Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12825
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Sentiment Analysis of Bank BCA Stock Price on Twitter Data Using LSTM and Genetic Algorithm Optimization

Rizki Tri Setiawan [1)*], Erwin Budi Setiawan [2)*]
[1,2)] Fakultas Informatika, Universitas Telkom, Bandung
[1)] Rizkitrisetiawan@students.telkomuniversity.ac.id, [2)] erwinbudisetiawan@telkomuniversity.ac.id

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

Abstract: In today's business world, there is significant development and emergence of various and diverse innovations. Therefore, every company needs to develop itself in various ways, one of which is going public. This involves a company selling a percentage of its value to the public in order to facilitate its growth in every aspect required. However, it is not easy for issuers to attract investors to invest their capital because each investor has different criteria in terms of investment unit value. Essentially, the stock price depends on the strengths and weaknesses of the company. Hence, in order to expand the market and manage customer relationships, information is needed as a decision support. One of the sources of information that can be used is Twitter, which includes positive, neutral, and negative opinions. This study employs the LSTM classification method and word embedding using GloVe, followed by Genetic Algorithm optimization, which is used to predict sentiment in tweets related to the BBCA stock. The model is built with classification using Long Short-Term Memory to determine the level of accuracy produced. Then, the word embedding method using GloVe is used, and the obtained results with the GloVe-LSTM method yield an overall accuracy score of 71%. Furthermore, the optimization method using Genetic Algorithm is applied to enhance the previous method, GloVe-LSTM, resulting in an accuracy of 87% with the best individual values of 111,170, 0.398, 93, etc., and the best fitness score of 0.8724.

Keywords: BBCA; GloVe; Genetic Algorithms; LSTM; Sentiment Analysis.

INTRODUCTION

The business world is currently experiencing rapid development, with many new companies emerging with competitive advantages. This has created intense business competition, requiring every company to develop its business in order to survive and compete with other companies. In order to expand their business, companies often need additional capital, which can be obtained through various means, one of which is by going public.

Of course, it is not easy for issuers to attract investors to invest their capital, as each investor has different criteria in terms of investment unit value. The stock price tends to fluctuate based on supply and demand (Dina P, Dewi A & Suaryana I, 2013). Those with surplus funds can invest them in real investments or financial investments (Made Wahyuliantini, 2015). Investors invest their capital for the future and expect returns in the form of stock returns. Stocks are considered to provide significant returns in a short period compared to other forms of investment (Syafira & Rikumahu, 2020).

Indonesia itself has 79 million social media users, including 17 million Twitter users in 2016 (Syafira & Rikumahu, 2020). To expand the market and manage customer relationships, we need

*name of corresponding author

information as decision support, and one source of information that can be used is social media. Twitter is a major social media platform with millions of users and provides unique insights. The data available on Twitter can help understand market sentiment objectively (Rustiana & Rahayu, 2017), (Hariyanto Wahyu & Maharani, 2020).

Word embedding is a technique for representing words as vectors. This technique can improve sentiment analysis accuracy or performance (Hariyanto Wahyu & Maharani, 2020). This research employs GloVe (Global Vectors) word embedding due to its superior performance in sentiment analysis when compared to other word embedding techniques like Skip-gram, and ect. The classification method used in this study is Long Short-Term Memory (LSTM) because LSTM has multiple layers, including one for word embedding using GloVe, and it has good performance for deep learning and sentiment classification of tweets. The Genetic algorithm is an optimization technique designed to imitate the processes observed in natural evolution (Wati, 2016).

In this study, the classification method used is Long Short-Term Memory (LSTM), with the addition of feature expansion using GloVe, and it will be optimized using genetic algorithm. Based on the scenario described above, the dataset used will be tweets data from Twitter regarding the stock price of BBCA. Further detailed explanations will be discussed in the following sections.

LITERATURE REVIEW

GloVe (Global Vektor)

Global Vectors (GloVe). GloVe is crucial for representing words and transforming them into vector form, which is essential for sentiment analysis (Nurdin et al., 2020). The words obtained during the data collection process may not be sufficient for analysis. Therefore, it is important to represent the statistical results of the data beforehand. GloVe, also known as Global Vector, is one of the models used to represent the vectorized representation of the collected words (Xiaoyan et al., 2022).

