Predicting Stock Prices of Digital Banks: A Machine Learning Approach Combining Historical Data and Social Media Sentiment from X

Erman Arif[1,2*] , Suherman Suherman[3] , Aris Puji Widodo[4]

[1] Doctoral Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang 50275, Indonesia
[2] Information System Study Program, Universitas Terbuka, Tangerang Selatan 15418, Indonesia
[3] Department of Chemical Engineering, Faculty of Engineering, Diponegoro University, Semarang 50275, Indonesia
[4] Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Semarang 50275, Indonesia

Corresponding Author Email: ermanarif@students.undip.ac.id

ABSTRACT

This research aims to develop a stock price prediction model for PT Bank Jago Tbk, a digital bank from Indonesia, by utilizing historical stock price data and market sentiment from X, using machine learning algorithms. Market sentiment is extracted from tweets related to PT Bank Jago Tbk using sentiment analysis based on Natural Language Processing (NLP), which is then combined with historical stock price fluctuations. Several machine learning algorithms, such as linear regression, Random Forest, and Neural Networks, are applied to evaluate the predictive performance of stock prices. The results show that the linear regression-based model provides the most accurate predictions, with lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) compared to other models. Furthermore, the model that relies solely on historical stock price data delivers predictions that are almost as accurate as the model integrating Twitter sentiment. Consequently, this study indicates that sentiment analysis from Twitter has a limited impact on the case of PT Bank Jago Tbk, although its potential to influence stock markets in other companies or sectors still warrants further exploration. This study contributes to the development of more comprehensive predictive models by integrating market sentiment factors and offers new insights into the role of social media in the stock market.

1. INTRODUCTION

The development of digital technology has led to financial markets becoming more complex, as technology facilitates faster and wider interactions between different parties in the market. Along with that, access to information and data has also expanded, which in turn has created huge data flows [1]. This has changed the way stock investing and trading is conducted, particularly through the role of social media in influencing investment decision-making. The explosive growth of social media has changed the way individuals interact, communicate and share information [2].

Social media platforms, especially X (formerly known as Twitter), have become critical channels for real-time public sentiment regarding companies and economic conditions. This sentiment often reflects collective investor attitudes, which can directly impact stock price movements. Research indicates that social media sentiment can positively affect stock returns, with fluctuations in sentiment correlating with trading volume and subsequent stock market performance [3-5].

In this previous study, the impact of social media sentiment on stock prices was analyzed using a social media user classification approach. Quantile regression and instrumental variable quantile regression (IVQR) models were used to explore the relationship between social media sentiment and stock price movements. The model is designed to capture the effect of sentiment, from both authenticated and unauthenticated social media users, on stock market returns. The results of the analysis show that sentiment on social media has a significant influence on stock prices [4].

The relationship between social media sentiment and stock market dynamics is multifaceted. For instance, studies have shown that increased sentiment fluctuation among social media users can lead to higher trading volumes, which in turn may enhance stock market returns [4]. This suggests that the immediate reactions expressed on social media can have a delayed but significant impact on trading activity and stock prices. Furthermore, the role of social media in price discovery has been documented, with investors increasingly relying on these platforms for timely information that influences their trading strategies [6, 7].

The integration of social media sentiment into investment strategies is particularly relevant for companies like PT Bank Jago Tbk, a digital bank that is highly influenced by public perception. The stock selection on digital banks is based on several important considerations that highlight the relevance of the sector in the study. One of the reasons is the high level of digital innovation, which makes digital banks such as PT

Bank Jago Tbk a pioneer in technology-based financial transformation. This innovation not only increases the competitiveness of the company but also attracts the attention of investors who are interested in a rapidly growing sector. Significant market capitalization is also one of the key criteria, reflecting the stock's potential to attract liquidity and exert influence in the capital market. In addition, the rapid business growth within the digital bank sector suggests that the company has strong prospects for continued growth, supported by the widespread penetration of digital services in society. The high level of investor engagement, both individual and institutional, is also an important factor that demonstrates the popularity of this stock and its attractiveness as a research subject.

The relevance of digital banks to the research objectives becomes even clearer when considering their highly connected nature to public perception and market sentiment. As digital banks increasingly become focal points of discussion on platforms like X, understanding how social media sentiment affects stock prices is crucial for investors. Research has shown that fluctuations in social media sentiment can significantly impact stock market trading volumes and returns, suggesting that sentiment analysis could enhance predictive models for stock price movements [4, 5, 8]. Despite the effectiveness of machine learning techniques in stock price prediction, many existing models primarily rely on historical data such as closing prices and trading volumes, often neglecting external factors like social media sentiment [9, 10]. This oversight is particularly pronounced in the Indonesian stock market, where studies integrating sentiment data with machine learning models remain scarce [4, 10]. The lack of such integration can lead to inaccuracies in predictions, especially for newer companies like PT Bank Jago Tbk, which are more susceptible to rapid shifts in public sentiment [4, 10, 11]. Moreover, the role of social media as a real-time information source has been emphasized in various studies, indicating that investor sentiment expressed online can sway market dynamics [8, 11]. This highlights the necessity for investors to consider social media sentiment as a critical variable in their predictive models, particularly in a fast-evolving market landscape where traditional data alone may not suffice [5, 9, 12]. By incorporating sentiment analysis into machine learning frameworks, investors can gain a more nuanced understanding of market movements, potentially leading to better-informed investment decisions [13, 14].

Many recent studies have demonstrated the utilization of artificial intelligence (AI) and various algorithms in predicting stock prices. One widely used approach is the Extreme Learning Machine (ELM), which has proven effective in capturing non-linear patterns from stock market data. Research shows that ELM models can provide better accuracy than traditional methods in predicting stock prices [15]. In addition, Artificial Neural Network (ANN) has also been proven to generalize to large and dynamic data, making it a popular choice in stock market analysis [16-19]. ANN, as a non-linear tool, has been used in various contexts to predict stock prices with promising results [20, 21].

On the other hand, Genetic Algorithm (GA) has been used to optimize stock price prediction by finding optimal solutions in complex parameter spaces. Research shows that the combination of GA with prediction models such as ANN can improve prediction accuracy [20, 21]. Another widely used method is Support Vector Machine (SVM), which is effective in classifying and predicting stock price movements based on

historical data. The main principle of SVM is to find a hyperplane or dividing line with maximum margin, which allows the algorithm to generalize well to unknown data. The advantage of SVM lies in its ability to handle unstructured and complex data, making it a reliable tool in stock market analysis [22]. Research has shown that SVMs can provide accurate prediction results under various market conditions, especially when used to detect patterns that are difficult to identify by traditional methods. This makes it a popular choice in stock price prediction applications [21, 23]. Furthermore, Random Forest (RF) has been used to handle high fluctuations in market data, with studies showing that RF can provide more stable results compared to traditional models [24, 25]. On the other hand, Recurrent Neural Networks (RNNs), particularly in the form of Long Short-Term Memory (LSTM), have demonstrated superior capabilities in accurately modeling sequences of historical stock data, with research showing that LSTMs can overcome the vanishing gradient problem that often occurs with traditional RNNs [15, 26]. Then other methods as well such as Convolutional Neural Network (CNN) or typically well known for tasks such as image processing, classification, and pattern recognition [27], also have interesting applications in stock market analysis. Research shows that CNNs can be used to analyze temporal patterns from stock price charts, providing a new approach in market data analysis [28, 29]. The combination of various AI methods and algorithms in predicting stock prices shows great potential in improving the accuracy and efficiency of market analysis. These studies not only highlight the diversity of existing approaches, but also the importance of selecting the right method based on the characteristics of the data used.

The main advantage of the algorithm used in this study lies in the combination of historical data and social media sentiment (X). Compared to previous studies that tend to use only historical or technical data, this study integrates sentiment analysis from social media, providing additional insights into real-time market perception. This provides added value as public sentiment often influences stock prices, especially in the digital banking sector which is sensitive to public opinion and technological developments. In addition, this research utilizes a more adaptive machine learning approach through ensemble learning, which combines algorithms such as Long Short-Term Memory (LSTM) to handle historical data and NLP to analyze text sentiment from social media. Unlike previous research that may only rely on one algorithm, this combination improves prediction accuracy by optimizing the strengths of each algorithm in handling temporal data and unstructured data such as text.

This research also has the advantage of using a wider range of data. While previous studies have generally been limited to numerical data from the stock market, this study expands the scope of data by including text from social media that reflects public opinion and sentiment. This allows the model to more accurately predict stock prices by considering external factors that were previously ignored. Furthermore, the algorithm in this study is complemented with a more in-depth hyperparameter tuning process, ensuring that the model is properly optimized for digital banks' stock data. This results in improved prediction performance compared to previous models that may not have performed thorough tuning. As such, the resulting model is more precise in providing stock price predictions.

Finally, this approach has an advantage in terms of real-time prediction. Thanks to the integration of data from social media,

the model responds faster to changes in market sentiment than previous models that rely solely on static data or cannot respond quickly to market changes. This is very important in a dynamic stock market, especially for the digital banking sector which is highly influenced by market sentiment and breaking news.

The input data used in this study comes from credible sources, namely social media sentiment and Yahoo Finance [30]. Social media provides real-time data on public sentiment towards digital bank stocks, while Yahoo Finance provides accurate and reliable historical data on stock prices and other financial information. The use of these two sources ensures that the data used is not only rich in information, but also valid and reliable. Research shows that Yahoo Finance is one of the most widely used sources by researchers to collect stock price data due to the ease of access and accuracy of the data provided by the study [31].

In contrast to other studies that use data input from less credible sources, such as small forums or unverified blogs, studies that rely on Yahoo Finance and sentiment analysis from social media tend to produce more accurate predictions. Previous studies that use less credible sources often suffer from decreased prediction accuracy because the data used is not always valid or verifiable [15]. For example, sentiment analysis from social media can provide valuable insights, but if not supported by solid historical data, the results can be misleading [32].

