

# Shoe Review Sentiment Analysis Using Machine Learning and Deep Learning with Word2vec

Sugiyanti<sup>a1</sup>, Arjun Fricco<sup>b2</sup>

<sup>a</sup>Magister of Computer Science, Faculty of Information Technology, Nusa Mandiri University  
Kramat Raya Steet No.18 RW.7 Kwitang Senen, Jakarta Pusat, 10450, Indonesia  
<sup>1</sup>14210253@nusamandiri.ac.id

<sup>b</sup>Department of Information Technology, Faculty of Information Technology, Sekolah Tinggi Ilmu  
Komputer Cipta Karya Informatika  
Jl. Radin Inten II No.8, RT.5/RW.14, Duren Sawit, Kec. Duren Sawit, Kota Jakarta Timur,13440,  
Indonesia  
<sup>2</sup>arjunfricco923@gmail.com

## Abstract

In this modern era, internet use is continually increasing and will run with an increasing amount of existing data, such as text data. Characteristics of unstructured text are a challenge in text processing feature extraction and encourage sentiment analysis research to be carried out. The availability of a lot of text data on the internet is a challenge in sentiment analysis because it requires a complex approach. This study uses the baseline Deep Learning (DL) method, namely Long-Short Term Memory (LSTM) and Convolutional Neural Network (CNN) with word2vec. It uses Machine Learning (ML), namely Random Forest (RF), Support Vector Machine (SVM), and Gaussian Naïve Bayes (GNB) with the proposed method, word2vec, and the dataset used is the shoe review dataset which consists of 389,877 reviews. From the discussion above, to carry out sentiment analysis, a more suitable method is to use the CNN baseline method with word embedding word2vec to get an accuracy value of 91.53%. The novelty of this study is the increase in the classification accuracy value from previous studies.

**Keywords:** *Sentiment Analysis, Shoe Review, Word2vec, DL, ML*

## 1. Introduction

Technology in this modern era is proliferating and must be connected to increased number of internet users. An increase in the number of internet users can affect an increase in the amount of data on the internet. With the internet, people can interact with other people through various platforms through text, images, and videos, making it possible to share ideas with others. The availability of abundant data on the internet, such as text data, has encouraged many researchers to research text mining and NLP (Natural Language Processing) [1]. NLP is a Machine Learning (ML) technology enabling computers to interpret, analyze, and understand human language [2].

Shoes are items that are almost needed by every human being. Shoes are economic goods because they are used to fulfil human needs in carrying out their activities. In determining the shoes to be purchased, you can first look at reviews from other customers. For that, it is necessary to have a sentiment analysis of shoe reviews.

Sentiment analysis uses text analysis, computational linguistics, and NLP to obtain semantic quantification of information. Sentiment analysis is one of the popular NLP studies often carried out by researchers. Sentiment analysis aims to know the opinions of a text to identify emotional information expressed by users in texts such as a product review, which can be used for decision-making in planning and making a marketing decision, increasing the number of customers, and expanding the business. Textual sentiment involves two types of emotions: positive and negative. Reviews and ratings given to the assessment and input of a product will affect other customers to evaluate the product to be used, and customers can make choices. Reviews that customers have given sometimes have incomplete and inappropriate reviews. With a lot of review data, it can make it difficult for

companies to find out if their product sales have more positive or negative. For this reason, conscious analysis is needed to determine whether customer reviews are positive or negative.

The sentiment analysis process is influenced by the dataset used; if the dataset is large, maximum and different handling is required. The review data that has been collected is processed to determine the response from each existing review, whether the review is negative or positive. For this reason, sentiment classification is needed to classify user reviews into positive and negative opinions. Classification can be done using ML. ML is the science of developing algorithms and statistical models computer systems use to create complex applications that can accurately classify and predict various data. ML learns from data distribution to make judgments about the incoming data. Based on the input features, the created model or classifier can then determine the accuracy of the results. This Machine Learning (ML) algorithm is categorized into reinforcement, unsupervised, semi-supervised, and supervised learning [3]. Apart from the ML method, this study will use another form, Deep Learning (DL). DL is an Artificial Intelligence (AI) method that teaches computers to process data like human intelligence. DL is part of ML which can recognize patterns and unstructured information.