LSTM (Long Short – Term Memory)

Long Short-Term Memory (LSTM) is a variant of the Recurrent Neural Network (RNN) that Hochreiter and Schmidhuber introduced in 1997. One of the challenges faced by traditional RNN is its limited ability to retain information during the learning process, especially when dealing with long sequences. This issue can be addressed by utilizing the LSTM method (Nurvania et al., 2021). LSTM addresses the vanishing gradient issue and enables enhanced information retention and learning across extended sequences, making it highly suitable for tasks like sequence prediction, language modeling, and sentiment analysis

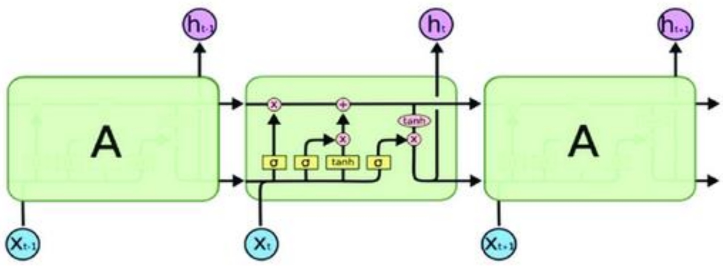

Fig 1. LSTM Gate

LSTM is made out of three doors, to be specific the Info Entryway (Rahman et al., 2021), Neglect Door, and Result Entryway, as represented in the gave picture. Each gate has a specific role in determining whether certain information should be stored or forgotten.

a. Input Gate: The flow of new data into the memory cell is controlled by the Input Gate. It surveys which values ought to be refreshed and added to the cell state in view of the ongoing info.

*name of corresponding author

b. Forget Gate: The Forget Gate plays a crucial role in deciding what information from the previous cell state should be ignored or erased. It takes into account both the current input and the previous hidden state to arrive at this conclusion.

c. Output Gate: The Output Gate regulates the output of the LSTM cell. It determines which information should be outputted, taking into account the current input and the updated cell state.

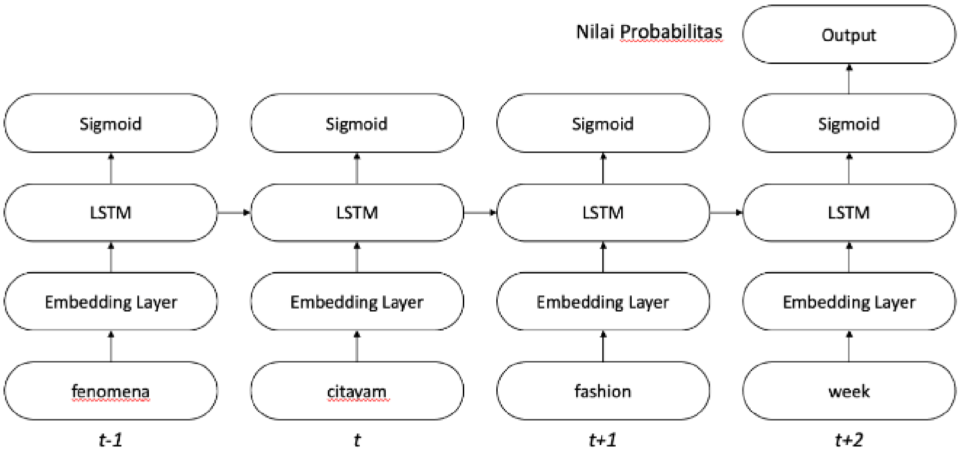

Fig 2. LSTM Layer

And the structure of the LSTM model consists of several interconnected layers based on the time steps present in each input data, as seen in Figure 3 (Farsiah et al., 2022). The LSTM structure can retain information from the previous training process using a control gate called the forget gate. The use of gates in the LSTM structure helps prevent the occurrence of vanishing/exploding gradients (Setiawan Irawati et al., 2020).