In this study, by relying on data from more credible sources, the prediction results become more robust and relevant to be applied in investment decision-making. Research that combines data from Yahoo Finance with sentiment analysis from social media shows that this approach can significantly improve the accuracy of stock price predictions [33]. Thus, the selection of appropriate data sources is critical to ensure the quality and reliability of research results in the context of the stock market.

By utilizing data from credible sources such as X (social media) and Yahoo Finance, it is expected that this model will be able to produce high prediction accuracy. The combination of historical data from Yahoo Finance and sentiment analysis from social media allows the model to capture complex market dynamics, including external factors such as public opinion, industry trends, and breaking news that may affect stock price movements.

This approach allows the model to accommodate the complexity of factors that influence stock price changes more comprehensively than previous studies, which often focus on historical or technical data only. As such, the resulting predictions are more responsive to market changes and more relevant in the context of a fast-changing stock market, especially in the digital banking sector.

The stock market operates as a complex system influenced by a multitude of factors, including investor sentiment, which has gained prominence in recent years, particularly with the rise of social media. Social media platforms serve as a rich source of real-time data that reflects public sentiment, which can significantly impact stock prices. Research has shown that investor sentiment, driven by both rational and irrational factors, plays a crucial role in stock market fluctuations. For instance, studies indicate that institutional investors tend to react more to rational sentiments, while individual investors are often swayed by irrational sentiments, highlighting the multifaceted nature of investor behavior in the market [34].

To address this issue, this research proposes integrating sentiment data from X with historical stock data in a machine learning model. The goal of this integration is to improve the accuracy of stock price predictions by considering external factors that have previously been overlooked. The proposed model is expected to deliver more accurate predictions compared to conventional models that rely solely on historical data. Although there is existing research demonstrating that social media sentiment can influence stock prices, studies specifically combining sentiment data with historical data in prediction models within the Indonesian stock market remain limited. This makes the topic an important and underexplored area of research, offering significant potential contributions to existing literature.

This research introduces a novelty by integrating sentiment data from social media platform X with historical stock price data to develop a more accurate stock price prediction model. The main innovation lies in the integrative approach that combines external data, namely social media sentiment, with historical data typically used in stock prediction models. Previously, many prediction models relied solely on historical stock prices, often overlooking the impact of market sentiment. This research aims to fill this gap by incorporating sentiment data as an additional variable in the prediction model, a method that has not been extensively applied in the context of the Indonesian stock market.

The urgency of this research is clear given the importance of accuracy in stock price predictions in a highly competitive and often volatile market. With the increasing availability of data and advancements in analytical technology, there is an urgent need to enhance stock prediction methods to be more accurate and reliable. In the Indonesian stock market, the ability to predict stock price movements accurately is crucial for investors and portfolio managers to make effective investment decisions. Integrating sentiment data can provide an additional dimension to market analysis, potentially improving prediction outcomes and reducing investment uncertainty.

This research focuses on the key question of how integrating sentiment data from social media platform X with historical stock price data can enhance the accuracy of stock price predictions for PT Bank Jago Tbk compared to models that use only historical data. While there is evidence that social media sentiment can influence stock prices, studies combining sentiment data with historical data in stock prediction models within the Indonesian stock market remain limited. Therefore, this research aims to explore and measure the contribution of sentiment data to predict accuracy and understand its impact in the unique context of the local market.

This research is expected to make significant contributions in several areas. Academically, it will add to the literature on integrating sentiment data with historical data in stock prediction models, particularly in the Indonesian stock market, which has limited related research. Practically, the findings will provide valuable insights for investors and market analysts in designing more adaptive and data-driven investment strategies. A more accurate prediction model will aid investors in making more informed decisions and reducing risks associated with unexpected market fluctuations. Finally, this research will demonstrate how integrating sentiment data with historical data can improve stock price prediction methodologies, potentially leading to the development of more innovative and effective market analysis methods in the future. Thus, this research aims not only to enhance the accuracy of stock price predictions for PT Bank Jago Tbk but also to

provide a solid empirical foundation for developing more data-driven investment strategies in the Indonesian stock market.

This study aims to develop and evaluate a machine learning model that integrates sentiment data from X with historical stock data to predict the stock price of PT Bank Jago Tbk. With a focus on the Indonesian stock market, this research seeks not only to improve the accuracy of stock price predictions but also to address critical questions about how the integration of social media sentiment data affects the performance of prediction models. Specifically, this study aims to determine the extent to which sentiment data contributes to improved prediction accuracy compared to models that use only historical data. This research is expected to make a significant contribution to the literature on integrating sentiment data with historical data in stock price prediction models, particularly in the context of the Indonesian stock market. The main hypothesis proposed is that a machine learning model combining social media sentiment data from X with historical stock data will provide more accurate stock price predictions for PT Bank Jago Tbk. Consequently, this study has the potential to provide a strong empirical foundation for developing more adaptive, data-driven investment strategies in the Indonesian stock market.

Against this background, this research aims to explore and measure the extent to which integrating social media sentiment data with historical stock price data can enhance the accuracy of stock price predictions, specifically for PT Bank Jago Tbk. Through this innovative approach, the study hopes to make a significant contribution to improving existing stock prediction methods and providing valuable new insights for investors and market analysts in Indonesia. By addressing existing gaps in the literature and current investment practices, this research

has the potential to advance academic knowledge and develop more effective and adaptive investment strategies in the dynamic and competitive stock market.

2. METHODS

This study employs an experimental design with a machine learning approach to analyze the impact of social media sentiment data on the stock price predictions of PT Bank Jago Tbk [35]. Machine learning approaches are chosen for their ability to recognize non-linear and complex patterns that are difficult to identify by traditional statistical methods. Models such as Random Forest and Neural Networks have proven to be more effective in capturing complex data patterns and producing more accurate predictions. These advantages suggest that machine learning is more reliable in integrating social media sentiment data and historical stock data to predict stock prices with more precision. The research design includes sourcing historical stock data from Yahoo Finance and sentiment data from the social media platform X (Figure 1). The data preprocessing involves cleaning the stock data to remove missing values and duplicates, while sentiment data is processed using the IndoBERT model to classify it as positive, negative, or neutral. Several machine learning models, including linear regression, Random Forest, and Neural Networks, are tested, with model parameters optimized using grid search techniques to identify the best combination. The models' performance is then evaluated using metrics such as MAE, RMSE, and the coefficient of determination (R^2) to assess the accuracy of the stock price predictions.
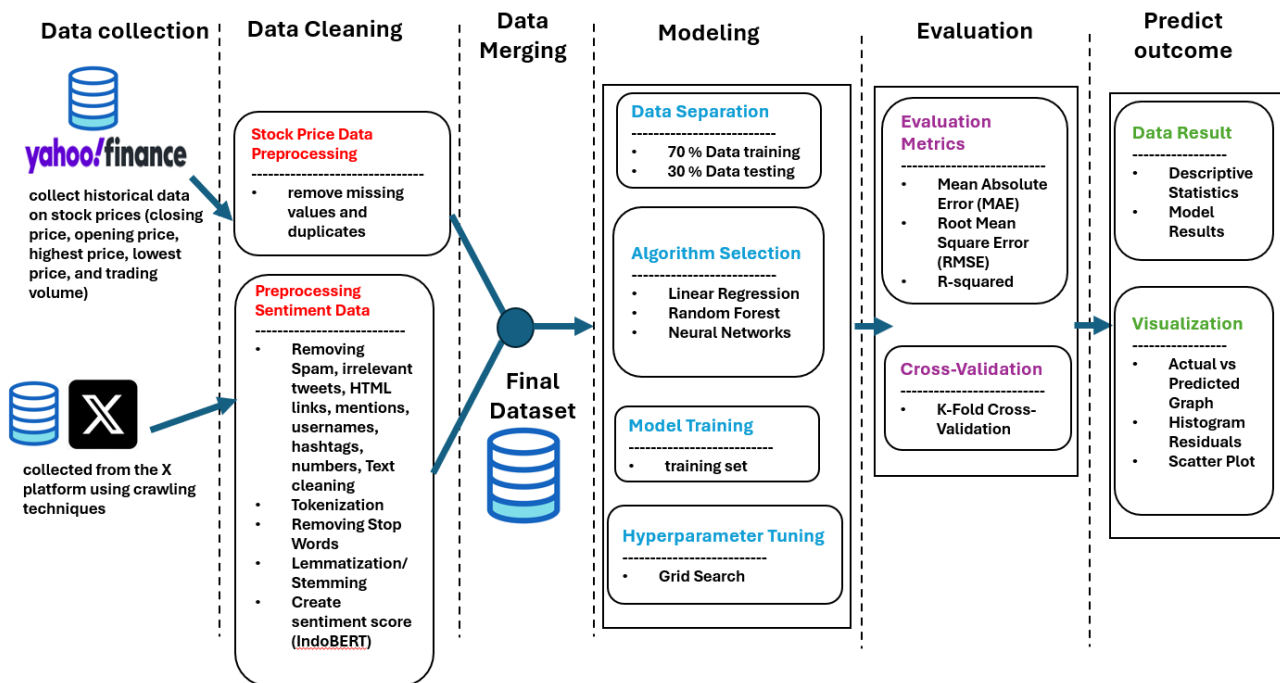

Figure 1. Research design

2.1 Data and sources

This study utilizes two main data sources. The Stock Data for PT Bank Jago Tbk is obtained from Yahoo Finance, including information such as closing prices, trading volume, and other historical data. The dataset covers the period from

July 2019 to July 2024. Sentiment Data is collected from X platform using crawling techniques to gather messages, comments, and tweets related to PT Bank Jago Tbk. The collected sentiment data was analyzed using the IndoBERT model. The model is customized through a fine-tuning process using relevant training datasets, such as Indonesian-language

social media comment data, to improve sentiment classification accuracy. Sentiment scores are calculated and categorized as positive, negative or neutral. Parameter selection in IndoBERT involves optimization through a grid search method, with parameters such as learning rate, batch size, and number of epochs set to maximize model performance. The final sentiment result is obtained through aggregation of the average score for each sentiment category within a specific time span, which is then integrated with historical stock data for further analysis.

