E-ISSN : 2963-4946
Vol. 1 No. 12 September 2023

Asian Journal of Social and Humanities
https://ajosh.org/

Utilization of LSTM (Long Short Term Memory) Based Sentiment Analysis for Stock Price Prediction

Muhammad Fajrul Aslim[1], Gerry Firmansyah[2], Habibullah Akbar[3], Budi Tjahyono[4], Agung Mulyo Widodo[5]
[1,2,3,4,5] Universitas Esa Unggul, Indonesia
Email: fajrul.aslim@gmail.com, gerry@esaunggul.ac.id, habibullah.akbar@esaunggul.ac.id, budi.tjahjono@esaunggul.ac.id, agung.mulyo@esaunggul.ac.id

* Correspondence: gerry@esaunggul.ac.id

KEYWORDS
Deep Learning, LSTM, Stock Price, Sentiment Analysis

ABSTRACT
This study aims to utilize sentiment analysis in predicting stock price movements. Sentiment analysis can provide information to investors to understand market sentiment. This study uses a text-based approach by pre-processing data, constructing a sentiment analysis model and evaluating model performance. The collected data is analyzed to identify the text's positive, negative, or neutral sentiments. The approach used in scoring sentiment analysis is the Text blob approach and the Lexicon approach. Differences in the results of the accuracy of the two Sentiment Analysis approaches with the LSTM model have an influence on the prediction results with a better increase in accuracy using the Lexicon Sentiment Analysis approach. Then the LSTM model is implemented to classify texts into the desired sentiment categories. The results of this study are insight into the use of sentiment analysis in predicting stock price movements. The implemented sentiment analysis model can be a useful predictive tool for investors and stock practitioners in making investment decisions.

Introduction

In the financial world, stocks have become one of the investment instruments more accessible to the public. However, the complex and often unpredictable fluctuations in stock prices have created challenges for investors and market analysts. Stock price prediction is crucial in the investment world as it aids market participants in making informed investment decisions. Patterns formed in the past and the belief that trading ideas will behave similarly as before are the fundamentals of technical analysis, which relates to the economic and social forces that can cause current market price deviations (V. Sharma et al., n.d.) (S. K. Sharma, 2021). Investors, financial analysts, and companies require accurate information to identify market trends, manage risks, and optimize investment portfolios (Jariwala et al., 2020).

Fundamental and technical factors in stock price prediction are undeniably essential for investors. Fundamental factors are based on a company's economic and financial aspects, and they can be unpredi (Liu et al., 2018). Technical analysis, on the other hand, relies on charts and complete historical data on stock prices, making stock price prediction crucial for investors' future investment decisions (Sarkar et al., 2020). Various stock price prediction techniques have been developed to address this issue, and one of the central focuses is Long Short-Term Memory (LSTM). LSTM is an artificial intelligence algorithm capable of recognizing long-term patterns in data and producing accurate predictions (Institute of Electrical and Electronics Engineers & IEEE Industrial Electronics Society, n.d.).

On the other hand, the development of internet technology has significantly impacted society's ability to interact and express their opinions. Nowadays, people have easier access to convey their opinions, views, and feelings through online platforms. News and information are disseminated to the public more easily and quickly. Sentiment analysis, which involves understanding the positive, negative, or neutral sentiments contained in text, has become increasingly important in analyzing how public opinions or news can influence stock prices (Guo, 2020). Therefore, sentiment analysis and stock price prediction research has become an intriguing and relevant topic in finance and information technology.

However, despite many separate studies in sentiment analysis and stock price prediction, there have been few studies that combine both areas. Integrating sentiment analysis into prediction algorithms like LSTM can provide deeper insights into how public sentiment can influence stock market behavior (IEEE Staff, 2016). Stock sentiment analysis poses several unique challenges (IEEE Staff, 2017). For example, market uncertainty, stock price volatility, and the influence of other factors like fake news or rumors can affect sentiment interpretation and stock price predictions (Aasi et al., 2021). Therefore, the goal of this research is to bridge this gap by analyzing the potential use of sentiment analysis as an additional factor in stock price prediction using LSTM algorithms (Birjali et al., 2021). LSTM has the ability to understand sentiment data in a long-term context and can handle issues related to sequential data, such as text, using specialized memory units to control information flow and address unstable gradient problems, especially when dealing with long sequences—a critical issue in sentiment analysis approaches for predicting stock price movements (Han et al., 2021).

Research Methods

In this study, a framework based on Method Improvement developed by Berndtsson et al. (2008), Dawson (2009), and Polancik (2007) was used. The independent variables studied include Sentiment, Data, Stock Price, and Prediction. While the bound variables used in prediction models are MAPE, RMSE, and RSquared.