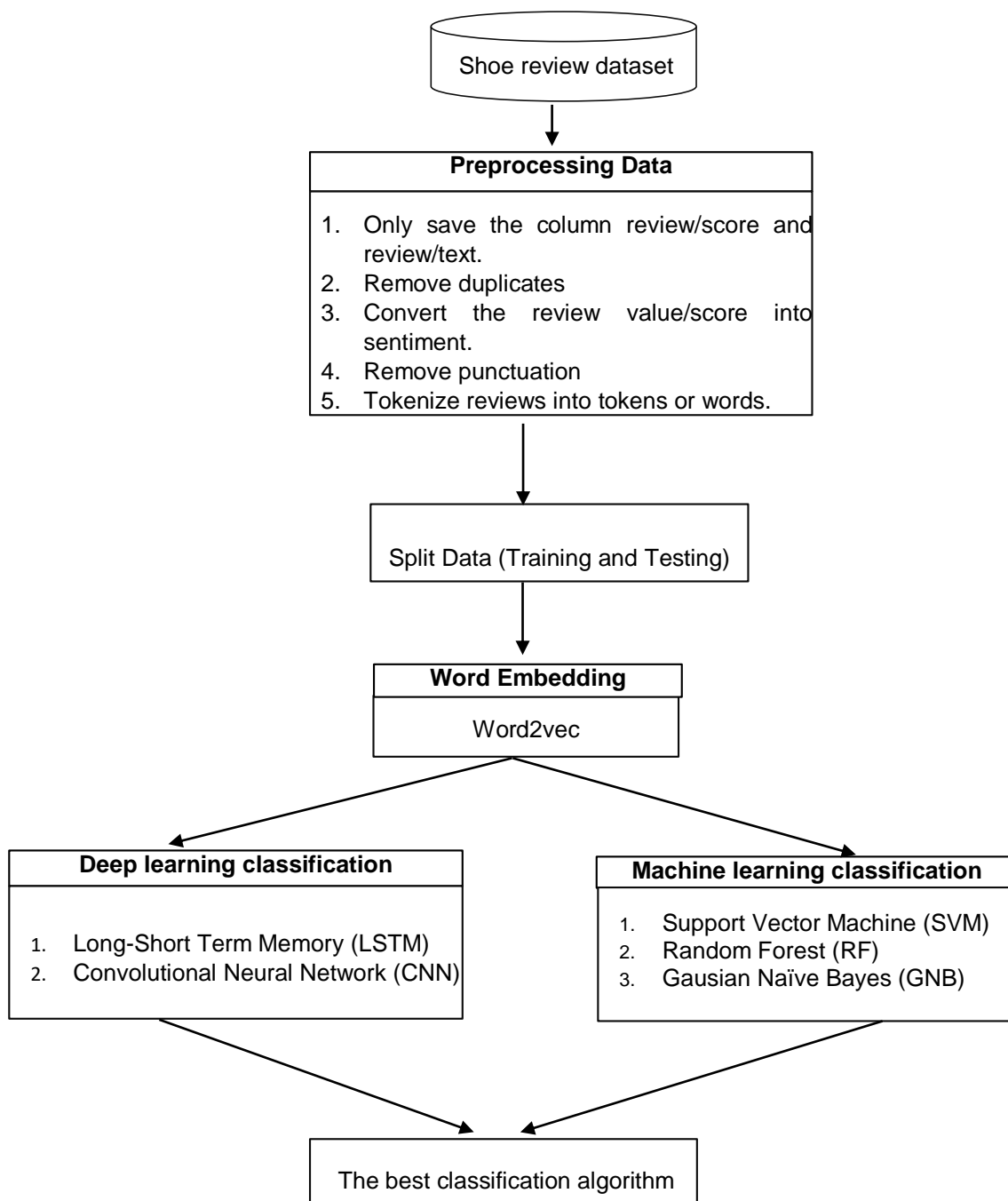
Feature engineering on textual data has the characteristics of unstructured text. For this reason, a feature engineering strategy is needed. One of the models that can be used is word embedding. Word embedding began to be developed around 2000. Word embedding works by mapping each word into a dense vector, and the vector will represent the word projection in that vector space. Word embedding can capture the meaning of word syntax and word semantics. One word embedding widely used by researchers is word2vec. Word2vec is a shallow neural network model that converts word representations into vectors [4]. Word2vec aims to group vectors of similar words in a vector space using a mathematical model so that word2vec can detect the exact meaning of the target. The word2vec model can be implemented using ML and DL [5].