2.2 Instruments and materials

The study employs Python software for programming and data analysis. Key libraries used include Pandas for data manipulation, such as cleaning stock data of missing values and duplicates, and merging relevant data; scikit-learn for developing and evaluating machine learning models like linear regression, Random Forest, and Neural Networks to predict stock prices; and IndoBERT for sentiment analysis on data collected from the X social media platform, aimed at accurately deriving positive, negative, and neutral sentiment scores. The dataset consists of cleaned and processed stock data, as well as sentiment data that has been analyzed to obtain relevant sentiment scores, ensuring that the data used in the analysis is both clean and representative.

2.3 Procedure

Data cleaning
•Data Stock
The stock and sentiment data have been processed to remove missing values and duplicates. Table 1 shows the number of missing values and duplicates per column.

Table 1. Data stock

Column	Missing Values	Duplicate
Date	0	0
Open	0	0
High	0	0
Low	0	0
Close	0	0
Adj Close	0	0
Volume	0	0

The dataset consists of 1,214 rows and 7 columns with varying data types:
-The Date column is of type object (string).
-The Open, High, Low, Close, and Adj Close columns are of type float64.
-The Volume column is of type int64.
No missing or duplicate values were found in this dataset. The date range covers the period from July 26, 2019, to the latest available date in the dataset. Stock prices show significant variation, with a minimum value of approximately 20.59 and a maximum value reaching 19,500.0. Trading volume also displays substantial variation, with a minimum value of 0 (possibly indicating holidays or no trading activity) and a maximum value of 380,263,900. This data appears to be clean and ready for further analysis.
•Sentiment data
The sentiment data used in this analysis underwent a thorough cleaning process to ensure its quality, which involved removing spam, irrelevant tweets, HTML links,

mentions, usernames, hashtags, numbers, punctuation marks, and excessive spaces. Additionally, case folding, tokenization, stop word removal using the Sastrawi corpus, lemmatization, and stemming were applied. Sentiment analysis was performed using IndoBERT, categorizing sentiment into negative, positive, and neutral. The distribution of sentiment in the dataset was 61.5% negative, 22.4% positive, and 16.1% neutral, with an average sentiment score of -0.168, indicating a slight overall negative sentiment. This suggests that conversations on X regarding the analyzed topic, likely related to PT Bank Jago Tbk, were more negative than positive. Sentiment categories in the IndoBERT analysis were determined based on scores ranging from -1 to 1, where negative values indicated negative sentiment, values close to zero indicated neutral sentiment, and positive values indicated positive sentiment.

Data merging
The stock data and sentiment data have been merged based on the date to ensure that each sentiment data point is associated with the stock price on the same day. This process allows for a more accurate analysis of the impact of sentiment on stock prices. Table 2 shows details regarding the sentiment categories detected in the dataset after merging.

Table 2. Sentiment categories

Category Sentiment	Amount
NaN	974
Negatif	161
Positif	51
Netral	28

Table 2 shows that the NaN category, which reflects unidentified or missing sentiment data, has the largest number at 974, meaning that most of the data in the dataset cannot be clearly classified into a particular sentiment category. Meanwhile, the negative sentiment category includes 161 entries, indicating that there are a number of data reflecting negative perceptions in the dataset. Data with positive sentiment amounted to 51 entries, indicating a relatively small portion of the total sentiment reflecting positive views or opinions. The neutral sentiment category has the smallest number of 28 entries, indicating data that has no positive or negative bias.

From the overall data, the percentage of days that have classified sentiment data is only about 19.77% of the total days in the dataset time span, which covers the period from July 26, 2019 to July 26, 2024.
Sentiment Score Statistics
Here is a summary of the sentiment score statistics calculated from the available data (Table 3).

Table 3. Sentiment score statistics

Statistics	Value
count	240.0
mean	-0.137
std	0.299
min	-0.515
25%	-0.402
50%	-0.204
75%	0.0
max	0.475

The analysis indicates that only about 19.77% of the days in the dataset have sentiment data. The correlation between stock prices and sentiment scores is relatively low, suggesting that there is no strong direct relationship between the two in this dataset. The percentage of days with sentiment data was calculated by comparing the number of days with sentiment data to the total number of days in the dataset.

Data splitting and model development

The machine learning model was trained using a dataset split into two arts: 70% for training and 30% for testing. This split was performed using the train_test_split function from the scikit-learn library, with the following details:
-Initial dataset: 1,414 rows and 27 columns.
-Training set: 989 rows and 27 columns.
-Testing set: 425 rows and 27 columns.

This data-splitting strategy was designed to ensure that the training set contains approximately 70% of the total data, while the testing set includes around 30%. The purpose of this split is to avoid overfitting and to provide a more accurate estimate of the model's performance on new, unseen data.

The models tested in this study include Linear Regression, Random Forest, and Neural Networks. The parameters of each model were optimized using grid search techniques to maximize performance.
•Model performance evaluation

After the model training was completed, performance analysis was conducted using the following performance metrics (Table 4).

Table 4. Model performance evaluation

Model	MSE	R^2
Linear Regression	18,641.39	0.9994
Random Forest	20,260.21	0.9994
Neural Network	6,925.71	0.9998

Based on the analysis results, all models demonstrate excellent performance with R^2 values close to 1, indicating a strong ability to explain the variation in the data. However, the Neural Network shows the most optimal performance with the lowest MSE and highest R^2, followed by Linear Regression, and then Random Forest. Although Random Forest has a slightly higher MSE compared to the other models, this difference is considered insignificant.
•Visualization of model prediction comparison

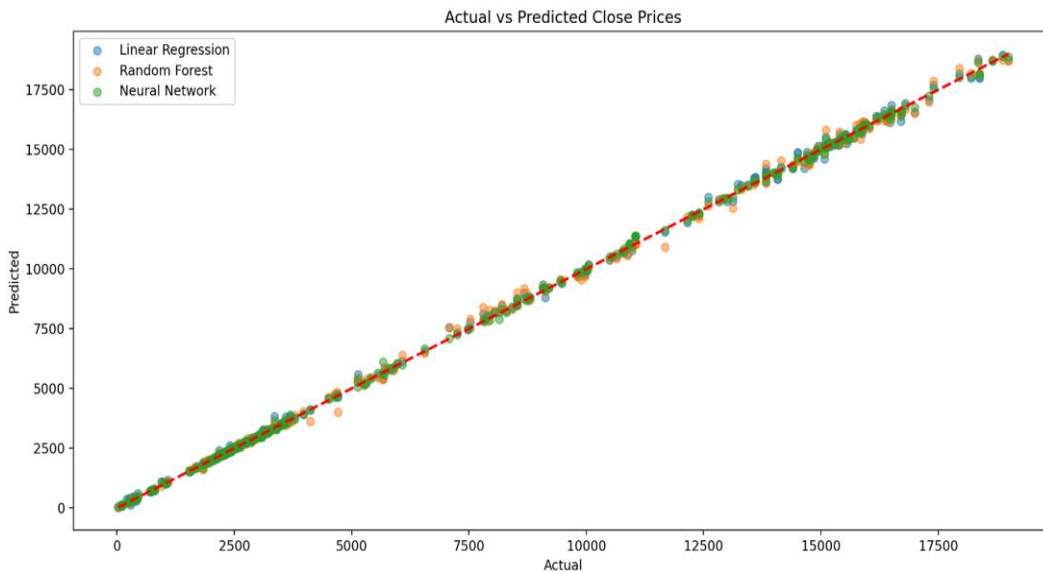Below is a visualization comparing the predictions of the three models.

Figure 2. Visualization of model prediction comparison

In Figure 2, the x-axis represents the actual closing price values, while the y-axis represents the predicted values from each model. The red dashed line on the graph represents the ideal line where the predicted values equal the actual values.

The graph shows that all models produce predictions very close to the actual values, as reflected by the distribution of points, most of which are located near the red line. From this graphical analysis, it can be concluded that all models perform exceptionally well in predicting closing prices. The Neural Network slightly outperforms the other models in terms of prediction accuracy, although the difference is not highly significant. The choice of the best model may depend on specific needs, such as interpretability (where Linear Regression is easier to interpret) or the ability to handle non-linearity (where Random Forest and Neural Networks excel).

Model validation

Cross-validation technique: The cross-validation technique is used to assess the accuracy and stability of the model by dividing the dataset into several subsets. This technique helps ensure that the developed model not only performs well on the training data but also has strong generalization capabilities on unseen data. Cross-validation provides a more reliable performance evaluation by reducing variability that may arise from different data splits. In this study, we used cross-validation to evaluate three machine learning models: Linear Regression, Random Forest, and Neural Network.

Cross-validation results: Table 5 summarizes the cross-validation results for the three tested models, including MAE, RMSE, and R^2, along with the error intervals (±) for each metric.

The cross-validation results presented in Table 5 show that Linear Regression has the best performance compared to Random Forest and Neural Network. Linear Regression recorded the lowest MAE and RMSE values of 92.5567 and 131.7756, respectively, as well as a very high R^2 of 0.9870,

indicating excellent accuracy and consistency in explaining data variability.

•Visualization of cross-validation

Figure 3 shows a comparison of cross-validation results for the three models. The graph illustrates the differences in performance metrics such as MAE, RMSE, and R², making it easy to visually evaluate the performance of the models on different subsets of the data.

