ICIC Express Letters ICIC International ©2024 ISSN 1881-803X
Volume **18**, Number **2**, February **2024** pp. **201–208**

STOCK PRICE PREDICTION USING ENSEMBLE LEARNING METHOD ON BIDIRECTIONAL LONG SHORT-TERM MEMORY MODEL WITH ATTENTION MECHANISM AND BERT'S FINANCIAL HEADLINES SENTIMENT

Valensius Ronald, Calvin Alfrido and Sani Muhamad Isa*

BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ valensius.ronald; calvin.alfrido }@binus.ac.id; *Corresponding author: sani.m.isa@binus.ac.id

Received February 2023; accepted May 2023

Abstract. *Stock price prediction may help investors in managing their stocks. To achieve that, several deep learning models have been studied and shown to be able to predict stock prices with relatively good results. Recently, sentiment analysis from various sources has been implemented by studies on the topic in order to achieve higher prediction power. Following the works that have been studied in this topic, we explore a deep learning method with added financial news sentiment for predicting stock prices. The proposed method uses ensemble learning on BiLSTM-AM models with added sentiment value from a fine-tuned BERT model. The model achieves an MASE score of 1.7153773.*

Keywords: Stock price, Time series forecasting, Financial news sentiment, Ensemble learning

1. **Introduction.** Predicting stock prices has always been a difficult yet rewarding challenge in the financial field. A series of bad investments will most likely lead to bankruptcy, and on the contrary, good investment decisions will bring fortune to the investor. As such, deciding on when and which stock to buy or sell requires deep, delicate, and intricate knowledge of the market.

With the recent rise of machine learning, this daunting task has become much more achievable than ever before. There are a lot of methods that can be used for predicting stock prices. Traditionally, investors would decide which stock to invest in based on either fundamental analysis, which focuses on the company's financial events (qualitative) and earnings (quantitative), or technical analysis, which focuses on the company's stock price movements [1]. The two methods of analysis can be translated into an algorithm that a machine can understand.

By feeding historical data into a machine learning model using different data sources and methods, researchers have been able to predict stock prices with varying accuracy. Traditional machine learning models typically achieve lower accuracy scores, while deep learning models are able to achieve better results [2]. Models that incorporate more complex addition with deep learning have also been recorded to achieve higher accuracy scores, as seen in [3,4,10]. These models would usually be satisfactory and could predict the rise and fall of prices in the market, but higher accuracy scores can still be achieved by using other methods.

Another data source that has been used to enhance predictions comes from the fundamental analysis factor, which is the company's financial events in the form of written news articles. Financial news gives investors a better understanding of the company's current events and standing in the market. Good news and bad news would correspond

DOI: 10.24507/icicel.18.02.201

to the rise and fall of stock prices [5]. Following this, works that incorporate financial news sentiment into stock price prediction were done and have been recorded to achieve better accuracy compared to methods that use only historical data [6,7].

Following the work that has been done on this topic, this study aims to explore a new and improved method for stock price prediction. This study proposes a method which incorporates financial news sentiment factors into an ensemble learning BiLSTM model with additional attention mechanism. By using these additional data and methods, our proposed model would be able to produce higher quality predictions and as such will help investors, both new and experienced, in making their decisions in the stock market. The proposed method is fed with the closing prices data of NETFLIX's stock, which is then evaluated by using the MASE score.

This paper is structured as follows: Section 2 describes a brief history and studies related to stock price prediction; Section 3 introduces the methods used in this paper; Section 4 details the simulation results and findings in this paper; and Section 5 presents our conclusion.

2. **Problem Statement and Preliminaries.** As of 2022, stock price prediction has been researched with some of the most sophisticated machine learning models and techniques to date. In this section, we review related works that include deep learning-based models or news articles as inputs to predictions.

2.1. **Historical price-based stock price prediction.** As seen in countless other works, deep learning has become the most prominent method of predicting stock prices as of now. Traditional machine learning models typically yield lower-accuracy results compared to deep learning models [2]. This advantage against traditional machine learning models leads to countless research and studies that used deep learning models as their method.

