similar to LSTM, GRU cells consist of an update gate  $z_t$ , a reset gate  $r_t$ , and a candidate activation  $h'_t$ . The formulas governing the operations of a GRU cell are as follows:

$$\begin{aligned}
 z_t &= \sigma(W_{zx} \cdot x_t + W_{zh} \cdot h_{(t-i)} + b_z) \\
 r_t &= \sigma(W_{rx} \cdot x_t + W_{rh} \cdot h_{(t-i)} + b_r) \\
 h'_t &= \sigma(W_{hx} \cdot x_t + W_{hh} \cdot (r_t \odot h_{t-1} + b_h)) \\
 h_t &= (1 - z_t) \odot h_{(t-1)} + z_t \odot h'_t
 \end{aligned}
 \tag{3}$$

where  $x_t$  represents the input at time step  $t$ ,  $\sigma$  denotes the sigmoid activation function,  $\odot$  indicates element-wise multiplication, and  $W$  and  $b$  are the weight matrices and bias vectors.

It was introduced in 2014 [1] that the GRU architecture was developed. Their seminal paper presented GRUs as a simplified variant of LSTM, demonstrating comparable performance on various sequential tasks while requiring fewer parameters. GRUs were shown to effectively capture long-term dependencies and exhibit better training efficiency than LSTM.

Since its introduction, GRU has gained widespread popularity and has been successfully applied to various domains, including natural language processing, speech recognition, and machine translation. Its efficiency, simplicity, and competitive performance have made GRU a popular choice for modeling sequential data [15].

In summary, GRU networks offer an efficient and effective solution for modeling sequential data. GRUs utilize gate mechanisms to capture long-term dependencies while maintaining computational efficiency [9]. Their simplicity and competitive performance have established GRUs as a valuable tool in various applications.

**4. RESULT AND DISCUSSION**

The Figure Data distribution for the review class, illustrates the data distribution for the "Overall" class in the dataset used in this study. This data distribution provides an overview of the sample distribution across the different "Overall" classes present in the dataset. In sentiment analysis and product review classification,

the "Overall" class represents the overall sentiment or rating of a product. This data distribution provides valuable insights into the balance or imbalance of sample sizes within each "Overall" class.

In Figure 1, the x-axis represents the available "Overall" classes, while the y-axis indicates the number of samples in each class. Data distribution reflects a balanced or imbalanced distribution. If there is a significant difference in the number of samples across each "Overall" class, the data distribution shows varying bar heights on the graph. Understanding the data distribution for the "Overall" class is crucial for comprehending the relationship between the sample sizes within each class and the classification outcomes generated by the model. In the case of imbalanced data distribution, the classification results in the majority class tend to have a greater impact on the overall model evaluation.

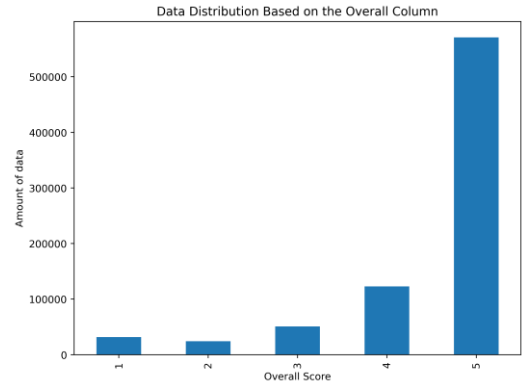


Figure 1: Data distribution for the review class

The results of the modeling using the ANOVA method are shown in ANOVA Test Results. This provides a clear visualization of the features that significantly influence the target and aims to evaluate the relative importance of each feature to the target variable to be predicted. If the ANOVA significance value is high, this indicates that the feature has a significant influence on the target. The visualization presented in Figure 2 shows that the "summary" feature has the greatest influence on the target (overall). Furthermore, the data used in this study only uses the "summary" and "overall" features for word embedding processing.