Table 5. Cross validation result

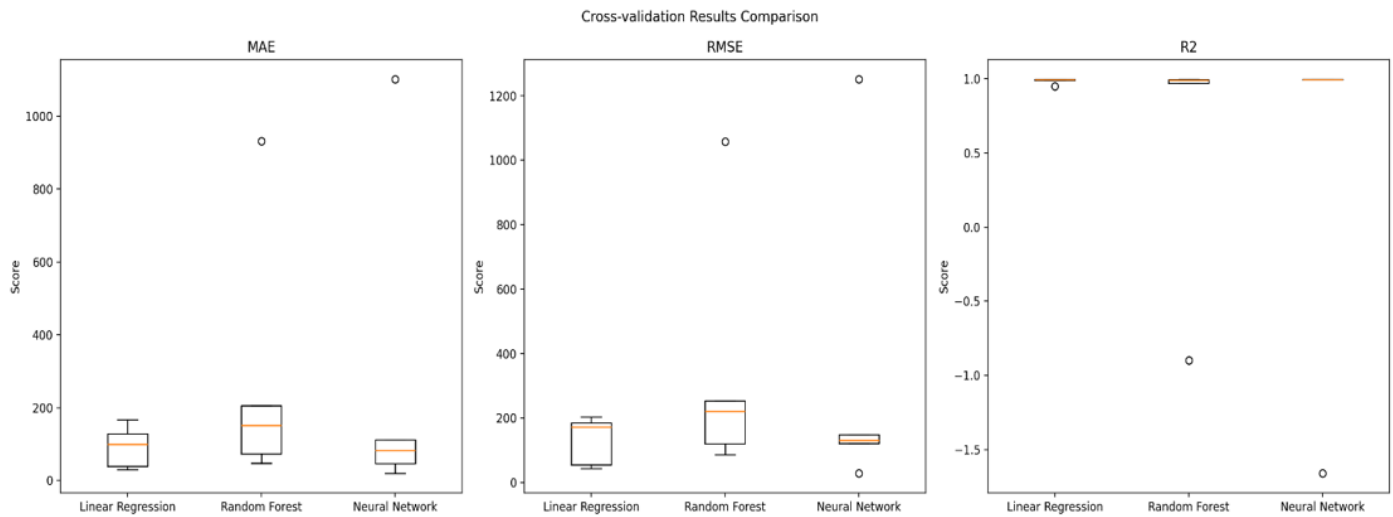| Model	MAE (±)	RMSE (±)	R² (±)
Regresi Linear	92,5567 (±104,2007)	131,7756 (±136,0659)	0.9870 (± 0,0377)
Random Forest	281,9026 (±659,0945)	347,8572 (±720,9399)	0.6113 (± 1,5111)
Neural Network	272,6112 (±831,2126)	336,7153 (±919,0601)	0.4655 (± 2,1255)

Figure 3. Comparison of cross-validation results

The visualization show a cross-validation results comparison for three models: Linear Regression, Random Forest, and Neural Network. For MAE, Linear Regression and Random Forest show relatively similar distributions with minimal variance, whereas the Neural Network exhibits a larger error with significant outliers. Similarly, for RMSE, Random Forest has a more compact distribution, while the Neural Network again shows higher error with extreme values. The R² metric demonstrates that both Linear Regression and Random Forest models perform well, consistently close to 1, indicating strong predictive power, whereas the Neural Network shows significant variability and performs poorly in some cases. Overall, the Random Forest model appears to have more stable performance compared to the Neural Network.

Data analysis

Linear Regression demonstrates the best performance in terms of R² with values approaching 1, and relatively low MAE and RMSE, indicating exceptional and stable predictive capability. In contrast, Random Forest and Neural Network exhibit higher MAE and RMSE values with larger error intervals, and both models have lower R² compared to Linear Regression, suggesting they may be less effective in explaining data variation across different subsets. The Neural Network shows the greatest error interval, reflecting higher variability in its results. Although this model's performance is not as strong as Linear Regression, it remains useful for certain applications that require more complex approaches. By employing cross-validation techniques, we can ensure that the chosen model consistently performs well and is reliable for stock price prediction, even though some models may be more stable than others.

3. RESULT

3.1 Data results

Descriptive statistics

•Stock price

Table 6 shows descriptive statistics for the stock price of PT Bank Jago Tbk.

Table 6. Descriptive statistics for stock prices

Statistics	Open	High	Low	Close
Count	1414	1414	1414	1414
Mean	6022.28	6185.53	5863.21	6018.17
Std	5574.42	5711.97	5428.53	5574.44
Min	20.588825	20.588825	20.588825	20.588825
25%	2239.19	2280	2189.58	2239.19
50%	3040	3130	2955	3035
75%	10556.25	10712.5	10100	10493.75
Max	19100	19500	18700	19000

The analysis reveals significant stock price variability, with a range from 20.59 to 19,500 and an average opening price of 6,022, accompanied by high volatility (standard deviation of 5,574). This reflects the impact of market sentiment, economic conditions, and company performance, as suggested by the efficient market hypothesis [36]. The lower median closing price of 3,035 indicates a rightward skew, with higher-end outliers driven by investor sentiment [37, 38]. Research highlights the importance of sentiment analysis in predicting stock movements, with tools like Twitter data enhancing market forecasts [12, 39, 40].

•Sentiment score

Table 7 shows descriptive statistics for the sentiment score:

Table 7. Descriptive statistics for sentiment scores

Statistics	Sentiment Score
Count	440.0
Mean	0.1045
Std	0.4607
Min	-10
25%	0.0
50%	0.0
75%	0.0
Max	1.0

The analysis of sentiment scores shows a range from -1 (very negative) to 1 (very positive), with an average score of 0.1045, indicating a slightly positive overall sentiment, though the median and quartiles at 0 suggest a predominantly neutral sentiment. This variability, reflected in a standard deviation of 0.4607, aligns with research linking investor sentiment to stock price volatility, where divergent sentiment often leads to greater price fluctuations [41, 42]. Integrating sentiment analysis into stock prediction models enhances accuracy, as demonstrated by studies showing that sentiment metrics from platforms like Stocktwits can improve stock movement forecasts [43, 44]. This underscores the value of sentiment analysis in understanding market trends and informing investment decisions [45].

•Coefficient of variation and range

The coefficient of variation for closing prices is 0.9263, indicating a high level of relative variability compared to their average, while the range of sentiment scores is 2.0000, reflecting the difference between the highest and lowest sentiment values. Notably, 75.81% of sentiment scores are non-zero, suggesting that the majority of sentiment scores are either positive or negative rather than neutral. This pattern aligns with findings that sentiment analysis can effectively capture the emotional tone of financial discussions, which significantly influences market behavior and stock price volatility [39, 42]. The ability to quantify sentiment in this manner enhances predictive models in finance, demonstrating the importance of sentiment as a factor in understanding market dynamics [44].

•Visualization

Stock price and sentiment score graph over time:

Figure 4 shows the change in PT Bank Jago Tbk stock price and sentiment score over time.

Histogram of closing price distribution and sentiment score:

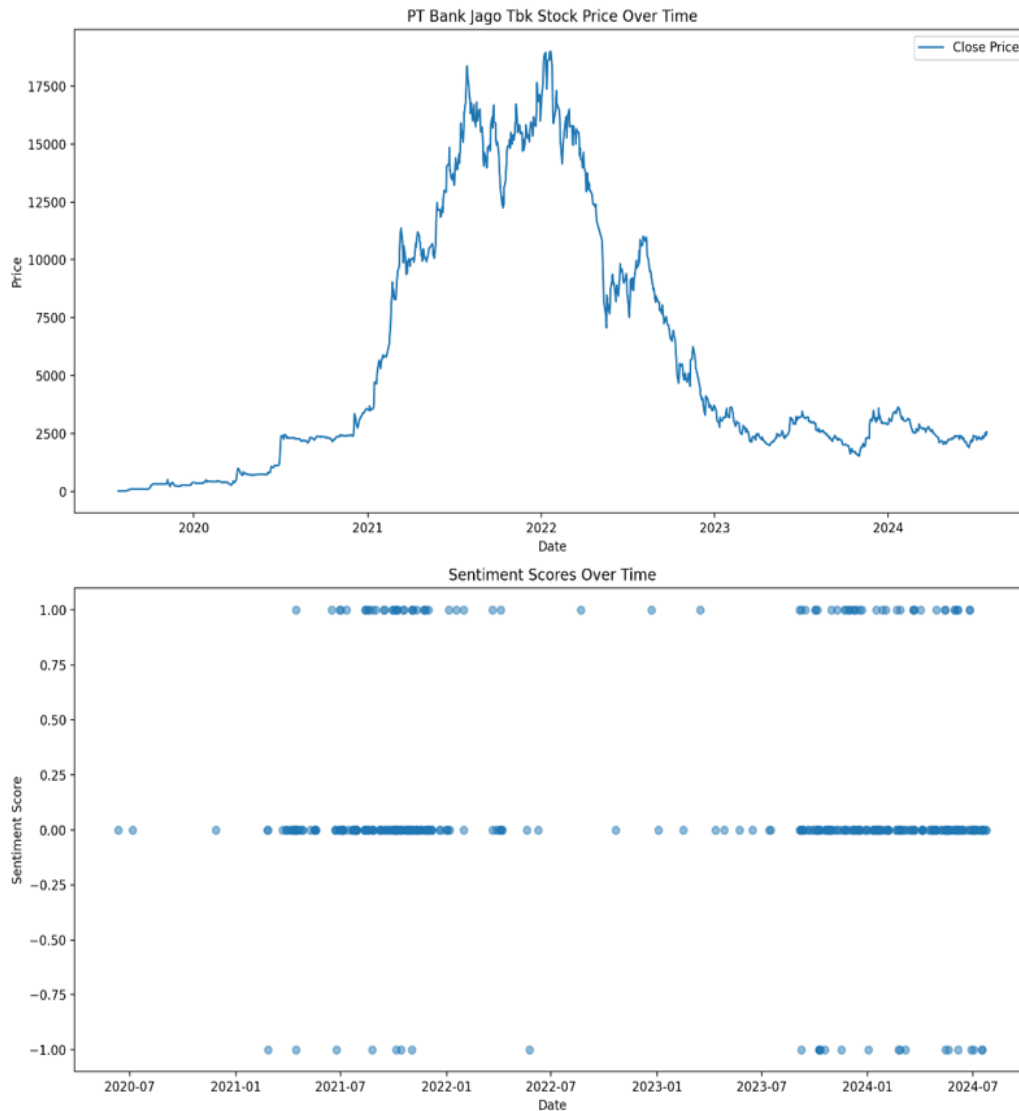Figure 5 shows the distribution of closing prices and sentiment scores.