Researchers have tested a lot of sentiment analyses built with various datasets about reviews. Research conducted by Murthy et al. in 2020 used Long Short Term Memory (LSTM) with datasets from IMDB and Amazon. The dataset is divided into train data and test data with the same comparison, namely 25,000 train data and 25,000 test data, and using the Adam optimizer, the results of this study produce an accuracy value of 85% [6]. In comparison, research conducted by A.Hassan et al. using deep learning called convLstm, which is a combination of the CNN and LSTM models and uses word2vec, where the dataset is taken from IMDB and Stanford, resulting in an accuracy value of 88.3% [7]. Research conducted by Chachra et al. used the Deep Convolution Neural Network (DCNN), and the dataset used was the Twitter dataset which had been labelled with positive and negative sentiments, and the result of this study was 80.69% [8].

In this study, we will explain the results of sentiment analysis on the shoe review dataset taken from Stanford using two DL methods, namely LSTM and CNN, using word2vec. In comparison, the second method uses ML, namely Random Forest (RF), Support Vector Machine (SVM), and Gaussian Naïve Bayes (GNB) using a word embedding word2vec. Most previous studies only used one research model, such as the DL method, and some did not use word embedding, such as word2vec. For this reason, this researcher will compare the shoe review sentiment analysis results using the ML and DL methods with word2vec with previous research.

## **2. Research Methods**

The research method can be seen in Figure 1, which illustrates the procedure followed in carrying out a shoe review sentiment analysis and obtaining the results:



**Figure 1.** Research Methodology

The steps taken in the first stage are collecting datasets to be used in research, then preprocessing the data, then splitting the data, then doing word embedding using word2vec, then classifying using DL and ML, and finally getting the best accuracy results. All analyzes were performed using Python version 3.9.7 and various modules, including "sklearn," "numpy," and "pandas", and all analyzes were completed.

### 2.1. Data Collection

The dataset is taken from the Stanford Amazon review, namely the shoe review dataset, which consists of 389,877 reviews. The dataset is taken from the address <https://snap.stanford.edu/data/web-amazon.html>. The dataset contains a review of shoe products, which consists of 10 attributes

## 2.2. Data Preprocessing

Data preprocessing is a technique used to convert raw data into a meaningful format so users can understand it more readily. Because the shoe review dataset is extensive and very susceptible to noise, missing, incomplete, and inconsistent because the review data comes from many people. So the dataset needs to be cleaned. As for the steps for cleaning the data in this study, the first of the ten attributes in the dataset will only be taken by the review/score and review/text columns because what will be analyzed is sentiment. Then delete duplicate data, then because it analyzes sentiment from the review, it must convert scores into sentiments. The score value of the original dataset is 1-5 because it is converted to sentiment into 0 and 1—value 0 for negative reviews and 1 for positive reviews. A value of 0 for a score  $\leq 3$ , and the rest is worth 1. And it results that there are 64,924 positive reviews and 15,412 negatives. After random merging, the data settlement value becomes 55,580. Then delete the existing punctuation, then tokenize the review into tokens or words. The following process is to remove stop words.