Genetic Algorithms

The Genetic Algorithm is an optimization algorithm that emulates specific processes observed in natural evolution. It is designed to solve complex problems by imitating the principles of natural selection, genetic inheritance, and survival of the fittest. By mimicking these processes, the Genetic Algorithm can efficiently explore and exploit the search space to find optimal solutions. It is a potent stochastic algorithm that has been successfully used to solve problems in machine learning and optimization. It is based on the concepts of genetic inheritance and natural selection. (Wiranata & Djunaidy, 2021). Hereditary Calculation, also known as Genetic Algorithm, utilizes techniques such as selection, crossover, and mutation to iteratively evolve a population of potential solutions and search for the optimal or near-optimal solution to a given problem. The capability to traverse the search space and identify globally optimal solutions renders it applicable to a broad spectrum of applications in machine learning and optimization.

Rank Spearman Correlation

Spearman rank correlation is a statistical technique that is used to evaluate the association between two independent variables without making assumptions about their distribution (Nugraha & Setiawan, 2023). Because of Spearman rank relationship being a non-parametric technique, it enjoys the benefit of not being impacted by the information conveyance, in this way diminishing aversion to commotion or exception information (Schober & Schwarte, 2018). Formula (1) shows the Spearman rank correlation.

$$rs = 1 - \frac{6\sum d_i^2}{n^3 - n^1} \quad (1)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12825
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In formula (1), it represents the formula for examining the relationship between sentiment and the movement of BBCA stock prices. Rank Spearman Correlation indicates the correlation coefficient that will be generated based on Table 2.

Table 2. Rank Spearman Correlation Interval

Interval	Strength
0.00 – 0.009	Relationship Ignored
0.10 – 0.39	Weak Relationship
0.40 – 0.69	Medium Relationship
0.70 – 0.89	Strong Relationship
0.90 – 1.00	Very Strong Relationship

Evaluation

In order to assess the constructed model's efficacy, evaluation metrics are used to measure the outcomes of classification tasks. The generated models' performance and quality are also evaluated using these metrics (Lee & Sibaroni, 2023). The assessment measurements utilized in this study followed.

a. The confusion matrix is a method for determining how well the classification model works by looking at how accurate its predictions are. It provides a tabular representation that summarizes the model's predictions and their alignment with the actual class labels (Lee & Sibaroni, 2023). Table 3 displays four conditions for testing the confusion matrix's accuracy.

Table 3. Confusion Matrix

Classificarion	**Positive Prediction**	**Negative Prediction**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

b. Taking into account the entire dataset, accuracy can be defined as the proportion of correct predictions made for both positive and negative values. It is calculated by dividing the number of accurate predictions by the total number of predictions. This condition considers assessing the precision of a model's predictions for all relevant data components. Equation (2) displays the formula used to determine accuracy.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (2)$$

c. Precision is the proportion of true positive predictions relative to the total number of positive outcomes. It measures the correctness of positive predictions made by a model. By calculating the precision, we can assess how well the model performs in correctly identifying positive instances compared to all instances it predicts as positive. Equation (3) displays the formula used to determine accuracy.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

d. Recall, commonly referred to as sensitivity or true positive rate, signifies the proportion of true positive predictions compared to the overall number of actual positive instances present in the dataset. It estimates the capacity of a model to accurately recognize positive occurrences out of all examples that are genuinely certain. By calculating recall, we can assess how well the model captures and recalls positive instances in the dataset. Equation (4) displays the formula used to determine accuracy.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12825
e-ISSN : 2541-2019
p-ISSN : 2541-044X

e. The F1-score is determined by computing the average of precision and recall to evaluate the performance of a model. Equation (5) displays the formula used to determine accuracy.

