Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Stock Price Correlation Analysis with Twitter Sentiment Analysis Using The CNN-LSTM Method

Muhammad Noer Ibnu Sina[1)*], Erwin Budi Setiawan[2)]
[1,2)]School of Computing, Study Program of Informatics, Telkom University, Bandung, Indonesia
[1)]mnoeribnu@student.telkomuniversity.ac.id, [2)]erwinbudisetiawan@telkomuniversity.ac.id

Submitted : Aug 7, 2023 | **Accepted** : Aug 9, 2023 | **Published** : Oct 1, 2023

Abstract: The intricate interplay between stock prices, reflecting a company's intrinsic value, and multifaceted factors like economic conditions, corporate performance, and market sentiment, constitutes a vital research domain. Grounded in sentiment analysis, our study deciphers public opinions from vast textual data to gauge sentiment, leveraging Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models. We focus on Bank Central Asia (BBCA), a prominent Indonesian banking institution, aiming to forecast stock price fluctuations by analyzing sentiment trends extracted from social media, especially Twitter. Meticulous experimentation, encompassing data segmentation, feature extraction, augmentation, and model refinement, yields significant enhancements in prediction accuracy. The CNN-LSTM model's performance improves from 73.41% to a robust 77.75% accuracy, with F1-scores rising from 73.00% to 75.42%. Importantly, strong correlations emerge between sentiment predictions and actual stock price movements, validated by Spearman correlation coefficients. Positive sentiment exhibits a substantial correlation of 0.745 with stock price changes, while negative sentiment exerts notable influence with a correlation coefficient of 0.691. In summary, our study advances the field of sentiment-driven stock price prediction, showcasing deep learning's effectiveness in extracting sentiment from social media narratives. The implications extend to understanding market dynamics and potentially integrating sentiment-aware strategies into financial decision-making. Future research directions could explore model transferability across financial contexts, real-time sentiment data integration, and interpretability techniques for enhanced practicality in sentiment-driven predictions.

Keywords: BBCA; CNN-LSTM; sentiment analysis; stocks; twitter

INTRODUCTION

Sentiment is a person's discernible opinion on a particular subject (Medhat, Hassan, & Korashy, 2014). It is possible to obtain public opinion or viewpoints by analyzing sentiment. Sentiment can be gleaned from a variety of sources, including social media comments (Hussein, 2018). Companies can use sentiments to determine whether their decisions, policies, or performance merit the term "good" (Cambria, Schuller, Xia, & Havasi, 2013). Twitter is a social media platform that enables users to voice their opinions on a variety of topics. According to (Medhat et al., 2014), Indonesia ranks sixth in terms of the number of Twitter users per country(Iftikar & Muyassar, 2022). This demonstrates that Indonesian Twitter users play a significant role in terms of engagement. Twitter in Indonesia is frequently used to communicate information and express opinions about company-related issues and stock.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Predicting the movement of stock prices is difficult due to the volatility of the stock market (Hussein, 2018). In the current economic climate, investing in equities is a common practice. Due to the influence of supply and demand, the value of shares fluctuates, and share prices can experience both increases and decreases. If demand is high, the price will increase, whereas if supply is high, the price will decrease. Macroeconomic fundamentals, fluctuations in the Rupiah exchange rate against foreign currencies, government policies, panic factors, and market manipulation factors are examples of external factors. The company's fundamentals, corporate actions, and performance are examples of internal factors. There is a correlation between Twitter sentiment and a company's stock price, according to research by (Sul, Dennis, & Yuan, 2014) and (Spencer, Santanen, & Ellis, 2014).

According to the Efficient Market Hypothesis (EMH), all available market information can affect stock prices (Fama, 1970). According to a study by (Sul et al., 2014), there is a correlation between the tone of tweets about a company and its stock price. Several researchers have conducted research on sentiment analysis, one study delved into this by analyzing data from Facebook comment sections. The collected information was categorized and subjected to regression analysis to examine the potential connection between sentiment expressed on Facebook and fluctuations in stock prices. The findings of the study demonstrated that a significant correlation exists only between optimistic sentiment and changes in stock prices (Pradana, Nurcahyo, & Saputro, 2020). Another study by (Gandhi, Malarvizhi Kumar, Chandra Babu, & Karthick, 2021) used the CNN method for sentiment analysis. The utilized dataset was derived from Twitter, particularly product evaluations. This study determined that the CNN method for text classification obtained an F-measure of 91.6% in the skip-gram scheme and an F-measure of 86.6% in the CBOW scheme. In a separate investigation, (Miedema, 2018) applied LSTM to a dataset of 50,000 instances. The reviews were categorized as either positive or negative, and the resulting model obtained an accuracy rating of 86.75%.

In this study, correlation analysis of stock prices utilizing the Spearman method and sentiment analysis utilizing the CNN-LSTM method are compared. Classes are separated into three categories: positive, negative, and neutral. In addition, the greatest result on the Chinese tourism review dataset was achieved by a lexicon-integrated two-channel CNN–LSTM model. Extensive experiments demonstrated that information from a sentiment lexicon and a parallel two-channel model significantly improved SA accuracy (Li, Zhu, Shi, Guo, & Cambria, 2020). Frequency-inverse document frequency (also known as TF-IDF) is a well-known method for determining the significance of a word in a document. Comparing TF-IDF and N-Gram, we can conclude that TF-IDF is the superior choice of features if a machine learning algorithm is to be used for text classification. When incorporating word2vec as a feature expansion, accuracy improves the most as the algorithm learns from the statistics of the number of times each pairing occurs (Ahuja, Chug, Kohli, Gupta, & Ahuja, 2019). Fine-tuning the model's hyperparameters enhances the model's performance on a validation set. In the context of machine learning, a hyperparameter is a parameter whose value is determined prior to beginning the learning process. In contrast, the values of model parameters are derived through data training. Model parameters are the weights and coefficients that the algorithm derives from the data(Nugroho & Setiawan, 2021). This method decreases sensitivity to noise or outliers, which can be measured in the ranks of data instead of actual values (Elgeldawi, Sayed, Galal, & Zaki, 2021).