The stock data used comes from PT GoTo Gojek Tokopedia Tbk with stock code GOTO, from January 2023 to June 2023. Sentiment data is obtained from online news portal CNBC Indonesia as it provides relevant economic and business information. The sentiment data of the subject must first be processed to clear it of noise. This process includes cleaning, tokenization, filtering, and stemming.

Polarity Calculation and Sentiment Classification, Each data from CNBC Indonesia is analyzed to determine its sentiment using a lexicon-based approach with Lexicon Inset. Each word in the sentence will be matched to the word in the lexicon, and the polarity score will be calculated for each sentence. Sentiment results are classified as positive, negative, or neutral (Shayaa et al., 2018).

The results of sentiment analysis are combined with a stock dataset by date, and the polarity score obtained is calculated to average for each date. Stock price predictions are made using the LSTM model without considering sentiment. Stock price predictions are carried out again using the LSTM model by including sentiment results as one of the factors.

LSTM models are evaluated using testing data using metrics such as MAPE, RMSE, and RSquared to measure prediction error and model accuracy. The evaluation results are analyzed and presented in visual or tabular form for easy interpretation. This research will end by concluding the results of the study and providing suggestions for further research. This method is used to examine the effect of sentiment on stock price predictions using the LSTM model (Cheng et al., 2019) (Kong et al., 2019) (Rafi et al., 2021).

Results and Discussions

4.1 Data Collection

4.1.1 Data CNBC Indonesia

The sentiment data used in this study is data crawled on websites https://www.cnbcindonesia.com/ with the search keyword "GOTO" and the time period January 2023 to June 2023. After doing so on the search form, data was obtained for a total of 411 news.

The process of crawling data is done using Python Programming Language (see attachment). The following is the process of crawling tweet data to retrieve sentiment tweet data and preprocessing steps, namely:

1. The process of crawling data using Python Programming Language with the help of the plugin BeautifulSoup
2. The keyword used for the search is "GOTO"
3. Search period from January 2023 to June 2023
4. The total data obtained is 411 data in the form of news.

4.1.2 Stock Data

The stock data used is GOTO companies in the period January 2023 to June 2023. The dataset was obtained from https://finance.yahoo.com website and downloaded in csv file format. The following is a visualization of GOTO stock data that can be obtained if displayed in graphic form.

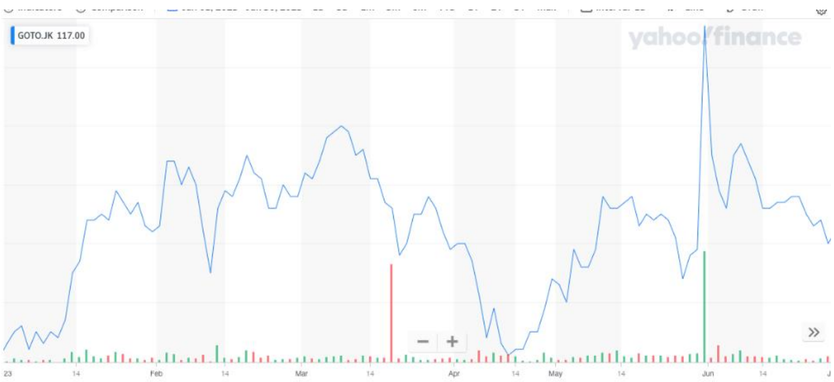

Figure 1. GOTO Shares for January 2023- June 2023 Period

4.2 Preprocessing Sentiment Data

News data obtained from CNBC is still raw data, so it is necessary to preprocess the data to obtain clean and structured data so that it can be used for sentiment analysis.

4.2.1 Cleaning Process

The cleaning process is carried out with the aim of cleaning news data from characters or elements that are not needed so that noise in the classification process will be reduced. The elements that are removed are hashtags, URL links, symbols, numbers, and punctuation, multiple spaces, and changing all words to lowercase. The process of removing punctuation, case folding and spelling correction is in the following table.

Tabel 1 Cleansing Data

Before	**After**
For information, last year the collection of funds reached Rp 260 trillion. Thus, the fundraising this year is targeted at Rp 170 trillion." The target for 2023 is IDR 170 trillion and compared to 2022, it is extraordinary to reach IDR 260 trillion. If we remove the GOTO et al layer, there will still be positive growth, but we will balance it between 2022 and 2023," explained Inarno.	To note, last year the fundraising reached IDR 260 trillion, thus this year's fund raising is targeted at IDR 170 trillion, the target for 2023 is IDR 170 trillion, and compared to 2022, it is extraordinary to reach IDR 260 trillion, if you release out the GoTo et al layer, there will still be positive growth, but think we are balanced between 2022 and 2023, explained Inarno

1. Tokenizing Process

Tokenizing is done to break down strings in news sentences into units of words that compose them. Basically this process does the hyphenation of sentences into words. The following process is described in the following table.