One of the more popular deep learning models used in past studies is the Long Short-Term Memory (LSTM) model. A study [8] compared the use of several LSTM combinations for stock price prediction. In the results, it was found that Bidirectional LSTM (BiLSTM) is better for low-volatility stocks, while CNN-LSTM is better for high-volatility stocks. Another study [9] was conducted that used a combination of CNN and BiLSTM. The study used data from Shanghai Composite Index, China Unicom, and CSI 300 and was able to achieve better prediction results compared to both CNN-LSTM and BiLSTM models.

Another deep learning method that was used for stock price prediction in the past is Recurrent Neural Network (RNN). A study [10] was conducted with RNN that was fed using 30 days of closing price data obtained from Tokyo Stock Exchange in order to predict the next day's price. Addition of Dynamic Time Warping (DTW) distance was used as an index for comparing the similarities of time-series data. The results showed that the proposed RNN model that applied DTW distance had better MAPE and RMSPE scores compared to the model that did not apply DTW. Another addition of Attention Mechanism (AM) [3,4] had also been researched and was able to achieve higher quality results compared to models that do not incorporate them. As such it can be concluded that addition of more complex methods may have a positive impact on a model's quality.

2.2. **News sentiment and historical price-based stock price prediction.** Investors play a massive role in the stock market. [3] suggested that future work on predicting stock prices should take the investors' views and emotions into consideration. To achieve this, there are variations of additional data that can be added to a dataset. One example is the investors' opinions and comments on social media (e.g., Twitter, and Reddit). The other example, which is the focus of this paper, is analyzing recent financial news.

Financial news is a great source to understand the current standing of a stock in the market. They are usually written in text, and as such, Natural Language Processing

(NLP) models may be deployed to gain information from them. [11] tested several NLP models and concluded that FinBERT was able to outperform Naïve Bayes and BERT classifiers for analyzing financial news.

Several other works have also incorporated financial news sentiment analysis as additional data to predict stock prices. A study that implemented lexicon-based sentiment analysis combined with LSTM on Hong Kong Exchange historical data and FINET news concluded that incorporating news sentiment and historical stock prices does improve accuracy compared to using them individually [7]. Another work using a FinBERT-LSTM model based on New York Times news and two years of historical NASDAQ price managed to reach an outstanding 98% accuracy, in which implementing sentiment news improves accuracy by 0.3640% [6].

3. **Methodology.** In this section, we explore several deep learning techniques and our implementation of said methods in our proposed model.

3.1. **Bidirectional Encoder Representations from Transformers.** Bidirectional Encoder Representations from Transformers (BERT) is a natural language processing model developed by Google. It is able to reach state-of-the-art accuracies by employing bidirectional training on a transformer, which is shown to be better at comprehending languages compared to single-directional models [12]. Base BERT models are available on Hugging Face, which researchers are able to download and fine-tune to their specific needs. For our needs, we acquired the base model "bert-base-uncased" and fine-tuned it to a combination of the FIQA and Financial Phrasebank datasets. The finished model is able to output the sentiment value embedding of a financial news headline, which includes negative, neutral, and positive.

3.2. **Attention mechanism.** An attention mechanism, or attention model, is a technique in deep learning that makes it possible for a neural network to focus on a specific component that is more important than the others. It is meant to follow how a human's cognitive attention works. Humans' vision is able to focus on more important areas, giving them importance and retrieving information from them. In the attention mechanism, focusing on a component works by giving each component weight values, which range from high values for important components to low values for insignificant components.

3.3. **Bidirectional Long Short-Term Memory.** Bidirectional Long Short-Term Memory (BiLSTM) is a variation of LSTM. A BiLSTM model combines two LSTM models, where one is used to understand sequential data in a time-forward manner and the other in a time-backward manner. In practice, a BiLSTM model is able to extract more information on how a feature is placed in its sequence compared to a standard LSTM model.