LITERATURE REVIEW

Sentiment analysis is one of the disciplines in the field of informatics that examines the processing of language in order to analyze a person's opinion, attitude, or evaluation regarding a specific topic. organization, subject, or particular product (Liu, 2010). CNN is a component of the Artificial Neural Network (ANN) that can accurately detect information (Rhanoui, Mikram, Yousfi, & Barzali, 2019). Researchers are currently enhancing CNN's capabilities in the field of NLP, including sentiment analysis (Luo, 2019). CNN is a fully connected layer in which the output of one layer serves as the input for the subsequent layer, typically comprising the input layer, the hidden layer, and the output layer (Rhanoui et al., 2019). Using Twitter data sets about the Indonesian government, the CNN method produced an accuracy of 95.56% in the study(Putra & Setiawan, 2023).

The LSTM (Long Short-Term Memory) method has a high degree of accuracy for text data, much like its predecessor, RNN, which has a distinct advantage when processing relatively lengthy data (long-

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12855 p-ISSN : 2541-044X

term dependency) (Hassan & Mahmood, 2018) and (Wang, Jiang, & Luo, 2016). As seen in the study (Azahra & Setiawan, 2023), sentiment analysis was conducted using political, social, and economic datasets. Using the LSTM procedure resulted in 88.97% accuracy with manual labeling and 97.80% accuracy with system labeling.

Comparing the CNN-LSTM hybrid to other methods such as CNN, LSTM, SVM, and NB-SVM, the accuracy of the CNN-LSTM hybrid on the IMDB dataset was 91% (Rehman, Malik, Raza, & Ali, 2019). In the investigation (Putri, Setiawan, & Sibaroni, 2023), using 1000 maximum feature values for TF-IDF on the aspect of acting yielded an accuracy of 94.80% and an F1-score of 94.74%. Using the SVM algorithm, the utilization of TF-IDF as an extraction feature and word2vec as an expansion feature yielded an accuracy of 68.56% in the study(Nugroho & Setiawan, 2021). One of the test scenarios utilized in the investigation (Alhakiem & Setiawan, 2022) was hyperparameter tuning. In addition to hyperparameter optimization, SMOTE and Feature Expansion were employed to achieve a signal F1-score of 96.48%.

METHOD

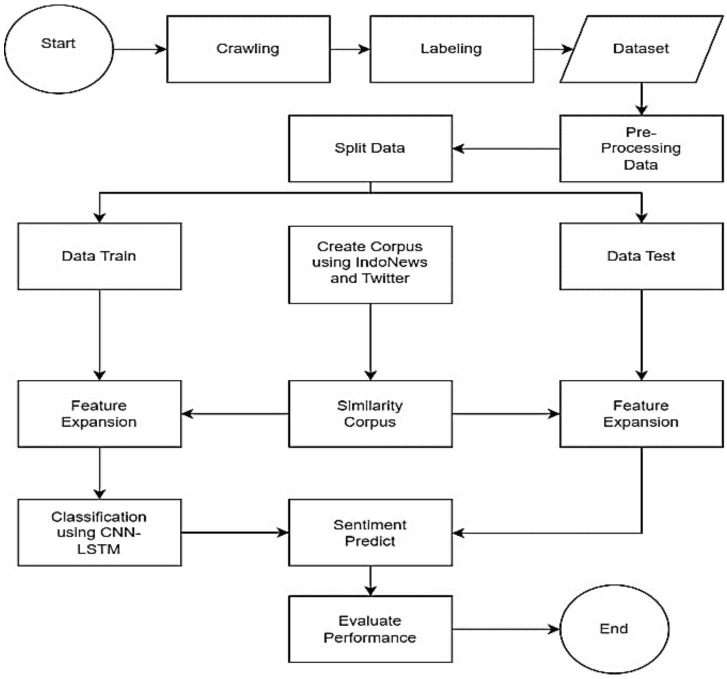

Figure 1. CNN-LSTM Design System for Analysis Sentiment

In Figure 1., the system begins by crawling data from Twitter containing the word 'BBCA', after which the data are manually labeled and divided into three classes: positive, negative, and neutral. After labeling is complete, proceed to the preprocessing phase. After the preprocessing phase is complete, the data is divided into training and test data. Data trains enter the process of Expansion Features, while data assessment enters the phase of Extension Features. The completed data train from the Extraction Feature stage enters the model classification and is then predicted using the Expansion Feature results. The evaluation of performance is conducted using the disorientation matrix.

Dataset

The dataset used in this research was obtained by crawling data on Twitter social media and obtaining 29,445 data tweets from January 2019 to February 2023. The data obtained is stored in a file with the CSV file format. At this stage labelling is carried out on the data that has been collected previously. Labelling is done manually using three sentiment classes, namely positive (2), negative (0), and neutral (1).