Table 2 Results of the Tokenizing Process

Before	**After**
To note, last year the fundraising reached IDR 260 trillion, thus this year's fund raising is targeted at IDR 170 trillion, the target for 2023 is IDR 170 trillion, and compared to 2022, it is extraordinary to reach IDR 260 trillion, if you release out the GoTo et al layer, there will still be positive growth, but think we are balanced between 2022 and 2023, explained Inarno	'for', 'known', 'year', 'past', 'collection', 'fund', 'achieve', 'rp', '260', 'trillion', 'with', 'thus', 'collection', 'fund', 'on', 'year', 'this', 'targeted', 'rp', '170', 'trillion', 'target', 'for', '2023', 'rp', '170', 'trillion', 'and', 'if', 'compared', '2022', 'indeed', 'extraordinary', 'achieve', 'rp', '260', 'trillion', 'apaka', 'release', 'out', 'layer', 'goto', 'et al', 'remain', 'ada', 'growth', 'positive', 'but', 'kira', 'we', 'balanced', 'between', '2022', 'and', '2023', 'clear', 'inarno'

2. Filtering Process

In the filtering process , important words are taken from the results of the previous process. Stopwords or words that lack meaning will be eliminated because they are not needed for sentiment analysis. Here are the results of the filtering process in the following table.

Table 3 Filtering Process Results

Before	After
'for', 'known', 'year', 'past', 'collection', 'fund', 'achieve', 'rp', '260', 'trillion', 'with', 'thus', 'collection', 'fund', 'on', 'year', 'this', 'targeted', 'rp', '170', 'trillion', 'target', 'for', '2023', 'rp', '170', 'trillion', 'and', 'if', 'compared', '2022', 'indeed', 'extraordinary', 'achieve', 'rp', '260', 'trillion', 'apaka', 'release', 'out', 'layer', 'goto', 'et al', 'remain', 'ada', 'growth', 'positive', 'but', 'kira', 'we', 'balanced', 'between', '2022', 'and', '2023', 'clear', 'inarno'	'Collection', 'Fund', 'Targeted', 'RP', '170', 'Trillion', 'Target', '2023', 'RP', '170', 'Trillion', 'Compared', '2022', 'Extraordinary', 'Achieve', 'RP', '260', 'Trillion', 'Release', 'Out', 'Layer', 'GoTo', 'et al', 'Growth', 'Positive', 'Balanced', '2022', '2023', 'Inarno'

3. Stemming Process

This process converts words into root words by removing word affixes in the form of prefixes and suffixes.

Table 4 Stemming Process Result

Before	After
'Collection', 'Fund', 'Targeted', 'RP', '170', 'Trillion', 'Target', '2023', 'RP', '170', 'Trillion', 'Compared', '2022', 'Extraordinary', 'Achieve', 'RP', '260', 'Trillion', 'Release', 'Out', 'Layer', 'GoTo', 'et al', 'Growth', 'Positive', 'Balanced', '2022', '2023', 'Inarno'	'Collection', 'Fund', 'Targeted', 'RP', '170', 'Trillion', 'Target', '2023', 'RP', '170', 'Trillion', 'Compared', '2022', 'Extraordinary', 'Achieve', 'RP', '260', 'Trillion', 'Release', 'Out', 'Layer', 'GoTo', 'et al', 'Growth', 'Positive', 'Balanced', '2022', '2023', 'Inarno',

4.3 Polarity Calculation and Sentiment Classification

The approach taken to examine polarity and sentiment classification is to use the Lexicon Inset Method and the NLP Textblob Method.

These two approaches have differences in processing sentiment, the Lexicon Inset method uses words while the Textblob NLP method uses sentences. The following describes the process of each approach.