3.4. **Ensemble learning.** Ensemble learning is a method in deep learning where several models are trained and combined together to achieve better accuracy compared to individual models. This method originated from the idea that if several models' predictions are combined together, they will be less prone to individual errors and achieve better prediction results.

3.5. **Financial headlines sentiment.** Using BERT that has been trained on news headline data from the combined FIQA and Financial Phrasebank dataset [13], an output vector that represents the sentiment of that news will be obtained. This vector generally represents three possible class labels of the inputted news, which are "negative", "neutral", or "positive". The vector explains how "negative" the news is, how "neutral" the news is, and how "positive" the news is. With this sentiment that has been gained using BERT on the acquired news data of a specific company stock, the proposed idea of this research believes that it shares some correlation with the ups and downs of the specified company's stock price. Hence, using the acquired news embedding in the form of a vector of 3 values representing "negativity", "neutrality", and "positivity" of said company, this research paper proposes an implicit attention mechanism to be used to combine obtained news embeddings and historical data of the specified company's stock price. In the proposed method of this research paper, only the latest news of said company in a day will be used for the news embedding relevant to that company for that specific date to maintain the simplicity and robustness of the proposed model.

3.6. **Ensemble model BiLSTM-AM with BERT.** Our proposed method is visualized in Figure 1.