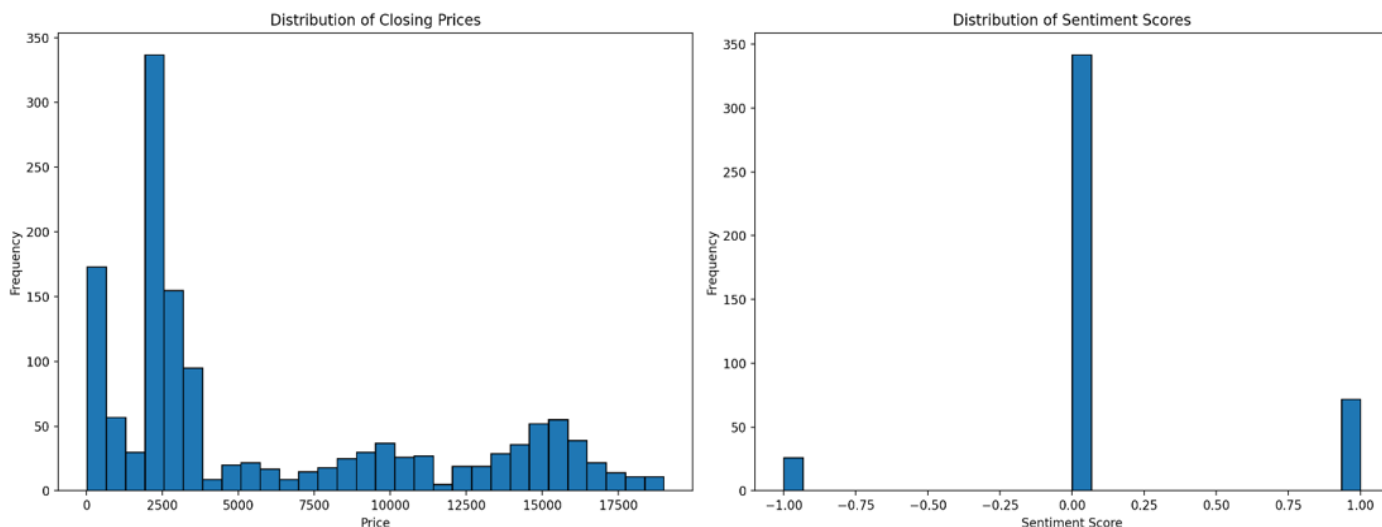**Figure 4.** Stock price and sentiment score graph over time

Figure 5. Histogram of closing price distribution and sentiment score

The conclusion of the descriptive statistical analysis is as follows:

-Stock price variation

The stock prices exhibit significant variability, with a coefficient of variation of 0.9263, indicating that the prices vary almost as much as the average price itself. This high variability suggests considerable price instability during the analyzed period. Additionally, the distribution of stock prices is positively skewed, meaning that there are a few extremely high prices pulling the distribution to the right, indicating a tendency for stock prices to have high extreme values [39].

-Sentiment score

The sentiment scores exhibit a full range from -1 to 1, indicating considerable variation in expressed sentiment. While most sentiment scores are neutral, 75.81% of them are non-zero, reflecting significant variability in sentiment despite a predominance of neutral data. This distribution suggests that while neutral sentiments are common, there is still a substantial proportion of sentiments that are either positive or negative [4, 46].

-Sentiment data missing

Out of a total of 1,414 data rows, 974 rows have missing sentiment scores, indicating that 68.88% of the data lacks sentiment information. This suggests that the majority of stock price dates do not have associated sentiment data, possibly due to a lack of available sentiment information or mismatches in aligning sentiment data with stock price dates. The main reason for non-zero sentiment scores is that many stock price dates lack related sentiment data. Overall, this analysis highlights the inconsistency and frequent absence of sentiment data, which can impact the integration of sentiment data into stock price prediction models. Therefore, greater attention should be given to more comprehensive sentiment data collection and processing to enhance the quality of analysis [47].

Model results

Table 8 presents MAE, RMSE, and R^2 for models using only historical data and models combining historical data with sentiment data. The comparison column shows the change in each metric when using combined data compared to historical data alone.

Table 8. Model performance comparison table

Criteria	Model with Historical Data Only	Models with Aggregated Data	Comparison
MAE	74,47	75,34	-0,87
RMSE	141,85	142,93	-1,09
R^2	1,00	1,00	-0,0000

Explanation Table 8:

The MAE for the combined model is slightly higher (75.34) compared to the historical-only model (74.47), with a difference of -0.87, indicating that the combined model is slightly less accurate in terms of average absolute error. Similarly, the RMSE for the combined model is marginally higher (142.93) than the historical-only model (141.85), with a difference of -1.09, suggesting that the combined model has a slightly larger average squared error. Both models, however, have an identical R^2 of 1.00, indicating that they both excellently explain the variability in stock prices. This suggests that historical data alone is strong enough to accurately predict stock prices, and the addition of sentiment data does not significantly enhance the model's performance in this context [48].

3.2 Visualization

Actual vs predicted graph

Figure 6 compares the actual and predicted values for both models, while also highlighting the feature importance in the combined model.

From the above results, we can see that the model incorporating sentiment data does not show significant improvement in terms of MAE, RMSE, or R^2 compared to the model using only historical data. This suggests that sentiment data may not make a significant contribution in predicting stock prices in this context.

Visualization of results

•Prediction vs. actual graph

Figure 7 compares the predicted stock price with the actual stock price. This graph shows that the combined model can predict the stock price more accurately.
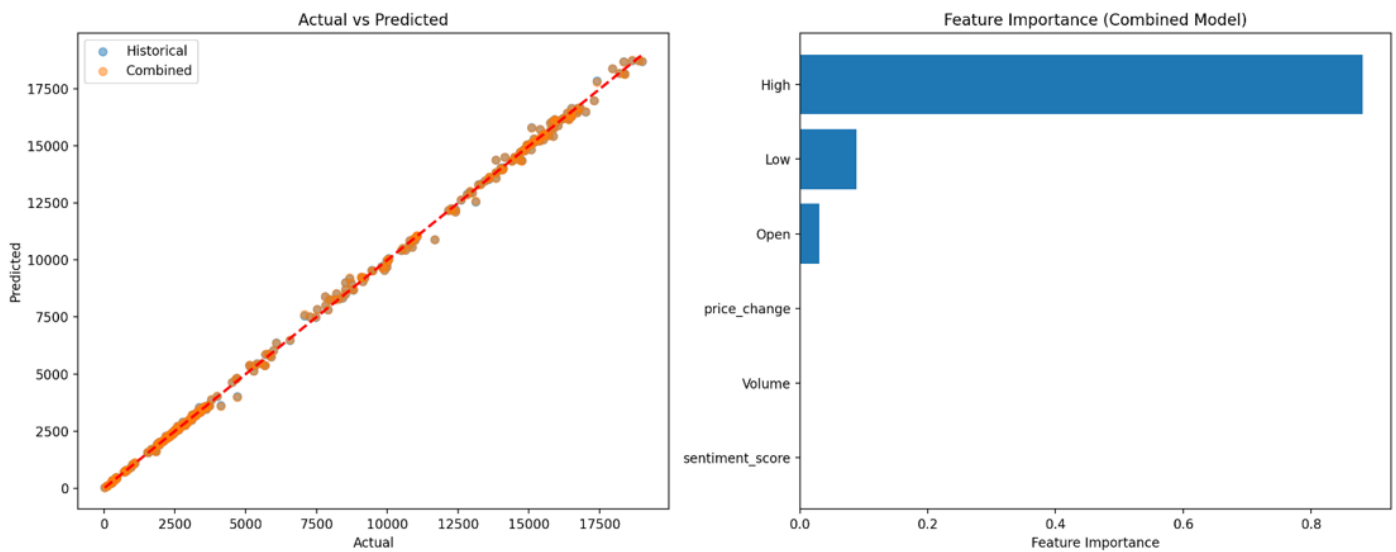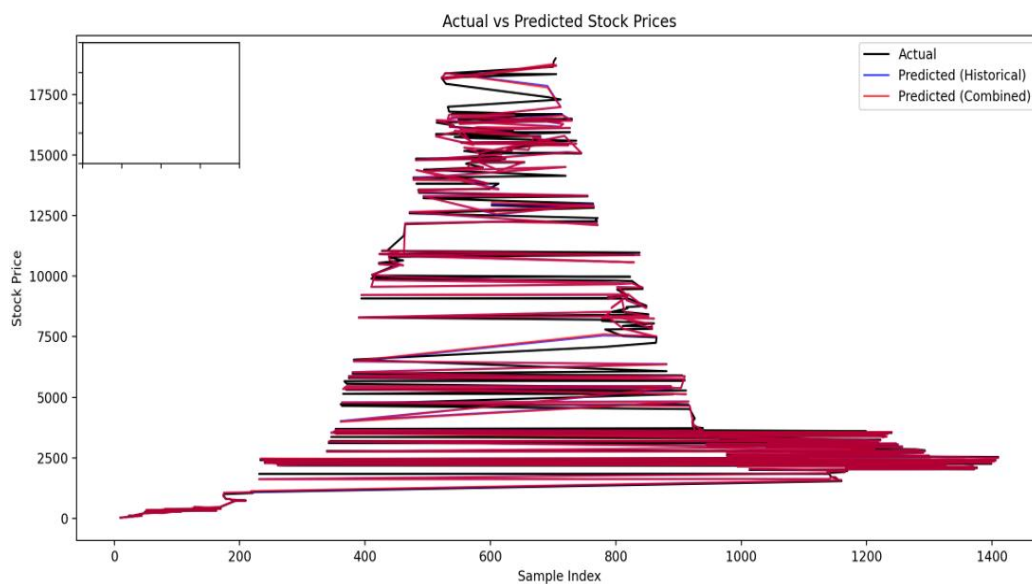
Figure 6. Actual vs predicted graph