4.3.1 Inset Lexicon Method

The approach process in the lexicon method for calculating polarity is to score each data according to the date. The table below describes the process of the data that is processed by scoring sentiment

Table 5 Lexicon Sentiment Score Results

Date	Title	Score

		Results
2023-01-02	OJK Targets 58 IPOs and IDR 170 T Funds from the Capital Market	75
2023-01-02	2022 Chaotic, Tech Sector Suffering Will Continue?	133
2023-01-02	JCI Doesn't Collapse Thanks to These 3 Stocks	5
2023-01-03	GOTO has been the God of Helping JCI for 2 days, what sign?	-9
2023-01-03	These 8 Stocks Make JCI Soar, Already Have It?	16

After scoring each word studied, the next process is to label the sentiment from the sentiment score data, before labeling, the data is averaged per date so that the results of grouping the senitmen by date are obtained, namely positive greater than 0, negative less than 0, and other than neutral. The following example of the process is described in the table

Table 6 Lexicon Sentiment Classification

Date	Result Score	Sentiment
2023-01-02	71,00	Positive
2023-01-03	3,50	Positive

4.3.2 NLP Textblob Method

The approach process in the lexicon method for calculating polarity is to score each data according to the date. The table below describes the process of the data that is processed by scoring sentiment

Table 7 Textblob Sentiment Score Results

Date	Title	Score Result
2023-01-02	OJK Targets 58 IPOs and IDR 170 T Funds from the Capital Market	0,121717
2023-01-02	2022 Chaotic, Will the Suffering of the Technology Sector Continue?	-0,007783
2023-01-02	JCI Doesn't Collapse Thanks to These 3 Stocks	0,127083
2023-01-03	GOTO has been the God of Helping JCI for 2 days, what sign?	0,005
2023-01-03	These 8 Stocks Make JCI Soar, Already Have It?	0,273611

After scoring each word studied, the next process is to label the sentiment from the sentiment score data, before labeling, the data is averaged per date so that the results of grouping the senitmen by date are obtained, namely positive greater than 0, negative less than 0, and other than neutral. The following example of the process is described in the table

Table 8. Textblob Sentiment Classification Results

Date	Score Results	Sentiment
2023-01-02	0,080339	Positive
2023-01-03	0,1393055	Positive

In this approach, the processing process is returned to a sentence. Then translated from English using translator to Indonesian, so the need for a translation process because this approach supports English

4.4 Combining Sentiment Results with Stock Prices

The following describes the process of combining sentiment results with stock prices to be predicted. The stock data used has indicators of date, opening stock price, highest stock price, lowest stock price, closing stock price, adjusted stock price, and number of transactions.

Table 9. Combining Lexicon Sentiment Results with Stock Prices

Date	Open	High	Low	Close	Adj Close	Volume	Sentiment
2023-01-02	91	94	91	93	93	897258100	71,00
2023-01-03	93	97	93	95	95	2761729900	3,50
2023-01-04	95	97	93	96	96	1816437300	31,00
2023-01-05	96	96	91	92	92	1795026400	0,00
2023-01-06	92	95	91	95	95	802624500	46,67

Table 10 Combining Textblob Sentiment Results with Stock Prices

Date	Open	High	Low	Close	Adj Close	Volume	Sentiment
2023-01-02	91	94	91	93	93	897258100	0,08
2023-01-03	93	97	93	95	95	2761729900	0,14
2023-01-04	95	97	93	96	96	1816437300	0,02
2023-01-05	96	96	91	92	92	1795026400	0,00
2023-01-06	92	95	91	95	95	802624500	0,07

4.5 Learning Model Outcomes (LSTM) and Analysis

The use of the LSTM Learning Model is carried out with 3 different indicators, but carried out simultaneously. These indicators are

1. Stock Price Prediction without Sentiment
2. Stock Price Prediction using Lexicon Sentiment
3. Stock Price Prediction using Textblob Sentiment

The stock prediction steps with LSTM are as follows:

- Normalize unused fields (Date)
- Save data before scaling (improve the stability of model performance training) and convert to python libs used (numpy, pandas) and scale by dividing the data range 0 and 1 from the minimum and maximum data to be tested.
- Divide the data into 2, namely training data and testing data (training data:

70%, testing data: 30%)
- Create a data function into 2 dimensions
- Determine the number of timesteps
- Create an LSTM model
- Optimization using Adam

The following are the results of the three indicators that were done with the LSTM model.