## 2.3. Word Embedding

Word embedding is one method for word representation. Word embedding was developed around 2000. Word embedding maps every word in a document, a dense vector. The vector will represent the projection of the word in the vector space so that the embedding word can capture the semantic and syntactic meaning of the word. One of the word embedding that is popular today is word embedding word2vec. Word2vec was created by Mikolov et al. in 2013, and today it is widely used in NLP research. The way word2vec works is by presenting words in vectors that can carry semantic meaning. Word2vec is an unsupervised learning that uses a neural network consisting of a hidden and fully connected layer. Word2vec relies on language and locale information. The existing semantics are learned from particular words and influenced by those around them. There are two word2vec algorithms, namely Continuous Bag-of-Word (CBOW) and Skip-gram. The CBOW model uses context to predict the target word, while the skip-gram model is a model that uses a word to indicate the target context. The shape of the CBOW model can be seen in Figure 2, and the form of the Skip-gram model can be seen in Figure 3:

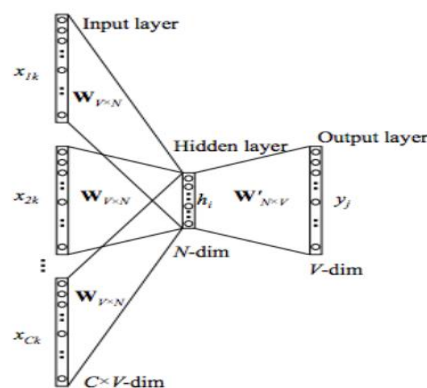


Figure 2. CBOW models

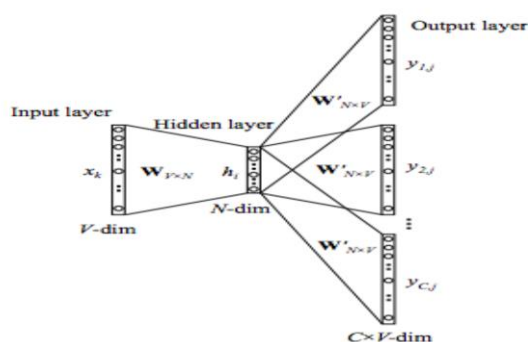
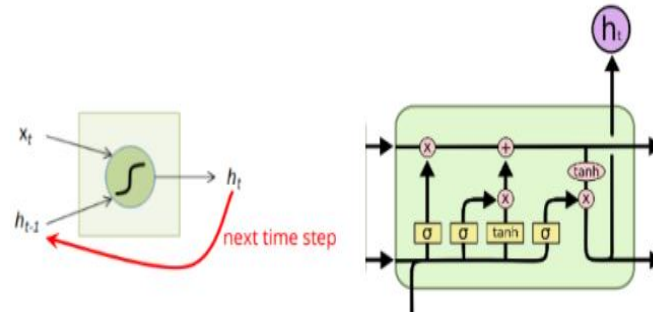


Figure 3. Skip-gram models

## 2.4. Classification Algorithm

### a. Long Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a deep learning method that can model long or short-term temporal dependencies in a time series. Where LSTM is a development of a Recurrent Neural Network (RNN), which can have the ability to remember values from the previous stage for future use purposes [9]. The RNN and LSTM methods have the same input, but the data processing function is different. Where the LSTM model has a more complex function than the RNN. Namely, the LSTM has four components, the forget gate, input gate, output gate, and cell states. The difference in the structure of RNN and LSTM can be seen in Figure 4:



**Figure 4.** Differences in the structure of RNN and LSTM

### b. Convolution Neural Network (CNN)

CNN is a popular deep-learning algorithm. The CNN algorithm is a class of feed-forward neural networks which uses multilayer perceptrons for the required preprocessing. Because CNN is the most critical factor, namely in removing feature extraction that can be trained according to the suitability of the task so that CNN can recognize new objects. The CNN layer extracts meaningful substructures used for prediction tasks [10]. The CNN algorithm must improve because the training process takes a long time.