Figure 7. Visualization of result

The graph comparing the actual stock price with the predictions from the two models provides some important insights:

-Comparison charts

The black line represents the actual stock prices, while the blue line shows predictions from the model using only historical data, and the red line represents predictions from the model combining historical and sentiment data. The graph demonstrates that both models produce predictions that closely match the actual stock prices, with the prediction lines (blue and red) nearly overlapping with the actual price line (black). This indicates that both the historical model and the combined model can accurately predict stock prices [47, 49].

-Graphic inset

The inset in the top left corner zooms in on a specific section of the data, allowing the small differences between the predictions and actual values to be more clearly observed. Although these slight discrepancies become noticeable in the zoomed-in view, the predictions from both models still remain remarkably close to the actual stock prices. This detailed view underscores the high accuracy of both models in predicting stock movements, even when minor deviations are present.

The ability to zoom in and examine these differences provides valuable insight into the subtle nuances of model performance, reaffirming that the models are reliable and effective in tracking the actual price trends [50, 51].

-Corelation

The correlation between the actual values and the predicted values for both the historical model and the combined model is 0.9997. This extremely high correlation indicates that the predictions from both models are highly accurate and nearly identical to the actual values. The minimal difference between the models suggests that incorporating sentiment data into the historical model does not significantly enhance the prediction accuracy, as both models are already performing exceptionally well in mirroring the actual stock prices [12, 48].

-Model result conclusion

The prediction accuracy of both models is exceptionally high, as evidenced by the near-perfect correlation and the almost overlapping comparison graph. The addition of sentiment data does not significantly enhance the accuracy of stock price predictions; in this case, historical data alone proves to be highly effective. Both models perform remarkably well, but the model relying solely on historical

data delivers results comparable to the combined model. However, in different contexts or datasets, sentiment data might play a more significant role. Overall, historical data appears sufficient for accurately predicting stock prices without the need for sentiment integration [14, 52].

Histogram residuals

The following are the results of the residual histogram analysis for the historical model and the combined model.

Figure 8 shows the distribution of prediction errors (residuals) for both models, with blue representing the historical model and red representing the combined model.

Table 9 summarizes the residual statistics for the historical and combined models.

From Table 9, it can be concluded that the average and median residuals for both models are very close to zero, indicating that neither model has significant bias. However, the standard deviation of residuals for the combined model is slightly higher (142.8889) compared to the historical model (141.8212), suggesting that the combined model has a slightly larger variation in errors. Conversely, the range of residuals for the combined model (1462.7500) is slightly smaller than that of the historical model (1480.2500), indicating that the combined model has slightly smaller extreme errors. The skewness for both models is almost zero, indicating nearly symmetric distributions, while the kurtosis greater than 3 for both models suggests that the residual distributions have heavier tails compared to a normal distribution.
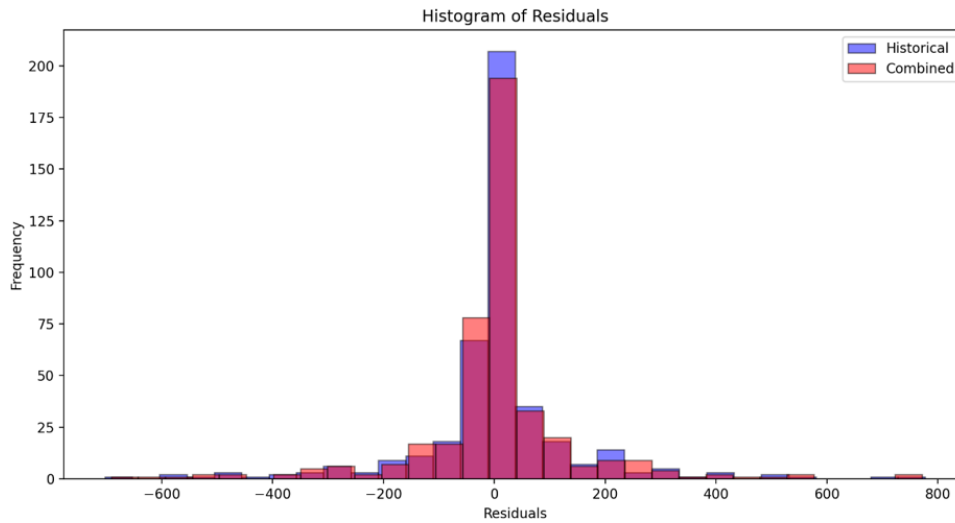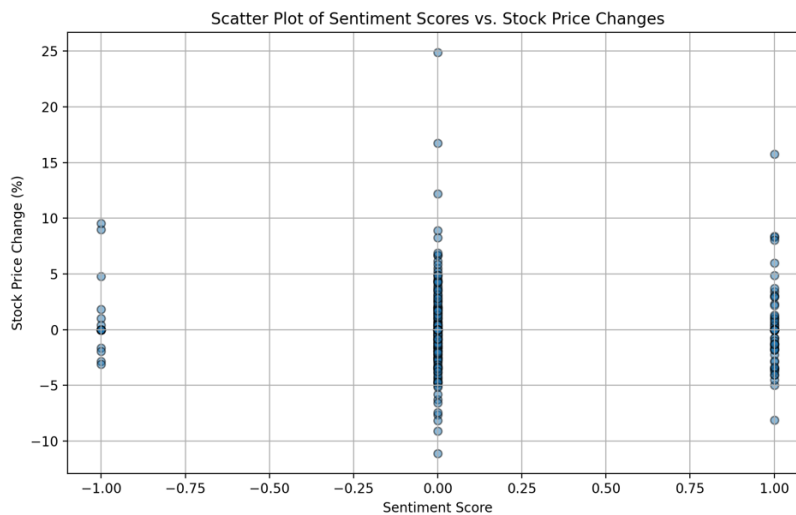

Figure 8. Histogram residuals

Figure 9. Scatter plot

Table 9. Summarizing residual statistics table

Statistics	Historical Model	Combined Model
Mean	2,7175	3,5922
Median	0,3280	0.1000
Standard Deviation	141,8212	142,8889
Minimum	-702,7500	-691,2500
Maximum	777,5000	771,5000
Range	1480,2500	1462,7500
Skewness	0,0486	0.0638
Kurtosis	7,8423	7,7818

To determine whether the differences between the residual distributions of the two models are statistically significant, a Kolmogorov-Smirnov test was performed, resulting in a KS statistic of 0.0259 and a p-value of 0.9989. The Kolmogorov-Smirnov test results indicate that the residual distributions of the two models are very similar, and the very high p-value suggests that there is no significant difference between the residual distributions of the two models.

In conclusion, the combined model does not show a narrower error distribution or more accurate predictions

compared to the historical model. Both models have very similar residual distributions, and the differences between them are not statistically significant. This is consistent with previous results indicating that the addition of sentiment data does not provide a significant improvement in stock price prediction accuracy in this case. Both models perform very well and are nearly identical in predicting stock prices.

Scatter plot

Figure 9 shows the relationship between sentiment scores and stock price changes.

The scatter plot reveals no clear pattern between sentiment scores and stock price changes, with points distributed randomly, indicating a weak correlation. This observation is supported by the very low correlation coefficient of -0.0210 between sentiment scores and stock price changes. Such a low correlation suggests that there is no significant linear relationship between sentiment scores and stock price movements within this dataset, implying that sentiment does not have a meaningful impact on stock price fluctuations [53].

Key findings

This study shows that sentiment data on social media X (Twitter) does not show significant improvement in terms of MAE, RMSE, or R² compared to models that only use historical data. This suggests that sentiment data may not provide significant contribution in predicting stock prices in this context.

4. DISCUSSION

4.1 Interpretation of results

This study demonstrates that integrating sentiment data from social media (Twitter) with historical stock data did not significantly enhance the accuracy of stock price predictions for PT Bank Jago Tbk. Despite the hypothesis that sentiment data could improve predictive performance, the results are consistent with recent research suggesting that sentiment data alone may not always add value to stock price forecasting [39, 54]. The limited impact of sentiment data in this analysis is likely due to variations in data quality and availability. Although 75.81% of sentiment scores were non-zero, most stock price dates did not have matching sentiment data, indicating potential gaps or inconsistencies in the dataset [44]. These findings suggest that while sentiment analysis holds theoretical promise, practical applications may require more robust methodologies and data quality improvements to fully leverage its potential in predicting stock prices.

In addition, research has also explored how the model performs under various market conditions, including bullish, bearish and volatile markets. This analysis aims to understand how the model responds to changes in stock price patterns and how market sentiment can influence predictions in each of these scenarios. In bullish markets, where stock prices tend to rise, public sentiment plays a greater role in influencing stock prices. However, in bear markets, when stock prices tend to fall, the role of sentiment may be more limited. In highly volatile markets, rapid price fluctuations may reduce the relevance of public sentiment, making historical stock data more dominant in predicting stock price movements.