Table 11. LSTM Model Results

#	Epoch	Batch	Without Sentiment			With Lexicon Sentiment			With Textblob Sentiment		
			MAPE	RMSE	R Square	MAPE	RMSE	R Square	MAPE	RMSE	R Square
	20	8	0,1797	0,1010	23,62%	0,1783	0,1002	24,80%	0,1781	0,0996	25,70%
	50	8	0,2068	0,1074	13,59%	0,2064	0,1070	14,30%	0,2056	0,1074	13,56%
	80	8	0,2108	0,1088	11,34%	0,2098	0,1086	11,57%	0,2107	0,1082	12,29%
	100	8	0,2098	0,1085	11,75%	0,2096	0,1082	12,27%	0,2101	0,1089	11,18%
	150	8	0,2092	0,1089	11,11%	0,2090	0,1086	11,56%	0,2092	0,1081	12,37%
	20	16	0,1668	0,0982	27,74%	0,1644	0,1003	24,58%	0,1652	0,0997	25,51%
	50	16	0,1826	0,0993	26,10%	0,1838	0,0992	26,31%	0,1861	0,1002	24,83%
	80	16	0,2026	0,1057	16,33%	0,2013	0,1048	17,79%	0,2032	0,1056	16,53%
	100	16	0,2087	0,1081	12,53%	0,2091	0,1084	11,98%	0,2073	0,1079	12,77%
	150	16	0,2096	0,1086	11,68%	0,2099	0,1082	12,26%	0,2097	0,1092	10,65%
	20	24	0,1933	0,1385	-43,82%	0,1862	0,1317	-30,01%	0,1841	0,1303	-30,01%
	50	24	0,1745	0,1024	21,41%	0,1754	0,0994	25,95%	0,1779	0,0987	26,99%
	80	24	0,1954	0,1025	21,34%	0,1951	0,1022	21,72%	0,1939	0,1027	20,91%
	100	24	0,2032	0,1051	17,30%	0,2055	0,1053	16,88%	0,2015	0,1060	15,83%
	150	24	0,2077	0,1086	11,66%	0,2087	0,1096	10,01%	0,2108	0,1084	11,97%
	20	32	0,2517	0,1869	-161,66%	0,3108	0,2290	-293,04%	0,3281	0,2402	-332,17%
	50	32	0,1732	0,1014	22,95%	0,1693	0,1024	21,50%	0,1715	0,0991	26,38%
	80	32	0,1820	0,1007	23,99%	0,1827	0,0995	25,79%	0,1855	0,0997	25,56%
	100	32	0,1931	0,1023	21,63%	0,1916	0,1020	22,10%	0,1906	0,1026	21,10%
	150	32	0,2068	0,1076	13,31%	0,2067	0,1072	13,95%	0,2073	0,1074	13,55%
	20	48	0,5713	0,3909	-1044,68%	0,5221	0,3614	-878,40%	0,5489	0,3775	-967,76%
	50	48	0,1667	0,0982	27,80%	0,1660	0,0988	26,89%	0,1657	0,0984	27,48%
	80	48	0,1703	0,1007	23,97%	0,1746	0,0999	25,23%	0,1726	0,1002	24,75%
	100	48	0,1740	0,0995	25,87%	0,1796	0,0987	27,02%	0,1801	0,0997	25,60%

150	48	0,1918	0,1012	23,25%	0,1917	0,1021	21,92%	0,1881	0,1009	23,73%

In the results of prediction experiments using the LSTM model, it is known that the indicators in the model testing process are:

1. Epoch = (20,50,80,100,150)
2. Batch = (8,16,24,32,48)

It is known that the best results were at epoch 20 and batch 16 of all three approaches. Next is to experiment 10 times at epoch 20 and batch 16 to see which approach is better. Here are the results.

Table 12 Results of 10 Epoch 20 and Batch 16 Experiments

#	Epoch	Batch	Without Sentiment			With Lexicon Sentiment			With Textblob Sentiment		
			MAPE	RMSE	R Square	MAPE	RMSE	R Square	MAPE	RMSE	R Square
1	20	16	0,1659	0,0994	25,96%	0,1680	0,0980	28,05%	0,1632	0,1024	21,45%
2	20	16	0,1662	0,0990	26,54%	0,1653	0,0992	26,23%	0,1641	0,0995	25,87%
3	20	16	0,1648	0,0989	26,76%	0,1646	0,1002	24,73%	0,1662	0,0988	26,87%
4	20	16	0,1687	0,0981	27,96%	0,1683	0,0980	28,07%	0,1655	0,0992	26,25%
5	20	16	0,1653	0,0990	26,60%	0,1661	0,0989	26,73%	0,1672	0,0970	28,32%
6	20	16	0,1668	0,0982	27,74%	0,1644	0,1003	24,58%	0,1652	0,0997	25,51%
7	20	16	0,1654	0,0987	26,97%	0,1644	0,0998	25,19%	0,1631	0,1015	22,81%
8	20	16	0,1629	0,1023	21,58%	0,1664	0,0992	26,33%	0,1668	0,0986	27,16%
9	20	16	0,1662	0,0992	26,32%	0,1644	0,0996	25,13%	0,1658	0,0995	25,79%
10	20	16	0,1693	0,0978	28,36%	0,1684	0,0973	29,02%	0,1672	0,0983	27,58%
AVERAGE			0,1662	0,0991	26,48%	0,1660	0,0991	26,41%	0,1654	0,0995	25,76%