$$F1 - Score = \frac{2*Recall*Precision}{(Recall+Precision)} \quad (5)$$

METHOD

System Design

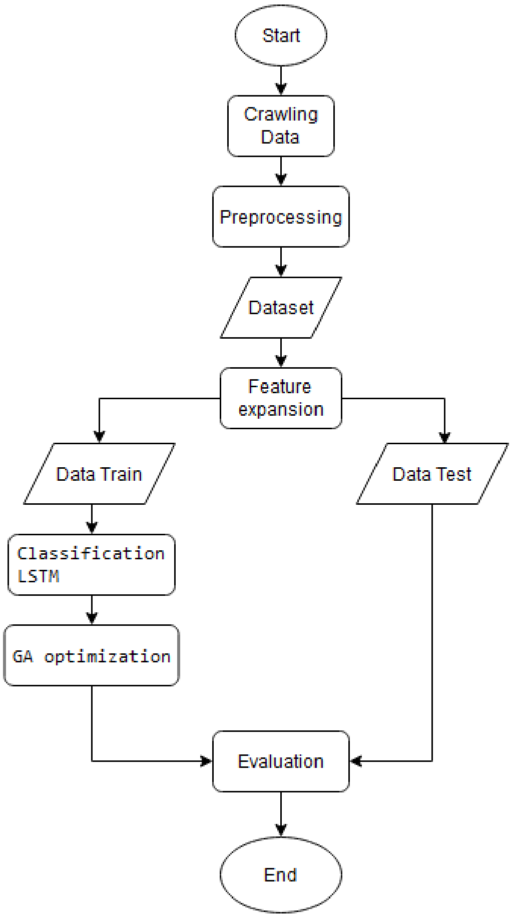

Fig 3. System Flowchart

Diagram in Figure 1 illustrates the system that will be built and executed. The process begins by preparing the dataset, which involves collecting Twitter data, removing duplicate data, and labeling the extracted data. The labeled data then proceeds to the preprocessing stage. After the data is processed through preprocessing, it undergoes modeling. The modeling stage involves utilizing word embedding with GloVe before performing classification using Long Short-Term Memory (LSTM). Following that, optimization using Genetic Algorithm is applied, and the model's performance is evaluated using evaluation metrics to assess its performance.

Dataset Preparation

The data is collected from the Twitter Application Programming Interface (API) to retrieve tweet data using keywords such as "saham bca" (BCA stock), "bank bca," and so on. In this stage, duplicate data will be removed, unnecessary columns will be discarded, and sentiment labeling will be done manually using positive, negative, and neutral labels for the tweets in Table 1.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12825
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 3. Labelled Dataset

No	Teks	Label
1	Udah paling bener beli BBCA aja gak usah tech tech segala	1 (Positif)
2	hbd bca	0 (Netral)
3	Kaget bgt bbca kok jadi 8800 ternyata error:))))	-1 (negatif)

Preprocessing Data

The preprocessing stage has the objective of transforming data from existing datasets into a more efficient and effective format, resulting in more accurate outcomes, reduced processing time, and smaller data size, while preserving the information within. This handled information can be used for building frameworks (YAHYADI & LATIFAH, 2022). Table 2 showcases a representation of preprocessed text. The following preprocessing steps were employed in this study:

a) Case Folding: Converting all letters in a sentence to lowercase, commonly applied at the beginning of a sentence, city names, personal names, etc.
b) Data Cleaning: Cleaning the data by performing various tasks, including unescaping HTML to remove HTML tags in sentences, removing URLs to eliminate links, removing mentions (words starting with "@"), removing punctuation marks, and removing numbers from sentences.
c) Stopword Removal: The Sastrawi library is utilized for the purpose of stopword removal, which involves reducing the dimensions of data by eliminating words that do not add any sentiment elements, such as personal pronouns, conjunctions, and prepositions.
d) Word Stemming: Converting words to their base form, which is crucial in text-based classification.
e) Tokenization: Splitting a sentence into individual word tokens or units, dividing the sentence into separate words referred to as tokens.

Feature Expansion

The input for this model is the data that has undergone the preprocessing steps in preprocessing data. GloVe is the model that will be used to cluster words that have similar meanings using the precomputed statistics during the data representation phase. This model will generate word vector spaces with meaningful sub-structures, which have been proven to achieve 75% accuracy on the analogy dataset.

Classification With Long Short – Term Memory (LSTM)

The input for this model is the data that has undergone preprocessing and feature expansion. The next step is to train the model with the training data and make predictions using the test data. The output of this method is the predicted sentiment label results for the test data, which will then be optimized using Genetic Algorithm.