Preprocessing

*name of corresponding author

DOI : https://doi.org/10.33395/sinkron.v8i4.12855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

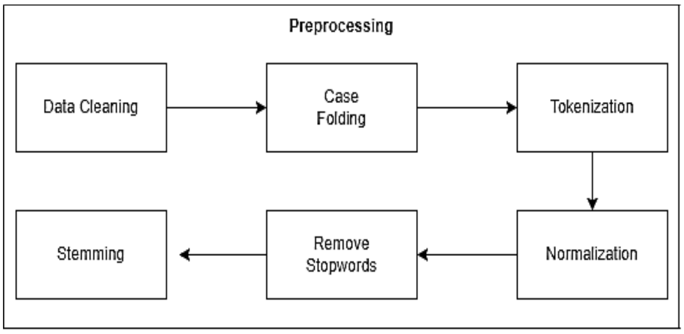

Figure 2. Preprocessing Stage

This pre-processing stage is the stage where the raw data obtained from crawling data is processed with the aim that the data can be used efficiently in the classification stage. In carrying out sentiment analysis it is very important to preprocess the data first to improve model performance in classifying (Alam & Yao, 2019). As shown in Figure 2. the first step is to the data cleaning section, we clean up sentences that contain elements that do not affect the sentiment analysis process. Such as emojis, symbols, mentions and URLs. Case folding section converts all the letters to lowercase. Tokenization is the process of separating sentences into words is carried out without being separated by spaces. Normalization section is changing non-standard words such as slang words, short words will be converted into standard words. Remove Stopwords is done by deletion of words in sentences such as conjunctions or words that do not have an important meaning. Stemming is mapping or decomposing word forms into basic words. At this stage the removal of prefixes, suffixes, or repeated words.

Feature Extraction

Term Frequency-Inverse Document Frequency (TF-IDF) is a method that can weight words by assessing the relevance of these words to the contents of the entire document. The word weighting is obtained by using the following equation:

$$idf_t = \log\left(\frac{D}{df_t}\right) \quad (1)$$

$$w_{dt} = tf_{d,t} \times idf_{d,t} \quad (2)$$

The equation incorporates several variables. 't' represents the frequency of a particular term in the document, and 'W' corresponds to the weight assigned to the document 'd' for that term. 'tf' indicates the total number of terms present in the document, while 'idf' represents the Inverse Document Frequency. When the word appears in a document, it can indicate to you how important that word is. The weighting of documents and words shall be high if the frequency of words in a document is high, while the frequency of words in all documents is low (Wahyuni, Prastiyanto, & Supraptono, 2017).

Feature Expansion

Feature expansion is a process that can expand the value of the features obtained in words that have a weight value from the TF-IDF process. Word2vec is a deep learning tool that uses the simplistic essence of neural networks (Zhang, Wang, Yu, & Wang, 2018). Word2vec can do word mapping into vectors or usually called Word Embedding (Faisal, 2019). Features that do not have a weight or value of 0 will be expanded using word2vec, where word2vec will give a value to words that have a value of 0 according to the similarity value that has been built using corpus similarity. This process will be carried out on all words contained in the document until they no longer have a value of 0 in their weighting value. Word2vec in this study is used as a similarity builder which contains words with the similarity of each word which is used as input to word2vec. The corpus used is 17,172 words which are the source of the data coming from the news media in Indonesia.

CNN

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Convolutional neural network (CNN) is a deep learning type of artificial neural network which is generally used for computer vision or image classification. But CNN also has a good performance on text classification (Jacovi, Shalom, & Goldberg, 2018). The CNN model for text classification uses word vectors as input from word embedding (Santos, Nedjah, & de Macedo Mourelle, 2017).

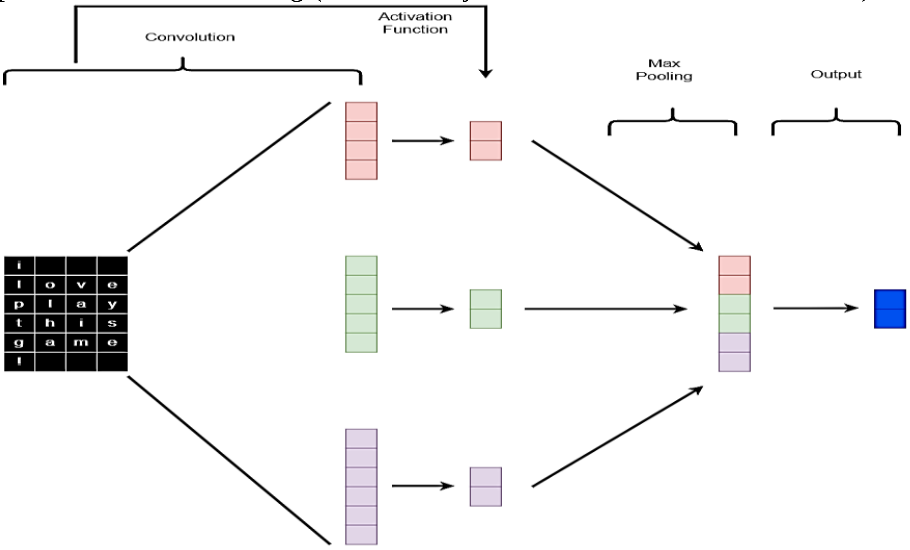

Figure 3. CNN Architecture (Santos et al., 2017)

CNN models generally use convolutional, pooling, and fully connected layers. In this case, namely text classification, generally in the development of the CNN model an embedding layer is used to convert the appropriate input text into a vector (Yadav & Vishwakarma, 2020).