After 10 experiments, the average calculation of each evaluation was carried out, and the following results were obtained:

1. Predictions with textblob sentiment have better MAPE results than MAPE predictions with lexicon sentiment with an evaluation value difference of 0.006. While predictions without sentiment have the lowest MAPE results.
2. A prediction with lexicon sentiment has the same RMSE result as a non-sentiment RMSE prediction with a difference of 0 values. Predictions with textblob sentiment have the lowest RMSE results.
3. Predictions without sentiment have better RSquare results than RSquare predictions with lexicon sentiment by a difference of 0.07%. While predictions with textblob sentiment have the lowest RSquare results.

Table 13. Comparison of Evaluation Metrics

Method	MAPE	RMSE	RSquare
Without Sentiment	0,1662	0,0991	26,48%
With Lexicon Sentiment	0,1660	0,0991	26,41%
With Textblob Sentiment	0,1654	0,0995	25,76%

The results of the tested model are depicted in the following graphic modeling :

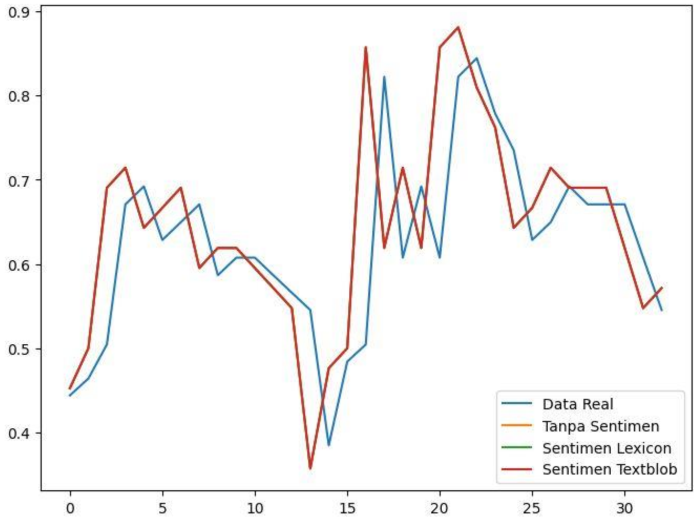

Figure 2. LSTM Model Research Results

Here's a comparison visualization of Evaluation Metrics illustrating RMSE, MAPE, and RSquared results.

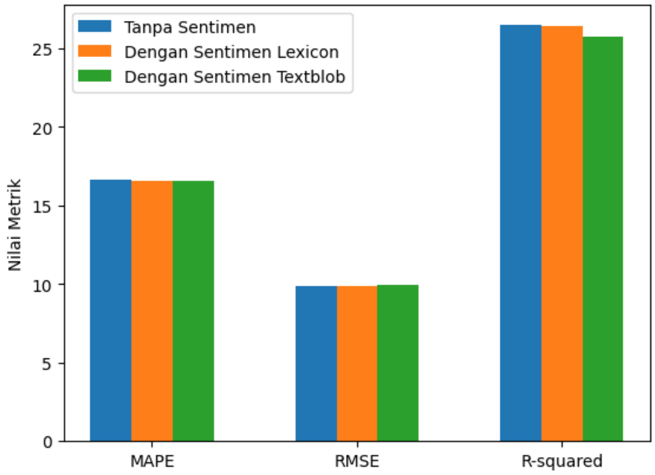

Figure 3. Evaluation Matrix Comparison

The addition of sentiment analysis indicators to stock predictions has an influence on RMSE evaluation results, MAPE being slightly better. However, it does not have a better influence on the results of RSquare evaluation.

4.6 Evaluation

In this study, there are several results of the author's analysis that become material for future research and limitations and constraints in current research.

4.6.1 Stock

1. The stock data used in this study is a determinant for the research objectives set at the beginning, so it is important to adjust to the time of research. So this study focuses on 1 LQ45 company data, namely GOTO.
2. Stock data used can be from various types of companies and more than one company to get more satisfactory results
3. Stock data can be obtained from other references other than yahoofinance and combined with other media if the research time is in accordance with

the number of data references taken and tested later

4.6.2 Sentiment

1. Sentiment data is crawled from the CNBC Indonesia news portal platform on the grounds that the data obtained can be more appropriate and accurate with the desired keywords. That is to discuss stocks, markets, market conditions, and companies that want to be researched.
2. The determination of the sentiment classification used depends on the type of research taken, some can only have 2 negative and positive indicators. However, for this study using 3 classifications, namely positive, negative, neutral because the classification results will be combined with stock data per date, and there is a possibility that there are certain dates that do not have data to be classified so that they will automatically be given zero or neutral indicators.
3. If you want a more accurate sentiment, you can use experts in accordance with the review language of the research to be studied, so that the results can be more accurate along with the majas and figurative words used in sentiment reviews (ICIEV.2017.8338584, n.d.).