Optimization Using Genetic Algorithms

The input for the optimization stage is the result of classification using LSTM and GA. This stage is conducted to observe the difference in accuracy results between the classification and the added optimization using Genetic Algorithm in terms of accuracy. The output of this method is the predicted sentiment label results for the test data, which will later be evaluated.

Evaluasi

The input for the optimization stage is the result of classification using Long Short-Term Memory and optimization Genetic Algorithms. This stage is conducted to observe the difference in accuracy results between the classification and the added optimization using Genetic Algorithm in terms of accuracy. The output of this method is the predicted sentiment label results for the test data, which will later be evaluated.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12825
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

Dataset and Labeling

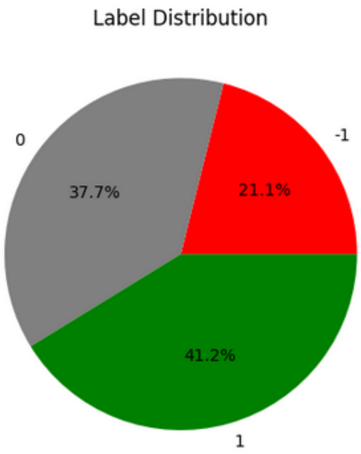

Fig 4. Label Distribution

From the data that have been collected and labeled for this study, labeling process in done manually, which involve three people for 1 data. The number of negative labels is 6332, the number of positive labels is 12374, and the number of netral labels is 113110. The distribution of these labels from the dataset can be seen in Figure 4. Label "-1" means the negative, "1" means the positive, and "0" means the netral

LSTM Classification Model Result

Table 4. LSTM Parameter

Parameter	Value
Bacth_size	512
Epoch	10
Val_split	0,1
Feature expansion	Feature expansion

The parameters used in this research to build the LSTM model are described in Table 4. The batch_size is set to 512, epoch is set to 10, and validation_split is set to 0.1. Additionally, the performance of the model will be compared when using the feature expansion technique using GloVe to represent each word in the dataset. The results for each performance metric will be presented in Table 5.

Table 5. LSTM Performance

Model	Accuracy(%)	Precision(%)			Recall(%)			F1-Score(%)		
		-1	0	1	-1	0	1	-1	0	1
LSTM	43	24	44	34	33	41	55	44	38	42
Glove-LSTM	71	43	81	66	54	65	79	60	72	69

From the experimental results conducted using three labels (-1 for negative, 0 for neutral, and 1 for positive) as shown in Table 5, the model utilizing GloVe and LSTM outperformed the LSTM model alone. The former achieved an average accuracy of 71%, precision of 63.333%, recall of 66%, and F1-score of 67%. In comparison, the LSTM model alone achieved an accuracy of 43%, precision of 34%, recall of 34%, and F1-score of 41.333%. These results indicate that with a large dataset and the inclusion of three labels (negative, neutral, and positive), the classification performance using LSTM alone is relatively low. However, by incorporating the GloVe word embedding method, which represents each word in the dataset, the LSTM model's accuracy is significantly improved.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12825 p-ISSN : 2541-044X

Genetic Algorithms

Table 6. Evolutionary GA Parameter

Evolutionary Parameter	Value
Cxpb	0.5
Mutpb	0.2
Ngen	10

Table 7. Genetic GA Parameter

Genetic Parameter	Value
Mate	Indpb = 0.5
Mutate	Mu = 0, sigma = 1, indpb = 0.2
Select	Tournsize = 3

The parameters for the GA model used in this study are described in Table 6 and 7. These parameters are set to mimic biological evolution processes or control the behavior of each population within the algorithm. The performance of the GA will be shown in Table 8. Table 8 will also compare the accuracy of each scenario built, including the LSTM model, Glove-LSTM model, and Glove-LSTM-GA model

Table 8. GA Performance

	Model	Accuracy	
	LSTM	43%	
	GloVe - LSTM	71% (+28%)	
	GloVe - LSTM - GA	87% (+16%)	

From the experimental results with the addition of optimization using the Genetic Algorithm (GA), it can be seen in Table 7 that using GA for optimization successfully achieved higher accuracy compared to the previous two scenarios. The LSTM model alone achieved an accuracy of 43%, while the Glove-LSTM model achieved 71% accuracy. However, with the parameters specified in Table 6, GA achieved the highest accuracy of 87% in sequential 27 with a total of 57 sequences. It obtained the best output shape of 100.111, dropout of 0.561, and the best individual with values such as (111.17040, 0.39841, 93, 0.2738, 109.5598, etc.). The GA also obtained the best fitness of 0.87241. On the other hand, in sequential 29, GA achieved the lowest accuracy of 38%. The process of obtaining accuracy, best individual, and best fitness required a total time of 210 minutes.