LSTM

Long Short-Term Memory (LSTM) is an enhanced version of the Recurrent Neural Network (RNN) model. While RNN has a limited amount of memory capacity, it is difficult to efficiently process data which has long term dependencies. LSTM can overcome this data by using loops to maintain relevance and not break linkages even with very long sequences (Seo, Kim, Kim, Mo, & Kang, 2020). The LSTM consists of several components, which include the forget gate, input gate, memory module and output gate. The output value of a hidden layer is calculated using these elements.

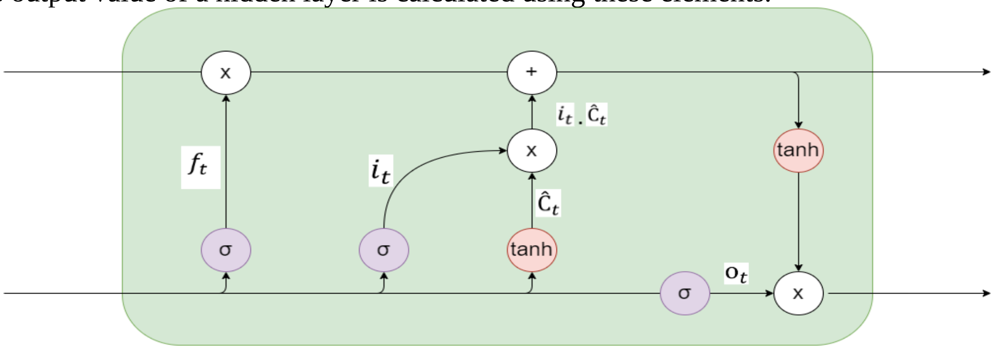

Figure 4. LSTM Architecture (Seo et al., 2020)

The information entered in the data will be analyzed and evaluated to establish if it should be retained or disposed of into memory cells, while performing the forget gate operation. In order to effectively manage the network's weight, this mechanism plays an important role. As a result, if the data are considered important for storage and 0 is indicated that it should not be taken into consideration, 1 shall be assigned.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The determination of which values to update shall be made by the input gates. After this decision, the new value of a vector will be computed in Tanh and stored in memory cells.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{4}$$

$$\mathcal{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{5}$$

A memory cell will perform the task of replacing the previous data stored in your memory with its new value. Combining the outputs from the forget and input gates produces this new value.

$$c_t = f_t \times c_{t-1} + i_t \times \mathcal{C}_t \tag{6}$$

The output gate shall decide whether the memory cell value should be used, at the last step. The output gate performs a transformation using the Tanh function to adjust the memory cell value when determining whether that value is relevant.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \tanh(c_t) \tag{8}$$

CNN-LSTM

CNN LSTM is a hybrid approach that combines Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. Using CNNs to extract relevant features from the input data and using LSTM for predictions of sequential information, it is leveraging its capabilities (Fadilah, Kusuma, Minarno, & Munarko, 2021).

In terms of image or video inputs, CNN-LSTM is commonly used to solve prediction problems. However, its effectiveness in the task of text-based data classification was also demonstrated. Embedding layers, convolutions, pooling layers, fully connected layers, dense layers and LSTM traces are the key architectural elements involved in CNN-LSTM.

Performance Evaluation

In the performance evaluation stage, it is usually calculated using a confusion matrix. This Confusion Matrix can describe the amount of test data in the form of actual values and predicted values (Normawati & Prayogi, 2021). The Confusion Matrix is generally shaped like Table 1(Novitasari & Purbolaksono, 2021), (Indriani, 2014).

Table 1. Confusion Matrix

Confusion Matrix		Predict	
		Positive	Negative
Actual Value	Positive	TP	FN
	Negative	FP	TN

The following terms are included in the table: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). A positive prediction that corresponds to the actual positive value is referred to as the TP. FP represents a positive prediction despite the actual value being negative. FN signifies a negative prediction when the actual value is positive. Lastly, TN is in line with the negative prediction, which is in line with the actual negative value.

Accuracy is a ratio that gives a correctly predicted value compared to all data.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{9}$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Precision measures the proportion of those values that have been correctly predicted to be positive in terms of all data points which are expected to be positive.

$$Precision = \frac{\mathrm{TP}}{TP+FP} \tag{10}$$

Recall is the ratio value of data that is predicted to be positive compared to all actual data with positive values.

$$Recall = \frac{\mathrm{TP}}{TP+FN} \tag{11}$$

The F1-Score provides a balanced evaluation by considering both precision and recall simultaneously.

Spearman Correlation Coefficients

Spearman Rank correlation is a statistical method used for assessing the correlation between two independent variables measured on an ordinal scale (Gauthier, 2001). Unlike parametric methods, the Spearman Rank correlation test is nonparametric, which means that the distribution of data is not affected. This method decreases sensitivity to noise or outliers, which can be measured in the ranks of data instead of actual values (Schober & Schwarte, 2018). The strength and direction of correlation are quantified by the Spearman Rank Correlation coefficient, described in the formula (Zar, 2005).

$$\boldsymbol{r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}} \tag{13}$$

The symbol "rs" represents the correlation coefficient itself in the Spearman Rank coefficient. In the coefficient calculation "d" means a difference in ranks for each pair, and "n" represents the total number of pairs to be analyzed. In addition, the degree of correlation can be divided into the following categories as set out in Spearman correlation Table.