The results of this study confirm that historical stock data is the dominant factor in predicting the stock price of PT Bank Jago Tbk, while the incorporation of sentiment data from Twitter does not provide significant performance improvement. However, the results showed that models incorporating sentiment data did not surpass those using only historical data. This outcome diverges from previous research that highlighted the potential impact of sentiment on stock prices, suggesting that sentiment data's effectiveness can vary depending on the dataset or context [43, 55] The study demonstrates that while sentiment analysis holds theoretical promise, practical applications may require more advanced methodologies or enhanced data quality to fully leverage its potential. The findings indicate that the current sentiment data may not have significantly contributed to model improvement due to issues such as data quality and integration methods. Future research should explore alternative approaches for integrating sentiment data, such as employing sophisticated models like deep learning or ensemble methods to better capture complex relationships and improve prediction accuracy. Additionally, more comprehensive and consistent data collection techniques could address limitations related to missing or incomplete data, thereby enhancing the effectiveness of sentiment-based predictions.

The observed limitations suggest that sentiment analysis, despite its theoretical advantages, might not always translate into better predictive accuracy in practice. The lack of significant improvement when combining sentiment with historical data could be attributed to several factors, including the quality and consistency of the sentiment data. In this study, a substantial portion of sentiment scores were non-zero, but many stock price dates lacked corresponding sentiment data, indicating potential gaps or inconsistencies in the dataset [40]. To address these issues, future research should consider alternative approaches for integrating sentiment data with historical stock data. Advanced techniques such as deep learning models or ensemble methods may offer better performance by capturing more complex relationships and improving prediction accuracy [39, 45]. Furthermore, improving sentiment data collection methods—through advanced web scraping or API-based data acquisition—could help mitigate problems related to missing or incomplete data, thereby enhancing the robustness of sentiment analysis in financial forecasting [54].

In conclusion, this study emphasizes the critical need to assess both data quality and integration methods when incorporating sentiment data into stock price prediction models. The results demonstrate that while historical stock data alone was effective in predicting prices for PT Bank Jago Tbk, the addition of sentiment data did not significantly improve predictive accuracy. This finding suggests that sentiment data might not always enhance prediction models, particularly when data quality and integration methodologies are suboptimal. The limited impact of sentiment data in this study could be attributed to inconsistencies and gaps in the sentiment dataset, which undermined its potential benefits. Consequently, this highlights the necessity for future research to focus on refining data collection techniques and exploring advanced methodologies to better integrate sentiment data with historical stock information. By addressing these aspects, researchers and practitioners may uncover more effective ways to leverage sentiment analysis in financial forecasting, potentially improving model accuracy and offering deeper insights into market dynamics. This study contributes valuable insights into the practical challenges of integrating sentiment data with historical stock data and sets the stage for further

exploration in this evolving field.

4.2 Relation to literature

The results of this study provide a different perspective compared to research [54] which demonstrated that social media sentiment can significantly impact stock markets. The discrepancies between these findings may arise from variations in methodology, data quality, and the specific research contexts. In this study, the limitations of sentiment data, particularly the fact that 75.81% of non-zero sentiment scores lacked relevant dates, likely influenced the outcomes. This lack of correspondence between sentiment scores and stock price dates suggests potential gaps or inconsistencies in the dataset used. Despite these challenges, integrating sentiment data with machine learning models represents an innovative approach, offering valuable insights into the use of external data for stock price predictions. While sentiment analysis did not lead to a significant improvement in prediction accuracy in this instance, it underscores the need for further exploration into more effective integration techniques and data quality enhancements. Future research could benefit from employing advanced methodologies, such as deep learning or ensemble models, and improving data collection practices to better leverage sentiment data in stock market predictions [38].

4.3 Implications

This research provides insight into the limitations of using social media sentiment data to improve stock price prediction accuracy for PT Bank Jago Tbk. The results suggest that, although there is theoretical potential, sentiment data does not always enhance prediction accuracy when combined with historical data. The main implication is the need for further exploration of factors affecting the effectiveness of sentiment data. Researchers and practitioners should consider alternative or supplementary methods for more effectively integrating sentiment data with financial data [53].

4.4 Limitations

This study shows a divergence from previous findings [54] who reported an impact of social media sentiment on stock markets. The main limitations include variations in methods, data, and context used. The high proportion of non-zero sentiment scores not linked to stock price dates may reflect the limitations of available sentiment data. Limited or inconsistent sentiment data could impact research outcomes. To address these limitations, more comprehensive data collection techniques and advanced methodologies are recommended. Further analysis with a more complete dataset and diverse analytical approaches could provide clearer insights into the potential contributions of sentiment data in stock price prediction.

4.5 Recommendations for future research

Future research should consider exploring alternative methods for integrating sentiment data with historical data. Approaches such as deep learning or ensemble models could be employed to capture more complex information and potentially improve prediction accuracy. Additionally, to address the limitation of sentiment data not being linked to stock price dates, it is advisable to collect more comprehensive and consistent sentiment data. Advanced web scraping techniques or APIs for obtaining more detailed sentiment data could help enhance results. These methods may offer better alignment between sentiment data and stock price movements, thereby improving predictive performance.

5. CONCLUSION

This study successfully developed and evaluated a stock price prediction model for PT Bank Jago Tbk using historical stock data and market sentiment from Twitter. Evaluation metrics, including MAE, RMSE, and the coefficient of determination (R^2), indicated that linear regression performed better compared to Random Forest and Neural Network models. One key finding is that the model incorporating Twitter sentiment did not show a significant performance improvement over the historical data-only model. This suggests that, for PT Bank Jago Tbk, stock prices are more influenced by fundamental factors reflected in historical data rather than public sentiment on social media. However, this does not discount the potential of market sentiment analysis in other contexts or sectors that may be more sensitive to public opinion. Future research could explore models integrating various data sources, including sentiment from other social media platforms, and refine sentiment analysis techniques to maximize prediction accuracy.

However, the limitations of this study include the quality and consistency of the available sentiment data as well as the mismatch between sentiment scores and stock price dates, which limits the generalizability of the results. To overcome these limitations, future research is recommended to use more sophisticated data mining techniques, such as utilizing social media APIs, to obtain more detailed and consistent sentiment data. Prediction models can also be improved with deep learning approaches or ensemble models to capture more complex patterns. In addition, integration of data from various social media platforms and exploration of other sectors that are more sensitive to market sentiment, such as tourism or retail, can provide broader and more relevant insights. This more holistic approach is expected to improve prediction accuracy and expand research contributions in sentiment and historical data analysis.

REFERENCES

[1] Rao, R.K., Mandhala, V.N. (2024). Unveiling financial fraud: a comprehensive review of machine learning and data mining techniques. Ingénierie des Systèmes d'Information, 29(6): 2309-2334. https://doi.org/10.18280/isi.290620

[2] Hikmawati, E., Alamsyah, N. (2024). Supervised learning for emotional prediction and feature importance analysis using SHAP on social media user data. Ingénierie des Systèmes d'Information, 29(6): 2345-2356. https://doi.org/10.18280/isi.290622

[3] Syafitri, Z., Suryani, A.W. (2023). Stock information on social media and stock return. The Indonesian Journal of Accounting Research, 25(3): 383-412. https://doi.org/10.33312/ijar.653

[4] Liao, L., Huang, T. (2023). The impact of social media sentiment on stock market based on user classification. In Digitalization and Management Innovation, pp. 1-16.

https://doi.org/10.3233/FAIA230002

[5] Rosa, L. (2023). Analyzing the effect of social media sentiment on stock prices. International Journal of Business and Management, 17(9): 20-27. https://doi.org/10.5539/ijbm.v17n9p20

[6] Fan, R., Talavera, O., Tran, V. (2023). Social media and price discovery: The case of cross-listed firms. Journal of Financial Research, 46(1): 151-167. https://doi.org/10.1111/jfir.12310

[7] Fan, R., Talavera, O., Tran, V. (2020). Social media, political uncertainty, and stock markets. Review of Quantitative Finance and Accounting, 55: 1137-1153. https://doi.org/10.1007/s11156-020-00870-4

[8] Gan, B., Alexeev, V., Bird, R., Yeung, D. (2020). Sensitivity to sentiment: News vs social media. International Review of Financial Analysis, 67: 101390. https://doi.org/10.1016/j.irfa.2019.101390

[9] Rautiainen, A., Jokinen, J. (2022). The value-relevance of social media activity of Finnish listed companies. International Journal of Accounting & Information Management, 30(2): 301-323. https://doi.org/10.1108/IJAIM-04-2021-0076

[10] Al Ridhawi, M., Al Osman, H. (2023). Stock market prediction from sentiment and financial stock data using machine learning. In the 36th Canadian Conference on Artificial Intelligence, Montréal, pp. 1-6.

[11] Yang, N. (2022). Social media and money: How our tweets affect the financial markets. Rangahau Aranga: AUT Graduate Review, 1(1). https://doi.org/10.24135/rangahau-aranga.v1i1.34

[12] Singh, S., Kaur, A. (2022). Twitter sentiment analysis for stock prediction. In 2nd International Conference on "Advancement in Electronics & Communication Engineering (AECE 2022), pp. 674-677.

[13] Shah, D., Isah, H., Zulkernine, F. (2018). Predicting the effects of news sentiments on the stock market. In 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, pp. 4705-4708. https://doi.org/10.1109/BigData.2018.8621884

[14] Chen, R., Dong, R. (2023). The relationship between twitter sentiment and stock performance: A decision tree approach. In Hawaii International Conference on System Sciences 2023 (HICSS-56), pp. 4850-4859.