Correlation Test Result

This correlation testing uses the Rank Spearman correlation test and uses the input data of BBCA stock closing prices from January 2, 2019, to February 20, 2023. The sentiment calculation is done by summing the total sentiment, which consists of 1 for positive sentiment, 0 for neutral sentiment, and -1 for negative sentiment. To observe the correlation between the total sentiment and BBCA stock prices, please refer to figures 5 and 6.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12825 p-ISSN : 2541-044X

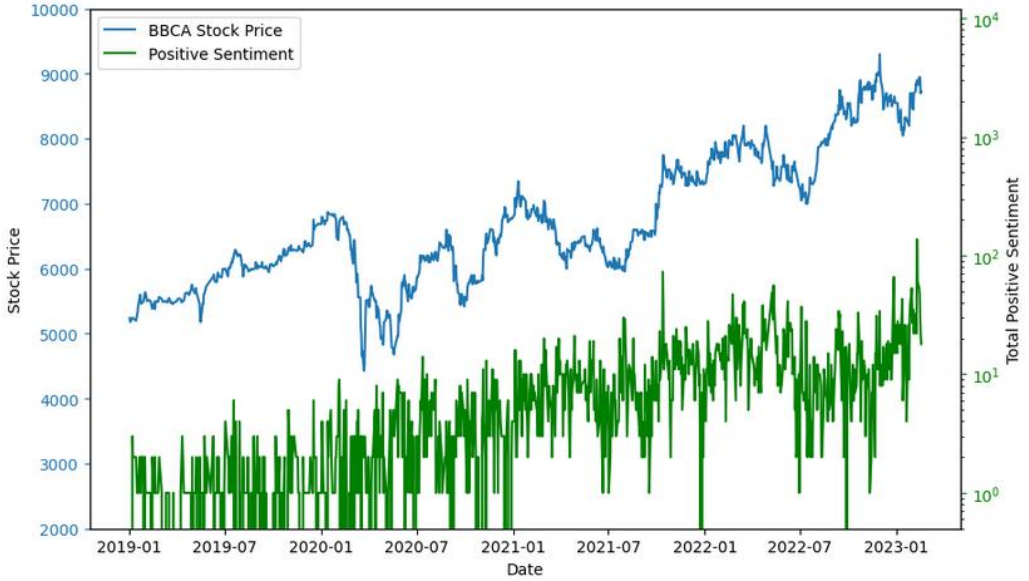

Fig 5. BBCA stock Price Chart with Predicted Positive Sentiment.

Figure 5 shows the prediction of positive sentiment from the correlation of BBCA stock prices. The stock prices appear to follow the amount of positive sentiment each year, although there was a significant decline in stock prices from January 2020 to July 2021. This is proven by the Spearman correlation test, which resulted in a correlation coefficient of 0.618.