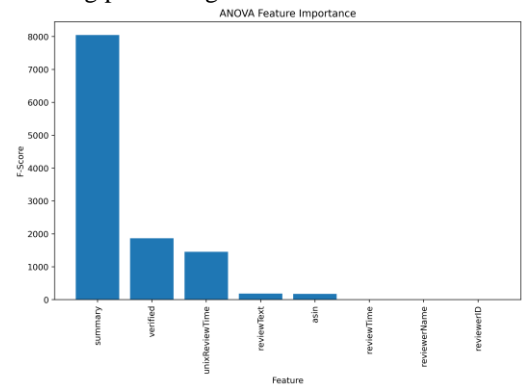


Figure 2. ANOVA Test Results

Figure Results of machine learning model accuracy, displays a visual representation of the test results obtained from the analysis of office product reviews using the LSTM, CNN, and GRU methods. The purpose of this test is to compare the performance of the three methods in the sentiment analysis of office product reviews. The results depicted in Figure 3 provide an overview of the relative performance of LSTM, CNN, and GRU in terms of accuracy. Through this visualization, it is easy to observe and compare the test results of the three methods.

Based on Figure 3, several conclusions can be drawn regarding this method's performance. For example, if CNN exhibits higher accuracy than LSTM and GRU, it may be more effective at office product review sentiment analysis. CNN's highest accuracy of 77.99% was obtained in the 3rd epoch. On the other hand, LSTM or GRU also performed well although not as well as CNN in this experiment. The highest accuracy value for LSTM is 77.96% while the highest accuracy value for GRU is 77.97%. Based on this data, it is evident that the accuracy of the three models is relatively equal, and that they are suitable for sentiment analysis of office products.

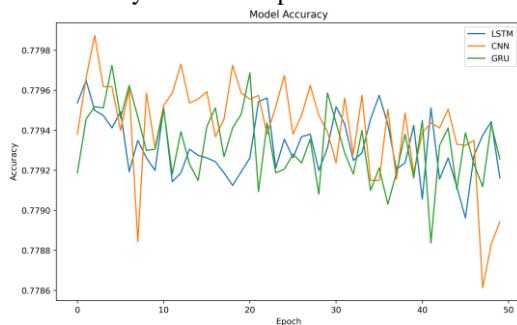


Figure 3. Results of machine learning model accuracy

In Figure Loss generated by machine learning, the x-axis indicates the number of epochs or iterations performed during training, and the y-axis indicates the model loss value. The model loss change curve in Figure 4 provides information about the extent to which the model experiences a decrease in error as the training progresses.

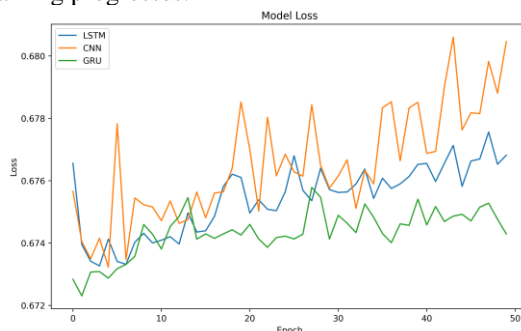


Figure 4. Loss generated by machine learning

When the model loss curve increases as the training progresses, this indicates that the model cannot reduce errors or errors during the learning

process. This can be caused by several factors, such as the complex model architecture, dataset size that is too small, or training conditions that are not optimal.

An increase in model loss may indicate that the model is unable to learn the patterns in the dataset properly or may experience overfitting, where the model "memorizes" the training data too much, so that it cannot be generalized properly to new data.

In the process of training a neural network, we can compare the time taken for an epoch, which represents a complete training cycle, across the three different types of architectures. Let us now examine the average time estimates for an epoch in each architecture. LSTM networks typically require approximately 152 s per epoch. LSTMs are generally used to process sequential data, and are known for their ability to handle long-term dependencies. The CNN architecture typically completed an epoch of approximately 120 s. The GRU network, similar to LSTM, showed an average epoch of approximately 148 s. The GRU model is often used for sequential data modeling and has the advantage of maintaining short-term information.