### c. Support Vector Machine (SVM)

Due to its effectiveness in handling linearly non-separable and sizeable dimensional data sets, SVM has been used extensively in many classification and regression tasks. SVM is an ML-based learning system that uses a hypothetical space as linear functions within a feature with a high dimension. SVM was first introduced by Vapnik in 1992, which has the principle of Structural Risk Minimization (SRM), which aims to find the best hyperplane. SVM divides the dataset into two classes, namely the class separated by a hyperplane, which has a value of 1, and the other has a value of -1 [2].

$$X_i \cdot W + b > 1 \text{ for } Y_i = 1 \quad (1)$$

$$X_i \cdot W + b < -1 \text{ for } Y_i = -1 \quad (2)$$

$X_i$  = With data

$W$  = Support vector weight value that is perpendicular to the hyperplane

$B$  = Biased value

$Y_i$  = 1st data class

### d. Random Forest (RF)

RF is a type of supervised learning that includes the ability to make predictions. RF is a classification that will result in the formation of multiple decision trees and is based on the values of random vectors sampled uniformly and independently. Aggregating trees were used for RF, and the majority would be selected. It also relies on a highly accurate classification that can handle substantial input variables without overfitting [11].

$$\text{Entropy}(S) = \sum_{i=1}^n -P_i \log_2(P_i) = \text{Set of cases} \quad (3)$$

$n$  = Number of partitions  $S$

$P_i$  = Portion of  $S$  to  $S$

## e. Gaussian Naïve Bayes (GNB)

NB, a collection of supervised ML techniques, forecasts membership probability for every class in the data. Conditional probability formula:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \quad (4)$$

Here,  $\mu_y$  and  $\sigma_y$  stand for the predictor distribution's mean and variance.

## 2.5. Evaluation

The author compares the accuracy of all of the currently used approaches and suggests the ML and DL model with the most outstanding performance using comprehensive features. Here is the formula to calculate the accuracy:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \times 100\% \quad (5)$$

TP = True Positive  
 TN = True Negative  
 FP = False Positive  
 FN = False Negative

## 3. Result and Discussion

This section will explain the results of the sentiment analysis research from the shoe review dataset. Where are the evaluation results of each method using ML and DL, namely LSTM, CNN, SVM, RF, and GNB, using word2vec to carry out sentiment classification, which is focused on the accuracy value? The dataset from the review shoe, which consists of 389,877 after pre-processing the data, becomes 55,580. The dataset is split data with a comparison of training and testing of 80:20. The results for the classification of the method LSTM with word embedding word2vec, tuning the parameters, namely batch size = 16, epoch = 5, and use adam optimizer, which can be seen in Table 1:

**Table 1.** LSTM experiment results with word2vec

| Epoch | loss   | acc    | val_loss | val_acc |
|-------|--------|--------|----------|---------|
| 1     | 0.5415 | 0.7201 | 0.4659   | 0.7228  |
| 2     | 0.5057 | 0.7229 | 0.4098   | 0.7229  |
| 3     | 0.4139 | 0.7721 | 0.3713   | 0.8474  |
| 4     | 0.3575 | 0.8343 | 0.3551   | 0.8635  |
| 5     | 0.3409 | 0.8428 | 0.3834   | 0.618   |

The LSTM model with word2vec has been trained five times by calculating accuracy and loss in the training sample, validation accuracy, and validation loss in the test sample. And the training that gives the highest accuracy value is the fifth, which produces an accuracy value of 84.28%, as seen in Table 1. So that the higher the training that can be given, the better the accuracy value will be.

While the results of using CNN with word2vec can be seen in Table 2:

**Table 2.** CNN experiment results with word2vec

| Epoch | loss   | acc    | val_loss | val_acc |
|-------|--------|--------|----------|---------|
| 1     | 0.4489 | 0.7972 | 0.3236   | 0.8577  |
| 2     | 0.3123 | 0.8683 | 0.3189   | 0.8613  |
| 3     | 0.2741 | 0.8873 | 0.332    | 0.8641  |
| 4     | 0.2447 | 0.9008 | 0.3473   | 0.8554  |
| 5     | 0.2148 | 0.9153 | 0.3831   | 0.8549  |

CNN model with word2vec has been trained five times by calculating the accuracy and loss of the training sample and the validation accuracy and loss of the test sample. And the training with the

highest accuracy value is the fifth, which produces an accuracy value of 91.53%. So that the higher the training provided, the better the accuracy value, which can be seen in Table 2.