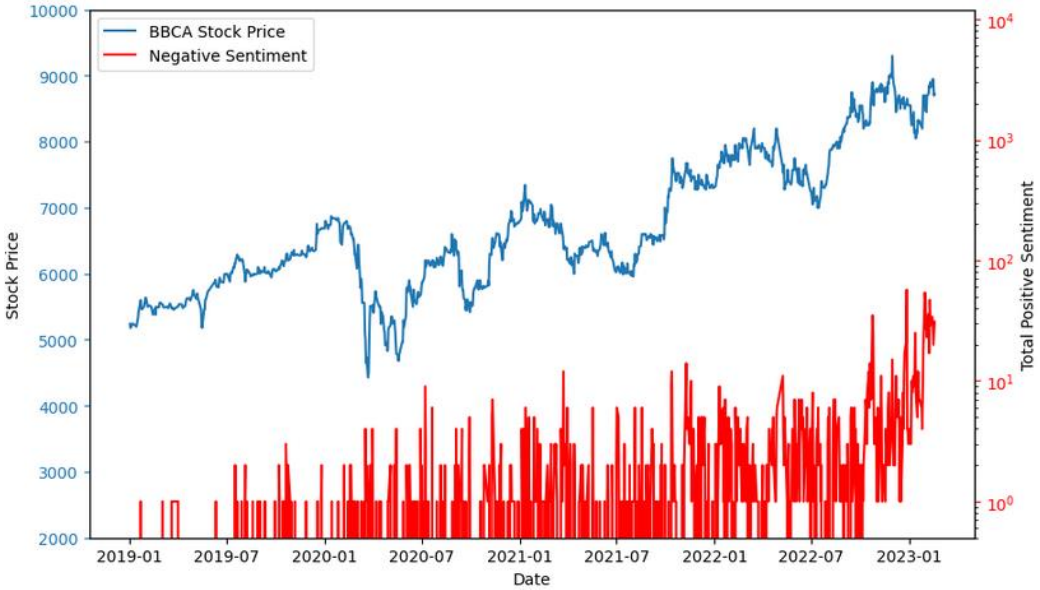

Fig 6. BBCA stock Price Chart with Predicted Negative Sentiment.

Figure 6 illustrates the negative sentiment of BBCA stock prices. The highest negative sentiment is observed from July 2022 to January 2023, indicating a significant decline in BBCA stock prices compared to previous months. This is supported by the Spearman correlation test, which resulted in a correlation coefficient of 0.684. However, sentiment from tweets is not the sole main factor affecting stocks. There are many other factors at play, such as government policies, the economy, and internal factors within the company itself. Nevertheless, sentiment from tweets can assist the public in gaining a deeper understanding of the prevailing situation.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12825
e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

The research utilized three different testing scenarios and three sentiment tweet labels. The first scenario applied Long Short-Term Memory for classification, the second scenario introduced feature expansion through Global Vectors, and the third scenario involved optimization using Genetic Algorithm. In total, 30,017 data samples were used in this study. The F1 scores showed improvement for each label when comparing scenarios 1 and 2. Moreover, the accuracy of all three scenarios increased as each new element was added. researchers assessed the relationship between sentiment in tweets and BBCA stock prices using the Rank Spearman Correlation, resulting in a correlation coefficient of 0.618 for the positive label and 0.684 for the negative label. Based on these results, both labels fall into the category of a medium relationship, as shown in table 2.

CONCLUSION

In this study, sentiment analysis is performed using Long Short-Term Memory classification and word embedding using Global Vectors, which is then optimized using genetic algorithms using a dataset of 30,017 tweets labeled as negative, neutral, and positive sentiments regarding BBCA stock on Twitter, Sentiments were calculated using Rank Spearman Correlation to measure the correlation between sentiment and BBCA stock prices, resulting in a correlation of 0.618 for positive sentiment and 0.684 for negative sentiment. These correlation values indicate a medium relationship with the movement of stock prices. The dataset was used to build three scenarios: LSTM, GloVe-LSTM, and GloVe-LSTM-GA models. The highest accuracy was achieved by the GloVe-LSTM-GA scenario with an accuracy of 87%, while LSTM achieved 43% accuracy and GloVe-LSTM achieved 71% accuracy. The optimization of the Genetic Algorithm improved the performance of the previous two scenarios by an increase of 16% in accuracy, and the use of GloVe improved the performance of the first scenario by an increase of 28% in accuracy. Based on these findings, it can be concluded that optimization successfully enhanced the performance of LSTM and GloVe-LSTM in analyzing the dataset of 30,016 tweets collected from Twitter. Further research is needed to provide more evidence for these conclusions by using larger or smaller datasets, more or fewer labels, and exploring other parameters to find optimal configurations and improve the performance of the built models.

REFERENCES

Dina, P., Dewi, A., & Suaryana, I. G. N. A. (2013). PENGARUH EPS, DER, DAN PBV TERHADAP HARGA SAHAM. *E-Jurnal Akuntansi Universitas Udayana*.