4.6.3 Production Results

1. In testing the LSTM model to determine stock price predictions and their effects on incoming sentiment is in accordance with the classification of sentiment polarity results, so it is known that positive sentiment affects stock price increases, vice versa negative sentiment affects stock price declines, and neutral sentiment reviews are classified as many from the results of this study
2. If you want accurate prediction results with the test metrics used, then make sure the sentiment data used and the sentiment categorization are appropriate and objective
3. When viewed from the evaluation results, the use of sentiment has a slightly better influence than predictions without using sentiment. However, because the difference is relatively small, it can be said in this study that predictions without sentiment are more efficient in their application. Further research is needed to improve prediction results using sentiment.
4. The results of observations made by researchers by observing charts of stock price movements in https://yahoo.finance.com, there are visualizations of increases and decreases in stock prices, from this data researchers observe public sentiment in https://cnbcindonesia.com the same time period as April-June 2023

And the sentiment results studied in that time period have a negative score when the stock price declines, and a positive score when stock price increases. However, not always the condition of public sentiment can affect stock price fluctuations, because there are several periods of time where there is a slight decline in stock prices and public sentiment obtained from the CNBC portal has a positive or neutral score.

In the following figure, in the May-June period, we know the highest price of the stock on May 31, 2023. And after testing the sentiment data obtained, the sentiment analysis results were positive.

Figure 4. Observation of Stock Price Increase and Decrease and Sentiment Relationship

4.6.4 Challenges and Obstacles

1. Sentiment Data Retrieval

The process of crawling data on CNBC Indonesia using Python is a little difficult, here are the researchers explaining and how researchers overcome these obstacles.

- There are some news data that cannot be crawled for reasons that researchers do not yet know. To overcome this, researchers analyze which data cannot be obtained, then crawl the data manually one by one on the official website.
- The search query applied to the CNBC Indonesia portal does not use a time range from when to when, only applies one specific date if you want to search by date. This is the effect of the first problem. To overcome this, there are two approaches that can be done. First, check the CNBC website and do a manual search to find out the data we want to be on any page. Second, check one by one by date

2. Stock Data

The stock data used will be data from January to June 2023 in one research object, namely the GOTO company. After testing stock predictions with this time frame, researchers can say that the prediction results are less than optimal. At least a longer range of stock data is needed, for example up to 2 years. But this cannot be done because it is constrained by obstacles / problems in retrieving Twitter data that have been explained by researchers in the previous discussion. So that the range of shares used is 6 months, because this does not affect the purpose of the study.

3. Sentiment Analysis

There are 2 approaches used in sentiment analysis, namely the Lexicon Inset approach and the NLP approach with Textblob. The difference is:

a. Lexicon
   - Analyze sentiment using Indonesian
   - The text analyzed is per word

b. Textblob
   - Analyze every sentence directly
   - The analyzed text must be in English. So that the sentiment data

obtained must first be translated into English which makes the sentiment processing estimation longer

Broadly speaking, both have different sentiment analysis results and the results of the study show that lexicon is better than textblob in terms of RMSE evaluation, and RSquare, while textblob is better than MAPE side.

4. Model LSTM

Researchers analyze from the results of the tested model, to produce significant predictions of the actual stock price market using datasets with longer periods of time, because currently researchers using datasets within a period of 6 months have good results but if training is done with more data then the results will be more accurate. Generally, LSTM is more often used for predicting stock price movements or time series forecasting than for direct stock value predictions. LSTM has the ability to capture temporal patterns and long-term relationships in time series data, making it more suitable for forecasting future stock price changes. However, LSTM can still be applied to direct stock value prediction with additional indicators such as sentiment analysis.

Conclusion

This study aims to explore the potential of sentiment analysis in predicting stock prices. By combining sentiment data from news portals with historical data on stock prices, and has tested the analysis for the relationship of public sentiment to stock prices. Sentiment analysis is one of the indicators in predicting stock prices, the process of collecting data, preprocessing data, to analyzing sentiment and predicting with the LSTM model, it is known that sentiment analysis indicators are one of the supporting considerations. After testing the model conducted in this study, it was discovered that sentiment analysis was not a major factor.