Table 1. Spearman Rank Correlation

Value Correlation	Correlation
$\rho = 0$	No Correlation
$0 < \lvert\rho\rvert \leq 0.19$	Very Weak Correlation
$0.20 \leq \lvert\rho\rvert \leq 0.39$	Weak Correlation
$0.40 \leq \lvert\rho\rvert \leq 0.59$	Moderate Correlation
$0.60 \leq \lvert\rho\rvert \leq 0.79$	Strong Correlation
$0.80 \leq \lvert\rho\rvert \leq 1.00$	Very Strong Correlation
1.00	Monotonic Correlation

RESULT

In this study, we conducted four testing scenarios to assess our model's performance. The first scenario involved evaluating the baseline through data split ratio application, examining model performance during data division into training and testing sets. In the second scenario, we enhanced the baseline by incorporating TF-IDF feature extraction, aiming to assess its impact on classification performance and word importance. The third scenario improved upon the baseline further by combining TF-IDF with Word2vec expansion, utilizing Word2vec's ability to capture contextual relationships among words. The fourth scenario aimed to optimize the CNN-LSTM architecture using hyperparameter tuning techniques. We sought performance enhancement by adjusting hyperparameters and have been experimenting with various combinations.

*name of corresponding author

Each scenario employed CNN-LSTM text classification models, leveraging Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for sequential data processing, particularly in text analysis. To ensure robust findings, we executed each scenario four times and calculated average accuracy and F1-score, providing comprehensive insights into model performance. Accuracy measured correct classification proportion, while the F1-score balanced precision and recall effectively.

Scenario 1

In this study, the primary objective was to determine the optimal data division for testing purposes. Different data splitting strategies were applied, such as a 90% training and 10% testing ratio, 80% training and 20% testing ratio, and so forth. The analysis involved an examination of a table displaying the performance of three distinct models: CNN-LSTM, CNN, and LSTM. The focus was on evaluating the accuracy of these models' predictions.

Table 2. Splitting Data Result

Test Size	CNN-LSTM		CNN		LSTM	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
90:10	**74.53**	**74.29**	**74.29**	**73.88**	**72.05**	**70.64**
80:20	72.55	72.66	72.67	72.28	69.40	67.94
70:30	71.88	71.87	71.81	71.37	69.53	68.96

Under the 90:10 data split, the CNN-LSTM model achieved an accuracy of approximately 74.53%, while the CNN and LSTM models achieved accuracies of about 74.29% and 72.05%, respectively. With an 80:20 split, both the CNN-LSTM and CNN models demonstrated comparable performance, achieving accuracies of around 72%. The LSTM model achieved a slightly lower accuracy of approximately 69.40%. Transitioning to a 70:30 split, the CNN-LSTM model attained an accuracy of about 71%, while the CNN and LSTM models achieved accuracies of around 71% and 69%, respectively.

Our findings showed that the best way to split the data for testing depended on the model we were using. It was clear that finding the right balance between training and testing data was important for how well our models could predict things. This information gave us a good starting point for the rest of our research.

Scenario 2

In the second scenario of our study, we introduced Term Frequency-Inverse Document Frequency (TF-IDF) as an extraction parameter to enhance the baseline model's capabilities in analyzing textual content. Our aim was to find the optimal number of features for maximum performance. We compared the model's performance with different maximum feature values: 500, 1.000, 5.000, 10.000, 15.000, 16.000, and 20.000, along with a baseline value matching the training dataset size from scenario 1. The results, summarized in Table 4, demonstrate the model's performance under varying feature counts.

Table 3. TF-IDF Result

Max Feature	CNN-LSTM		CNN		LSTM	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
500	72.86	72.56	73.44	73.03	72.42	71.88
1000	**73.41**	**73.00**	73.00	72.78	73.37	72.88
5000	72.73	72.47	74.56	74.29	**74.09**	**73.71**
10000	71.54	71.33	75.11	74.68	73.17	72.87
15000	71.64	71.55	74.77	74.41	72.56	72.28
16000	72.05	71.96	**75.17**	**74.74**	73.78	73.36
20000	71.30	71.18	74.80	74.54	73.00	72.69
26500	70.89	70.59	74.83	74.46	72.73	72.40

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Among these models, the feature count of 1.000 proved to be the most conducive to the CNN-LSTM model, resulting in an accuracy of 73.41%. For the CNN model, an optimum accuracy of 75.17% was attained with 16.000 features. Meanwhile, the LSTM model demonstrated its highest accuracy of 74.09% when utilizing 5.000 features.

Scenario 3

In this third scenario, we introduce Word2Vec as an expansion feature into the model from the previous scenario. Our experiment focuses on corpus similarity, using Indonesian news and tweet corpus. We rank words from each corpus: top 1, top 2, top 3, top 5, top 10, and top 20. The outcomes of this scenario are summarized in Table 5.

Table 4. Feature Expansion Result

Top (n) word	CNN-LSTM		CNN		LSTM	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
1	**75.34**	**75.13**	**77.79**	**77.44**	76.77	76.30
2	74.32	73.97	76.63	76.14	76.57	76.15
3	75.28	74.87	76.36	76.00	75.78	75.30
5	75.07	74.69	77.24	76.91	76.94	76.67
10	74.60	74.33	77.35	77.08	**77.31**	**77.15**

Remarkably, the results revealed intriguing dynamics among these models. The CNN-LSTM model achieved its highest accuracy when utilizing a singular top word, showcasing an impressive accuracy of 75.34%. Similarly, the CNN model demonstrated its peak accuracy, reaching 77.79%, also with just one top word. Intriguingly, the LSTM model's optimal accuracy was achieved with a broader scope, incorporating the top 10 words, resulting in a commendable accuracy of 77.31%.