Farsiah, Laina., Misbullah, A., & Husaini. (2022). ANALISIS SENTIMEN MENGGUNAKAN ARSITEKTUR LONG SHORT-TERM MEMORY(LSTM) TERHADAP FENOMENA CITAYEM FASHION WEEK. *Jurnal Pendidikan Teknologi Informasi*, 86–94.

Hariyanto Wahyu, Dicky., & Maharani, Warih. (2020). Analisis Snetimen pada Media Sosial Twitter Berbahasa Indonesia dengan Metode GloVe. *Jurnal E-Proceeding of Engineering*.

Lee, H. M., & Sibaroni, Y. (2023). *JURNAL MEDIA INFORMATIKA BUDIDARMA Comparison of IndoBERTweet and Support Vector Machine on Sentiment Analysis of Racing Circuit Construction in Indonesia*. https://doi.org/10.30865/mib.v7i1.5380

Made Wahyuliantini, N. (2015). *PENGARUH HARGA SAHAM, VOLUME PERDAGANGAN SAHAM, DAN VOLATILITAS RETURN SAHAM PADA BID-ASK SPREAD Anak Agung Gede Suarjaya (2)*.

Nugraha, M. L., & Setiawan, E. B. (2023). *JURNAL MEDIA INFORMATIKA BUDIDARMA Bank Central Asia (BBCA) Stock Price Sentiment Analysis On Twitter Data Using Neural Convolutional Network (CNN) And Bidirectional Long Short-Term Memory (BI-LSTM)*. https://doi.org/10.30865/mib.v5i1.2293

Nurdin, A., Anggo, B., Aji, S., Bustamin, A., & Abidin, Z. (2020). PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS. *Jurnal TEKNOKOMPAK*, *14*(2), 74.

Nurvania, Jovita., Jondri., & Lhaksamana Muslim, Kemas. (2021). Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-term Memory(LSTM). *Jurnal E-Proceeding of Enginnering*.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12825
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Rahman, M. Z., Sari, Y. A., & Yudistira, N. (2021). *Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)* (Vol. 5, Issue 11). http://j-ptiik.ub.ac.id

Rustiana Program Studi Sistem Komputer Perguruan Tinggi Raharja, D., & Rahayu Magister Teknologi Informatika Perguruan Tinggi Raharja, N. (2017). ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES. *Jurnal SIMETRIS*, *8*.

Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Setiawan Irawati, Esther., Ferdianto, Adriel., Santoso, Joan., Kristian, Yosi., Gunawan., Sumpeno, Surya., & Purnamo Hery, Mauridhi. (2020). Analisis Pendapat Masyarakat rerhadap Berita Kesehatan Indonesia Menggunakan Pemodelan Kalimat LSTM. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*.

Syafira, L., & Rikumahu, B. (2020). *ANALISIS KORELASI SENTIMEN PADA TWITTER TERHADAP ABNORMAL RETURN SAHAM (STUDI KASUS PADA SAHAM INDEKS LQ45 DI TWITTER) Jurnal Mitra Manajemen (JMM Online)*. *4*(9), 1322–1335.

Wati, Risa. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes. *Jurnal Evolusi*.

Wiranata, R. B., & Djunaidy, A. (2021). Optimasi Hyper-Parameter Berbasis Algoritma Genetika Pada Ensemble Learning Untuk Prediksi Saham Yang Mempertimbangkan Indikator Teknikal & Sentimen Berita. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(3). http://jurnal.mdp.ac.id

Xiaoyan, Li., C. Raga, Rodolfo., & Xuemei, Shi. (2022). GloVe-CNN-BiLSTM Model for Sentiment Analysis on Text Reviews. *Sensors*.

YAHYADI, A., & LATIFAH, F. (2022). ANALISIS SENTIMEN TWITTER TERHADAP KEBIJAKAN PPKM DI TENGAH PANDEMI COVID-19 MENGGUNAKAN MODE LSTM. *Journal of Information System, Applied, Menagement, Accounting Adn Research*.

*name of corresponding author