There is a significant correlation between positive, negative, or neutral sentiment. But this relationship is complex because it is influenced by other factors such as news, market conditions, and companies. Sentiment as a supporting factor, which can provide additional information for market participants. Positive sentiment can also affect investors' outlook and potentially boost demand for stocks.

The level of prediction accuracy, in sentiment analysis that provides early indications for stock price predictions has varying levels of prediction accuracy. The performance of predictive models by combining sentiments tends to be better compared to models that rely solely on sentiment. The sentiment approach that researchers use from two different approaches, namely Textblob and Lexicon, is a reinforcement in taking perspectives on the relationship of sentiment analysis which is one of the factors in stock price predictions.

Changes in stock prices are often influenced by external factors and market volatility. Therefore, sentiment should be used as one of the supporting factors in making investment decisions. Future research could explore more about how sentiment can be integrated with other factors in stock price predictions.

References

Aasi, B., Imtiaz, S. A., Qadeer, H. A., Singarajah, M., & Kashef, R. (2021, April 21). Stock price prediction using a multivariate multistep LSTM: A sentiment and public engagement analysis model. 2021 IEEE International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021 - Proceedings. https://doi.org/10.1109/IEMTRONICS52119.2021.9422526

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems, 226. https://doi.org/10.1016/j.knosys.2021.107134

Cheng, L. C., Huang, Y. H., & Wu, M. E. (2019). Applied attention-based LSTM neural networks in stock prediction. Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018, 4716–4718. https://doi.org/10.1109/BigData.2018.8622541

Guo, Y. (2020). Stock Price Prediction Based on LSTM Neural Network: The Effectiveness of News Sentiment Analysis. Proceedings - 2020 2nd International Conference on Economic Management and Model Engineering, ICEMME 2020, 1018–1024. https://doi.org/10.1109/ICEMME51517.2020.00206

Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2021). A Review of Deep Learning Models for Time Series Prediction. In IEEE Sensors Journal (Vol. 21, Issue 6, pp. 7833–7848). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/JSEN.2019.2923982

ICIEV.2017.8338584. (n.d.).

IEEE Staff. (2016). 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES). IEEE.

IEEE Staff. (2017). 2017 International Conference on Computer and Applications (ICCA). IEEE.

Institute of Electrical and Electronics Engineers, & IEEE Industrial Electronics Society. (n.d.). Proceedings, 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS 2019) : Howards Plaza Hotel, Taipei, Taiwan, 06-09 May, 2019.

Jariwala, G., Agarwal, H., & Jadhav, V. (2020, November 6). Sentimental Analysis of News Headlines for Stock Market. 2020 IEEE International Conference for Innovation in Technology, INOCON 2020. https://doi.org/10.1109/INOCON50539.2020.9298333

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. IEEE Transactions on Smart Grid, 10(1), 841–851. https://doi.org/10.1109/TSG.2017.2753802

Liu, S., Liao, G., & Ding, Y. (2018). Stock transaction prediction modeling and analysis based on LSTM. Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications, ICIEA 2018, 2787–2790. https://doi.org/10.1109/ICIEA.2018.8398183

Rafi, S. H., Al-Masood, N., Deeba, S. R., & Hossain, E. (2021). A short-term load forecasting method using integrated CNN and LSTM network. IEEE Access, 9, 32436–32448. https://doi.org/10.1109/ACCESS.2021.3060654

Sarkar, A., Sahoo, A. K., Sah, S., & Pradhan, C. (2020). LSTMSA: A Novel Approach for Stock Market Prediction Using LSTM and Sentiment Analysis. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA).

Sharma, S. K. (2021). An Overview on Neural Network and Its Application. International Journal for Research in Applied Science and Engineering Technology, 9(8), 1242–1248. https://doi.org/10.22214/ijraset.2021.37597

Sharma, V., Singh, M., Galgotias College of Engineering and Technology, Galgotias College of Engineering and Technology. Department of Computer Science and

Engineering, Galgotias College of Engineering and Technology. Department of Information Technology, Institute of Electrical and Electronics Engineers. Uttar Pradesh Section, & Institute of Electrical and Electronics Engineers. (n.d.). Proceedings, IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking : (ICACCCN) : on 12th-13th Oct, 2018.

Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Seuk Wai, P., Wai Chung, Y., Piprani, A. Z., & Al-Garadi, M. A. (2018). Sentiment analysis of big data: Methods, applications, and open challenges. IEEE Access, 6, 37807–37827. https://doi.org/10.1109/ACCESS.2018.2851311