[15] You, Z. (2022). Evaluation of two models for predicting Amazon stock based on machine learning. BCP Business & Management, 34: 39-47. https://doi.org/10.54691/bcpbm.v34i.2862

[16] Abhishek, K., Khairwa, A., Pratap, T., Prakash, S. (2012). A stock market prediction model using Artificial Neural Network. In 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), Coimbatore, India, pp. 1-5. https://doi.org/10.1109/ICCCNT.2012.6396089

[17] Guresen, E., Kayakutlu, G., Daim, T.U. (2011). Using artificial neural network models in stock market index prediction. Expert systems with Applications, 38(8): 10389-10397. https://doi.org/10.1016/j.eswa.2011.02.068

[18] Alva, A.I., Rodríguez, C. (2024). Assessing offshore oil spills using remote sensing and geospatial artificial intelligence (GeoAI): A systematic literature review. Ingénierie des Systèmes d'Information, 29(6): 2105-2113. https://doi.org/10.18280/isi.290602

[19] Shi, Y. (2023). Research on the stock price prediction using machine learning. Advances in Economics, Management and Political Sciences, 22: 174-179. https://doi.org/10.54254/2754-1169/22/20230307

[20] Bin Khamis, A., Yee, P.H. (2018). A hybrid model of artificial neural network and genetic algorithm in forecasting gold price. European Journal of Engineering and Technology Research, 3(6): 10-14. https://doi.org/10.24018/ejers.2018.3.6.758

[21] Chong, W.U., Peng, L.U.O., Yongli, L.I. (2015). Stock price forecasting: hybrid model of artificial intelligent methods. Engineering Economics, 26(1): 40-48. https://doi.org/10.5755/j01.ee.26.1.3836

[22] Pangaribuan, J.J., Thames, J., Barus, O.P. (2024). Comparative analysis of classification of dry nut types using the support vector machine and linear discriminant analysis methods. Ingénierie des Systèmes d'Information, 29(6): 2231-2242. https://doi.org/10.18280/isi.290613

[23] Zhang, J., Teng, Y.F., Chen, W. (2019). Support vector regression with modified firefly algorithm for stock price forecasting. Applied Intelligence, 49: 1658-1674. https://doi.org/10.1007/s10489-018-1351-7

[24] Vijh, M., Chandola, D., Tikkiwal, V.A., Kumar, A. (2020). Stock closing price prediction using machine learning techniques. Procedia Computer Science, 167: 599-606. https://doi.org/10.1016/j.procs.2020.03.326

[25] Mokhtari, S., Yen, K.K., Liu, J. (2021). Effectiveness of artificial intelligence in stock market prediction based on machine learning. arXiv preprint arXiv:2107.01031. https://doi.org/10.5120/ijca2021921347

[26] Liu, H., Qi, L., Sun, M. (2022). Short-term stock price prediction based on CAE-LSTM method. Wireless Communications and Mobile Computing, 2022(1): 4809632. https://doi.org/10.1155/2022/4809632

[27] Kamiel, B.P., Saputri, A.D., Muizza, Z.H., Yobioktabera, A. (2024). Smart harvest: Web-integrated ripeness detection for apples with CNN algorithm. Ingénierie des Systèmes d'Information, 29(6): 2181-2190. https://doi.org/10.18280/isi.290608

[28] Lee, J., Kim, R., Koh, Y., Kang, J. (2019). Global stock market prediction based on stock chart images using deep Q-network. IEEE Access, 7: 167260-167277. https://doi.org/10.1109/ACCESS.2019.2953542

[29] Mehtab, S., Sen, J. (2019). A robust predictive model for stock price prediction using deep learning and natural language processing. arXiv preprint arXiv:1912.07700. https://doi.org/10.36227/techrxiv.15023361.v1

[30] Abayomi-Alli, A., Abayomi-Alli, O., Misra, S., Fernandez-Sanz, L. (2022). Study of the Yahoo-Yahoo Hash-Tag tweets using sentiment analysis and opinion mining algorithms. Information, 13(3): 152. https://doi.org/10.3390/info13030152

[31] Li, A.W., Bastos, G.S. (2020). Stock market forecasting using deep learning and technical analysis: A systematic review. IEEE Access, 8: 185232-185242. https://doi.org/10.1109/ACCESS.2020.3030226

[32] Nguyen, T.H., Shirai, K., Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, 42(24): 9603-9611. https://doi.org/10.1016/j.eswa.2015.07.052

[33] Olotu, S.I. (2023). A multivariate LSTM-based deep learning model for stock market prediction. Applied and Computational Engineering, 2: 965-973. https://doi.org/10.54254/2755-2721/2/20220602

[34] Roy, P.P., Rao, S., Zhu, M. (2022). Mandatory CSR expenditure and stock market liquidity. Journal of Corporate Finance, 72: 102158. https://doi.org/10.1016/j.jcorpfin.2022.102158

[35] Maciejewski, M.L. (2020). Quasi-experimental design. Biostatistics & Epidemiology, 4(1): 38-47. https://doi.org/10.1080/24709360.2018.1477468

[36] Pagolu, V.S., Reddy, K.N., Panda, G., Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. In 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, India, pp. 1345-1350. https://doi.org/10.1109/SCOPES.2016.7955659

[37] Charoenrook, A., Daouk, H. (2014). Conditional skewness of aggregate market returns: Evidence from developed and emerging markets. Global Economy and Finance Journal, 7(1): 96-121.

[38] Engle, R., Mistry, A. (2014). Priced risk and asymmetric volatility in the cross section of skewness. Journal of Econometrics, 182(1): 135-144. https://doi.org/10.1016/j.jeconom.2014.04.013

[39] Wang, Z. (2023). Investor sentiment analysis based on comment text for stock price prediction. BCP Business & Management, 38: 2710-2716. https://doi.org/10.54691/bcpbm.v38i.4178

[40] Gumus, A., Sakar, C.O. (2021). Stock market prediction by combining stock price information and sentiment analysis. International Journal of Advances in Engineering and Pure Sciences, 33(1): 18-27. https://doi.org/10.7240/jeps.683952

[41] Siganos, A., Vagenas-Nanos, E., Verwijmeren, P. (2017). Divergence of sentiment and stock market trading. Journal of Banking & Finance, 78: 130-141. https://doi.org/10.1016/j.jbankfin.2017.02.005

[42] Guo, Q. (2023). The relationship between investor sentiment and stock market price. Frontiers in Business, Economics and Management, 9(2): 124-129. https://doi.org/10.54097/fbem.v9i2.9139

[43] Farimani, S.A., Jahan, M.V., Milani Fard, A. (2022). From text representation to financial market prediction: A literature review. Information, 13(10): 466. https://doi.org/10.3390/info13100466

[44] Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L.T., Trajanov, D. (2020). Evaluation of sentiment analysis in finance: from lexicons to transformers. IEEE Access, 8: 131662-131682. https://doi.org/10.1109/ACCESS.2020.3009626

[45] Razouk, A., Falloul, M.E.M., Harkati, A., Touhami, F. (2023). Performance evaluation of technical indicators for forecasting the Moroccan stock index using deep learning. Indonesian Journal of Electrical Engineering and Computer Science, 32(3): 1785-1794.

https://doi.org/10.11591/ijeecs.v32.i3.pp1785-1794

[46] Agarwal, B. (2023). Financial sentiment analysis model utilizing knowledge-base and domain-specific representation. Multimedia Tools and Applications, 82(6): 8899-8920. https://doi.org/10.1007/s11042-022-12181-y

[47] Das, N., Sadhukhan, B., Chatterjee, R., Chakrabarti, S. (2024). Integrating sentiment analysis with graph neural networks for enhanced stock prediction: A comprehensive survey. Decision Analytics Journal, 10: 100417. https://doi.org/10.1016/j.dajour.2024.100417

[48] Sarkar, A., Chakraborty, S., Ghosh, S., Naskar, S.K. (2022). Evaluating impact of social media posts by executives on stock prices. In Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation, Kolkata, India, pp. 74-82. https://doi.org/10.1145/3574318.3574339

[49] Kim, J.N., Park, H.J., Kim, M.S., Choi, Y.J., Kim, E., Park, J.H., Hong, S.W. (2023). Sonographic and elastographic findings after arthroscopic rotator cuff repair: A comparison with clinical results. The British Journal of Radiology, 96(1143): 20210777. https://doi.org/10.1259/bjr.20210777

[50] Johnson, Z.L., Ammirati, M., Wasilko, D.J., Chang, J.S., et al. (2023). Structural basis of the acyl-transfer mechanism of human GPAT1.Nature Structural & Molecular Biology, 30(1): 22-30. https://doi.org/10.1038/s41594-022-00884-7

[51] Miller, A., York, E.M., Stopka, S.A., Martínez-François, J.R., et al. (2023). Spatially resolved metabolomics and isotope tracing reveal dynamic metabolic responses of dentate granule neurons with acute stimulation. Nature Metabolism, 5(10): 1820-1835. https://doi.org/10.1038/s42255-023-00890-z

[52] Zheng, X. (2023). Stock price prediction based on CNN-BiLSTM utilizing sentiment analysis and a two-layer attention mechanism. Advances in Economics, Management and Political Sciences, 47: 40-49. https://doi.org/10.54254/2754-1169/47/20230369

[53] Ho, T.T., Huang, Y. (2021). Stock price movement prediction using sentiment analysis and CandleStick chart representation. Sensors, 21(23): 7957. https://doi.org/10.3390/s21237957

[54] Chen, Y., Zhu, S., He, H. (2021). The influence of investor emotion on the stock market: Evidence from an infectious disease model. Discrete Dynamics in Nature and Society, 2021(1): 5520276. https://doi.org/10.1155/2021/5520276

[55] Vanstone, B.J., Gepp, A., Harris, G. (2019). Do news and sentiment play a role in stock price prediction? Applied Intelligence, 49(11): 3815-3820. https://doi.org/10.1007/s10489-019-01458-9