## 5. CONCLUSION

The test results show that three methods, namely LSTM, CNN, and GRU play a significant role in the sentiment analysis of office product reviews. Each method has strengths and weaknesses when dealing with this task. CNN has the highest accuracy with the shortest epoch cycle time. LSTM accuracy is the lowest and takes the longest time. GRU has medium accuracy with medium epoch time.

The selection of an appropriate method depends on the specific context of the sentiment analysis and the characteristics of the dataset. Based on the test results, it is suggested that LSTM, CNN, and GRU be considered for sentiment analysis of office product reviews, choosing the method that best suits the research objectives and requirements.

This study serves as an initial foundation for the further development of sentiment analysis of office product reviews. Future research may involve exploring other methods, optimizing models, incorporating additional features, or adapting to more specific contexts and types of product review.

## 6. REFERENCES

- [1] J. Panthati, J. Bhaskar, T. K. Ranga, and M. R. Challa, "Sentiment Analysis of Product Reviews using Deep Learning," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2018, pp. 2408–2414. doi: 10.1109/ICACCI.2018.8554551.
- [2] R. Thakur, "Customer engagement and online reviews," *J. Retail. Consum. Serv.*, vol. 41, pp.

- 48–59, 2018, doi: <https://doi.org/10.1016/j.jretconser.2017.11.002>.
- [3] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, “A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis,” *Multimed. Tools Appl.*, vol. 78, no. 18, pp. 26597–26613, 2019, doi: 10.1007/s11042-019-07788-7.
- [4] N. Suryana, A. S. H. Basari, and others, “Generate Contextual Insight of Product Review Using Deep LSTM and Word Embedding,” in *Journal of Physics: Conference Series*, 2020, p. 12006.
- [5] X. Liang and J. Li, “Optimize Statistical Analysis via Propensity Score Matching and Repeated-Measures Analysis of Variance,” *Cardiovasc. Interv.*, vol. 16, no. 3, pp. 361–362, 2023.
- [6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.
- [7] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *3rd International Conference on Learning Representations (ICLR 2015)*, Computational and Biological Learning Society, 2015, pp. 1–14.
- [8] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [9] K. Cho *et al.*, “Learning Phrase Representations using {RNN} Encoder{--}Decoder for Statistical Machine Translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing ({EMNLP})*, Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734. doi: 10.3115/v1/D14-1179.
- [10] A. Vaswani *et al.*, “Attention is All you Need,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)
- [11] Yu, Z., Guindani, M., Grieco, S.F., Chen, L., Holmes, T.C. and Xu, X., 2022. Beyond t test and ANOVA: applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. *Neuron*, 110(1), pp.21-35.
- [12] Ding, X., Zhang, X., Han, J. and Ding, G., 2022. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11963-11975).
- [13] Zha, W., Liu, Y., Wan, Y., Luo, R., Li, D., Yang, S. and Xu, Y., 2022. Forecasting monthly gas field production based on the CNN-LSTM model. *Energy*, p.124889.
- [14] Zhang, W., Li, H., Tang, L., Gu, X., Wang, L., & Wang, L. (2022). Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks. *Acta Geotechnica*, 17(4), 1367-1382.
- [15] Zhang J, Jiang Y, Wu S, Li X, Luo H, Yin S. Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. *Reliability Engineering & System Safety*. 2022 May 1;221:108297.
- [16] Huda M, Supriyati E, Listiyorini T. Analisis sentimen pengguna youtube terhadap tayangan #matanajwamenantiterawan dengan metode naïve bayes classifier, JIKO (Jurnal Informatika dan Komputer), Vol. 5, No. 1, April 2022, hlm. 1-6, DOI: 10.33387/jiko