Scenario 4

In the fourth phase of our study, we aimed to make our model even better by adjusting its settings. Our researcher tweaked important factors, like the hidden layers (128 to 256) and batch size (32 to 64), hoping to improve accuracy and the F1-score. The Table 6. below summarizes the results from this hyperparameter tuning, comparing it with our previous experiments: Split data, Max Feature, and Feature Expansion. We tested three model variants: CNN-LSTM, CNN, and LSTM.

Table 5. Hyperparameter Tuning

Scenario	CNN-LSTM		CNN		LSTM	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
Split data	**74.53**	**74.29**	**74.29**	**73.88**	**72.05**	**70.64**
Max Feature	**73.41**	**73.00**	**75.17**	**74.74**	**74.09**	**73.71**
Feature Expansion	**75.34**	**75.13**	**77.79**	**77.44**	**77.31**	**77.15**
Hyperparameter Tuning	**75.75**	**75.42**	**77.82**	**77.49**	**77.45**	**77.15**

DISCUSSIONS

In our research, we explored combinations of CNN and LSTM models. When used together, the CNN-LSTM combination achieved varying levels of accuracy and F1-Scores across scenarios, including split data (74.53% accuracy, 74.29% F1-Score), max feature adjustment (73.41% accuracy, 73.00% F1-Score), feature expansion (75.34% accuracy, 75.13% F1-Score), and hyperparameter tuning (75.75% accuracy, 75.42% F1-Score). Notably, the CNN model demonstrated improved performance, surpassing the combined approach in some cases, particularly during feature expansion and hyperparameter tuning. These findings highlight the impact of feature manipulation and hyperparameter optimization on the predictive capabilities of CNN and LSTM models, offering insights for effective model configuration.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12855
e-ISSN : 2541-2019
p-ISSN : 2541-044X

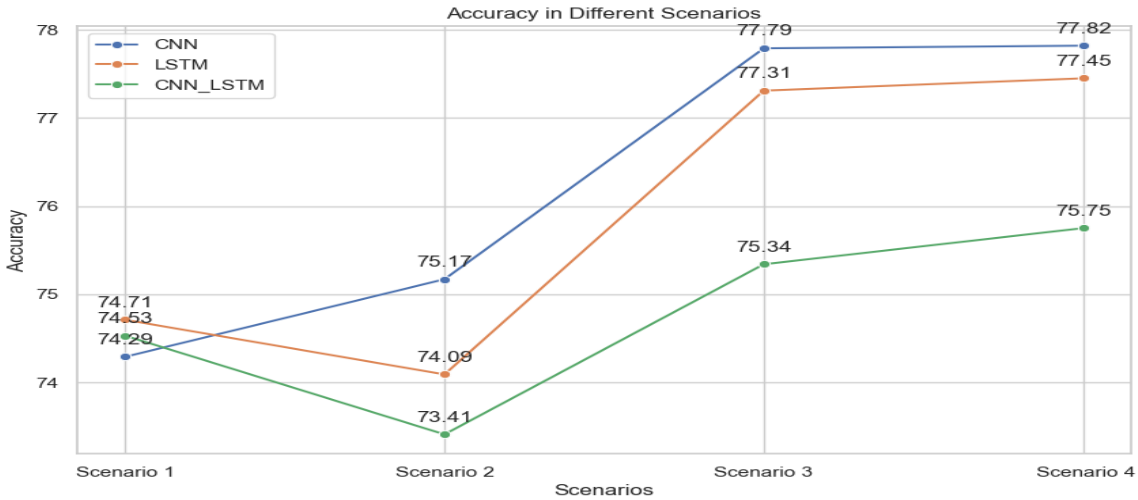

Figure 5. Accuracy from Each Scenario

Then we're going to perform a correlation analysis with Spearman Correlation. The sentiment anticipated by the model is taken into consideration to determine the values that will be evaluated. A graph showing the positive sentiment towards the BCA bank's stock price is shown in Figure 6.

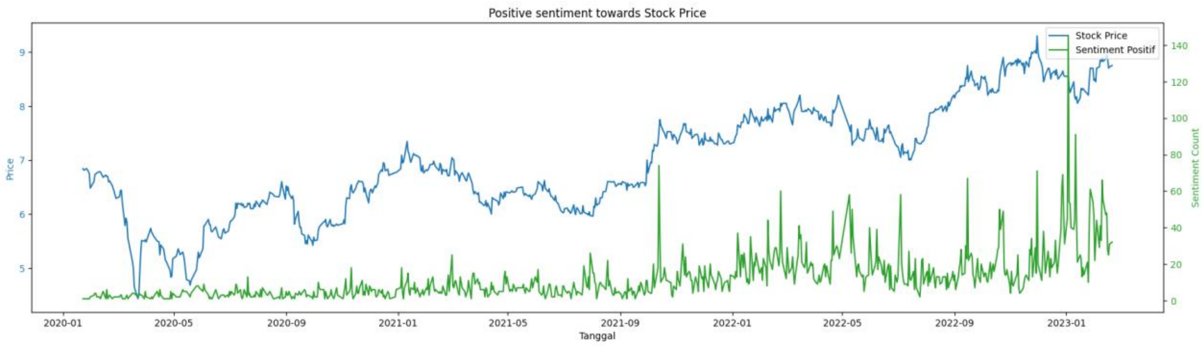

Figure 6. Visualization of Positive Sentiment on Stock Prices

Positive sentiment predictions regarding BBCA stock price, which have been tested by Spearman correlation testing, are shown in Figure 7. The test yielded a correlation value of 0.745, placing it within the Strong Correlation category based on the Spearman correlation table's range of 0.60 - 0.79.