The classification results using the proposed method, namely by using the word2vec method with the ML method, namely using SVM, RF, and GNB, which can be seen in Table 3:

**Table 3.** Experimental results of the word2vec proposed method with ML

| Classification | accuracy | Recall | F1-Score |
|----------------|----------|--------|----------|
| SVM            | 0.79     | 0.82   | 0.76     |
| RF             | 0.84     | 0.75   | 0.77     |
| GNB            | 0.69     | 0.73   | 0.67     |

#### 4. Conclusion

The research on sentiment analysis of shoe review using the DL method, namely LSTM and CNN with word embedding word2vec, and the ML method, namely SVM, RF, and GNB with word embedding word2vec, has been successfully carried out. The accuracy results using DL with word2vec have an accuracy value above 80% when trained several times, but the CNN method has a higher accuracy value than the LSTM method. So the DL method with word2vec is better than using ML. However, the ML method, namely RF with word2vec, has a reasonably high accuracy value, reaching an accuracy value of 84%. From the discussion above, to carry out sentiment analysis, a more suitable method is to use the CNN method with the word embedding word2vec, obtaining an accuracy value of 91.53%. Still, the dataset used in this study differs from previous research. Future research can be carried out by adding other DL and ML methods and using feature selection.

#### References

- [1] Z. Zhang, T. Jasaitis, R. Freeman, R. Alfrjani, A. Funk, and R. Court, "Mining Healthcare Procurement Data Using Text Mining and Natural Language Processing -- Reflection From An Industrial Project. (arXiv:2301.03458v1 [cs.CL])," *arXiv Comput. Sci.*, 2022.
- [2] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Inf.*, vol. 10, no. 4, pp. 1–68, 2019, doi: 10.3390/info10040150.
- [3] J. Wang, "A hybrid machine learning model for sales prediction," *Proc. - 2020 Int. Conf. Intell. Comput. Human-Computer Interact. ICHCI 2020*, pp. 363–366, 2020, doi: 10.1109/ICHCI51889.2020.00083.
- [4] A. Nurdin, B. Anggo Seno Aji, A. Bustamin, and Z. Abidin, "Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks," *J. Tekno Kompak*, vol. 14, no. 2, p. 74, 2020, doi: 10.33365/jtk.v14i2.732.
- [5] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 728–735, 2021, doi: 10.1016/j.procs.2021.01.061.
- [6] Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, and Mounika Bagadi, Mounika Belusonti, "Text based Sentiment Analysis using LSTM," *Int. J. Eng. Res.*, vol. V9, no. 05, pp. 299–303, 2020, doi: 10.17577/ijertv9is050290.
- [7] A. Hassan and A. Mahmood, "Deep Learning approach for sentiment analysis of short texts," *2017 3rd Int. Conf. Control. Autom. Robot. ICCAR 2017*, pp. 705–710, 2017, doi: 10.1109/ICCAR.2017.7942788.
- [8] A. Chachra, P. Mehndiratta, and M. Gupta, "Sentiment analysis of text using deep convolution neural networks," *2017 10th Int. Conf. Contemp. Comput. IC3 2017*, vol. 2018-Janua, no. August, pp. 1–6, 2018, doi: 10.1109/IC3.2017.8284327.
- [9] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 1394–1401, 2019, doi: 10.1109/ICMLA.2018.00227.
- [10] A. K. Sharma, S. Chaurasia, and D. K. Srivastava, "Sentimental Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 1139–1147, 2020, doi: 10.1016/j.procs.2020.03.416.

- [11] S. Getrudis, I. Sadipun, I. G. Ngurah, A. Cahyadi, and M. Cs, "Analisis Algoritma Random Forest Dalam Memprediksi Penyakit Jantung Koroner," vol. 11, no. 4, pp. 757–764, 2023.