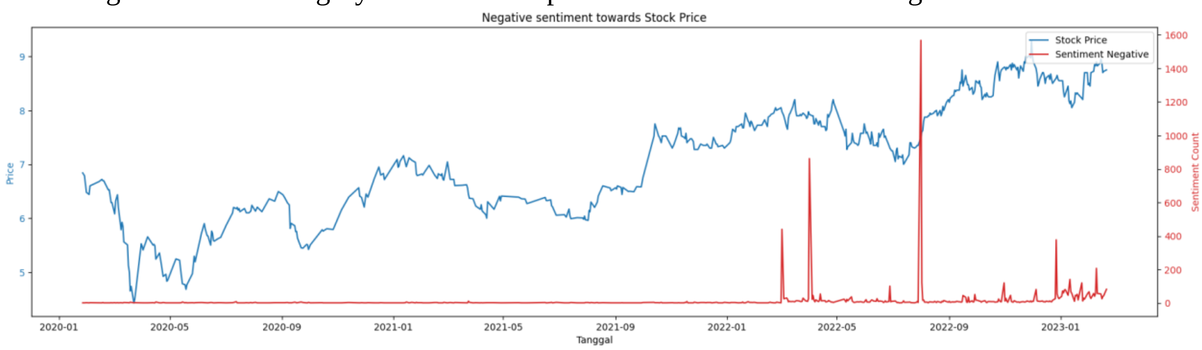

Figure 7. Visualization of Negative Sentiment on Stock Prices

The results of the prediction of negative sentiment based upon this model are shown in Figure 6, And Figure 7. Subsequently, a Spearman correlation test was carried out, resulting in a correlation coefficient of 0.691. According to a correlation table, which has an interval between 0.60 and 0.79, this value is included in the strong correlation category.

*name of corresponding author

CONCLUSION

This comprehensive study aimed to enhance the performance of a CNN-LSTM text classification model through various testing scenarios. These scenarios encompassed data split ratios, TF-IDF feature extraction, Word2Vec expansion, and hyperparameter tuning. When used together, the CNN-LSTM combination achieved varying levels of accuracy and F1-Scores across scenarios, including split data (74.53% accuracy, 74.29% F1-Score), max feature adjustment (73.41% accuracy, 73.00% F1-Score), feature expansion (75.34% accuracy, 75.13% F1-Score), and hyperparameter tuning (75.75% accuracy, 75.42% F1-Score). Notably, the CNN model demonstrated improved performance, surpassing the combined approach in some cases, particularly during feature expansion and hyperparameter tuning. These findings highlight the impact of feature manipulation and hyperparameter optimization on the predictive capabilities of CNN and LSTM models, offering insights for effective model configuration. The study culminated in a sentiment analysis, where Spearman correlation tests revealed intriguing insights based on the sentiment predictions of the CNN-LSTM model. The positive sentiment predictions regarding BCA bank's stock price exhibited a substantial correlation coefficient of 0.745, indicating a strong positive relationship between the model's sentiment assessment and the actual stock prices, as depicted in Figure 6. Moreover, the model's negative sentiment predictions also demonstrated a notable correlation coefficient of 0.691, reinforcing the strong correlation between the model's sentiment evaluations and corresponding stock price trends, as illustrated in Figure 7. These robust correlations further validate the predictive capabilities of the CNN-LSTM model in capturing underlying sentiment trends, providing valuable tools for stock price prediction. Overall, the findings from this study underscore the significance of feature engineering and hyperparameter tuning in optimizing CNN and LSTM models, offering valuable insights for effective model configuration and application in predicting stock price sentiment, and highlighting the substantial correlation between sentiment predictions and actual stock price movements.

REFERENCES

Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). *The impact of features extraction on the sentiment analysis*. In *Procedia Computer Science* (Vol. 152). doi:10.1016/j.procs.2019.05.008

Alam, S., & Yao, N. (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. *Computational and Mathematical Organization Theory*, *25*(3), 319–335. doi:10.1007/s10588-018-9266-8

Alhakiem, H. R., & Setiawan, E. B. (2022). Aspect-Bas1ed Sentiment Analysis on Twitter Using Logistic Regression with FastText Feature Expansion. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(5), 840–846. doi:10.29207/resti.v6i5.4429

Azahra, N. M., & Setiawan, E. B. (2023). Sentence-Level Granularity Oriented Sentiment Analysis of Social Media Using Long Short-Term Memory (LSTM) and IndoBERTweet Method. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI)*, *9*(1), 85–95. doi:10.26555/jiteki.v9i1.25765

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, *28*(2), 15–21. doi:10.1109/MIS.2013.30

Elgeldawi, E., Sayed, A., Galal, A. R., & Zaki, A. M. (2021). Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. *Informatics*, *8*(4). doi:10.3390/informatics8040079

Fadilah, W. R. U., Kusuma, W. A., Minarno, A. E., & Munarko, Y. (2021). Classification of Human Activity Recognition Utilizing Smartphone Data of CNN-LSTM. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*. doi:10.22219/kinetik.v6i2.1319

Faisal, M. R. (2019). *Ekstraksi Fitur Menggunakan Model Word2vec Untuk Analisis Sentimen Pada Komentar Facebook Effect of features Generated from additional segments in protein sequence classification View project IT Asset Management View project*. Retrieved from https://www.researchgate.net/publication/343057288

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383. doi:10.2307/2325486

*name of corresponding author

Gandhi, U. D., Malarvizhi Kumar, P., Chandra Babu, G., & Karthick, G. (2021). Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). *Wireless Personal Communications*. doi:10.1007/s11277-021-08580-3

Gauthier, T. D. (2001). Detecting trends using Spearman's rank correlation coefficient. *Environmental Forensics*, *2*(4). doi:10.1006/enfo.2001.0061

Hassan, A., & Mahmood, A. (2018). Convolutional Recurrent Deep Learning Model for Sentence Classification. *IEEE Access*, *6*, 13949–13957. doi:10.1109/ACCESS.2018.2814818

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, *30*(4), 330–338. doi:10.1016/j.jksues.2016.04.002

Iftikar, A., & Muyassar. (2022). *Analisis Sentimen Twitter: Penanganan Pandemi Covid-19 Menggunakan Metode Hybrid Naïve Bayes, Decision Tree, dan Support Vector Machine*.

Indriani, A. (2014). *Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier*. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*. Retrieved from www.bluefame.com,

Jacovi, A., Shalom, O. S., & Goldberg, Y. (2018). Understanding Convolutional Neural Networks for Text Classification.

Li, W., Zhu, L., Shi, Y., Guo, K., & Cambria, E. (2020). User reviews: Sentiment analysis using lexicon integrated two-channel CNN–LSTM family models. *Applied Soft Computing Journal*, *94*. doi:10.1016/j.asoc.2020.106435

Liu, B. (2010). Sentiment analysis and subjectivity.

Luo, L. xia. (2019). Network text sentiment analysis method combining LDA text representation and GRU-CNN. *Personal and Ubiquitous Computing*, *23*(3–4), 405–412. doi:10.1007/s00779-018-1183-9

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. doi:10.1016/j.asej.2014.04.011

Miedema, F. (2018). *Sentiment Analysis with Long Short-Term Memory networks*.

Normawati, D., & Prayogi, S. A. (2021). *Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter*. *Jurnal Sains Komputer & Informatika (J-SAKTI* (Vol. 5).

Novitasari, F., & Purbolaksono, M. D. (2021). Sentiment Analysis Aspect Level on Beauty Product Reviews Using Chi-Square and Naïve Bayes. *OPEN ACCESS J DATA SCI APPL*, *4*(1), 18–030. doi:10.34818/JDSA.2021.4.72

Nugroho, N. A., & Setiawan, E. B. (2021). Implementation Word2Vec for Feature Expansion in Twitter Sentiment Analysis. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(5), 837–842. doi:10.29207/resti.v5i5.3325

Pradana, M. G., Nurcahyo, A. C., & Saputro, P. H. (2020). PENGARUH SENTIMEN DI SOSIAL MEDIA DENGAN HARGA SAHAM PERUSAHAAN. *Edutic - Scientific Journal of Informatics Education*, *6*(2). doi:10.21107/edutic.v6i2.6992

Putra, D. T., & Setiawan, E. B. (2023). Sentiment Analysis on Social Media with Glove Using Combination CNN and RoBERTa. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(3), 457–563. doi:10.29207/resti.v7i3.4892

Putri, S. I., Setiawan, E. B., & Sibaroni, Y. (2023). JURNAL MEDIA INFORMATIKA BUDIDARMA Aspect-Based Sentiment Analysis on Twitter Using Long Short-Term Memory Method. doi:10.30865/mib.v7i2.5637

Rehman, A. U., Malik, A. K., Raza, B., & Ali, W. (2019). A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. *Multimedia Tools and Applications*, *78*(18). doi:10.1007/s11042-019-07788-7

Rhanoui, M., Mikram, M., Yousfi, S., & Barzali, S. (2019). A CNN-BiLSTM Model for Document-Level Sentiment Analysis. *Machine Learning and Knowledge Extraction*, *1*(3), 832–847. doi:10.3390/make1030048

Santos, I., Nedjah, N., & de Macedo Mourelle, L. (2017). *Sentiment analysis using convolutional neural network with fastText embeddings*. In *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)* (pp. 1–5). IEEE. doi:10.1109/LA-CCI.2017.8285683

*name of corresponding author

Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, *126*(5). doi:10.1213/ANE.0000000000002864

Seo, S., Kim, C., Kim, H., Mo, K., & Kang, P. (2020). Comparative Study of Deep Learning-Based Sentiment Classification. *IEEE Access*, *8*, 6861–6875. doi:10.1109/ACCESS.2019.2963426

Spencer, D. H., Santanen, E. L., & Ellis, T. J. (2014). *Introduction to Advances in Teaching and Learning Technologies Minitrack*. In *2014 47th Hawaii International Conference on System Sciences* (pp. 2–2). IEEE. doi:10.1109/HICSS.2014.10

Sul, H. K., Dennis, A. R., & Yuan, L. (2014). *Trading on twitter: The financial information content of emotion in social media*. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 806–815). IEEE Computer Society. doi:10.1109/HICSS.2014.107

Wahyuni, R. T., Prastiyanto, D., & Supraptono, D. E. (2017). *Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi*.

Wang, X., Jiang, W., & Luo, Z. (2016). *Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts*. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 2428–2437). Osaka, Japan: The COLING 2016 Organizing Committee. Retrieved from https://aclanthology.org/C16-1229

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, *53*(6), 4335–4385. doi:10.1007/s10462-019-09794-5

Zar, J. H. (2005). Spearman Rank Correlation. In *Encyclopedia of Biostatistics*. Wiley. doi:10.1002/0470011815.b2a15150

Zhang, C., Wang, X., Yu, S., & Wang, Y. (2018). *Research on Keyword Extraction of Word2vec Model in Chinese Corpus*. In *Proceedings - 17th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2018*. doi:10.1109/ICIS.2018.8466534

*name of corresponding author