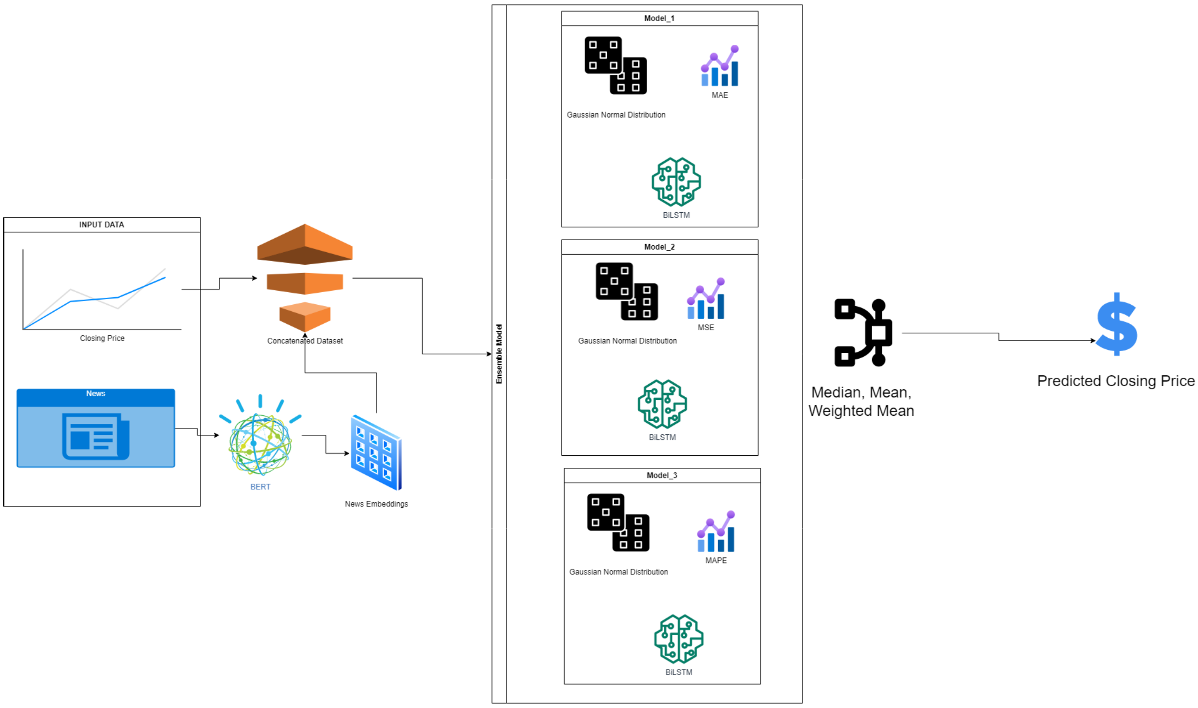

FIGURE 1. Proposed ensemble BiLSTM-AM with added BERT sentiment

In the proposed method of this research, BERT was used to obtain the news embedding in the form of a vector of normalized values. With the help of the implicit attention mechanism, these outputs in the form of news embedding were then combined with a sequence of time-series closing stock prices. In this research, a window size of 7 is used as the default window size for the historical data of closing stock prices. Using the concatenated data as input, several BiLSTM models are then trained, with the difference being in their respective loss functions, to achieve the "decision of the crowd" effect. In this research paper, several popular loss functions for time-series forecasting are used, which are MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and MASE (Mean Absolute Scaled Error). The equations for each loss function are computed as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{3}$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| \tag{4}$$

$$MASE = \frac{MAE}{\frac{1}{n-1}\sum_{t=2}^{n}|y_t - y_{t-1}|} \tag{5}$$

These several BiLSTM models are then assembled using several popular combining methods for ensemble models (the median, mean, and weighted mean, which will then be compared) of their predicted closing stock price. The output of the proposed model in this research paper is a predicted closing price of a stock within the horizon of 1 day given time-series closing price data of the previous 7 days of the specified stock.

4. **Results and Analysis.** There are several phases in the process of the proposed method in this research paper. The first phase of the proposed method is the preprocessing of the dataset used for training the BERT model. The second phase is then training and testing the BERT model to be able to output the desired news embedding from the news input. The third phase is to obtain news embeddings from the trained BERT model and preprocess them along with the stock price dataset. The fourth phase is then to train and test the performance of the ensemble model. All models in the ensemble model in this research are trained using the same hyperparameter except for the loss function of each model. The dataset used for training the two models (BERT and ensembled model) used in this research paper is different because the aim of the two models is different. While BERT is used to obtain the news embedding of the news input, the embedded models are used to actually predict the closing price of a specific company.

The evaluation scores of the models used in the ensembled model, along with the evaluation scores of several ensembled models (consisting of the same models) with different combination methods, are listed in Table 1.

Table 1. Performance evaluation of each model

Model	MAE	MSE	RMSE	MAPE	MASE
Model_0 (MAE)	13.5260725	334.10193	18.27846	5.3015814	2.091102
Model_1 (MSE)	11.450924	231.55664	15.21699	4.7293706	1.7702883
Model_2 (MAPE)	19.593906	561.334	23.69249	8.316271	3.029176
Ensembled model (Median)	12.326321	285.46265	16.89564	4.922573	1.9056228
Ensembled model (Mean)	12.22978	279.11877	16.70685	4.8828735	1.8906978
Ensembled model (Weighted Mean)	11.095737	227.69586	15.08959	4.4742575	1.7153773

This study aims to predict stock prices using an ensemble of different models. The models used in this study include MAE (Model_0), MSE (Model_1), and MAPE (Model_2), as well as three ensembled models using different methods (Median, Mean, and Weighted Mean). The results of the study suggest that ensembling different models can lead to improved performance compared to using a single model, and that the choice of ensemble method can have a significant impact on performance.

The individual models used in this study (Model_0, Model_1, and Model_2) all performed reasonably well, with MAE values ranging from 11.450924 to 19.593906, and MSE values ranging from 231.55664 to 561.334. However, all of these values were outperformed by the ensembled models. It is worth noting that Model_1 had the lowest MSE value among the individual models, which suggests that it might be a useful component to include in an ensemble model.

The ensembled models generally performed better than the individual models. The Weighted Mean ensemble had the best value across all metrics, which suggests that it might be the most appropriate method for predicting stock prices. The MASE values for all models suggest that they are performing better than a naive forecast model that always predicts the last observed value, and the MAE and RMSE values suggest that the

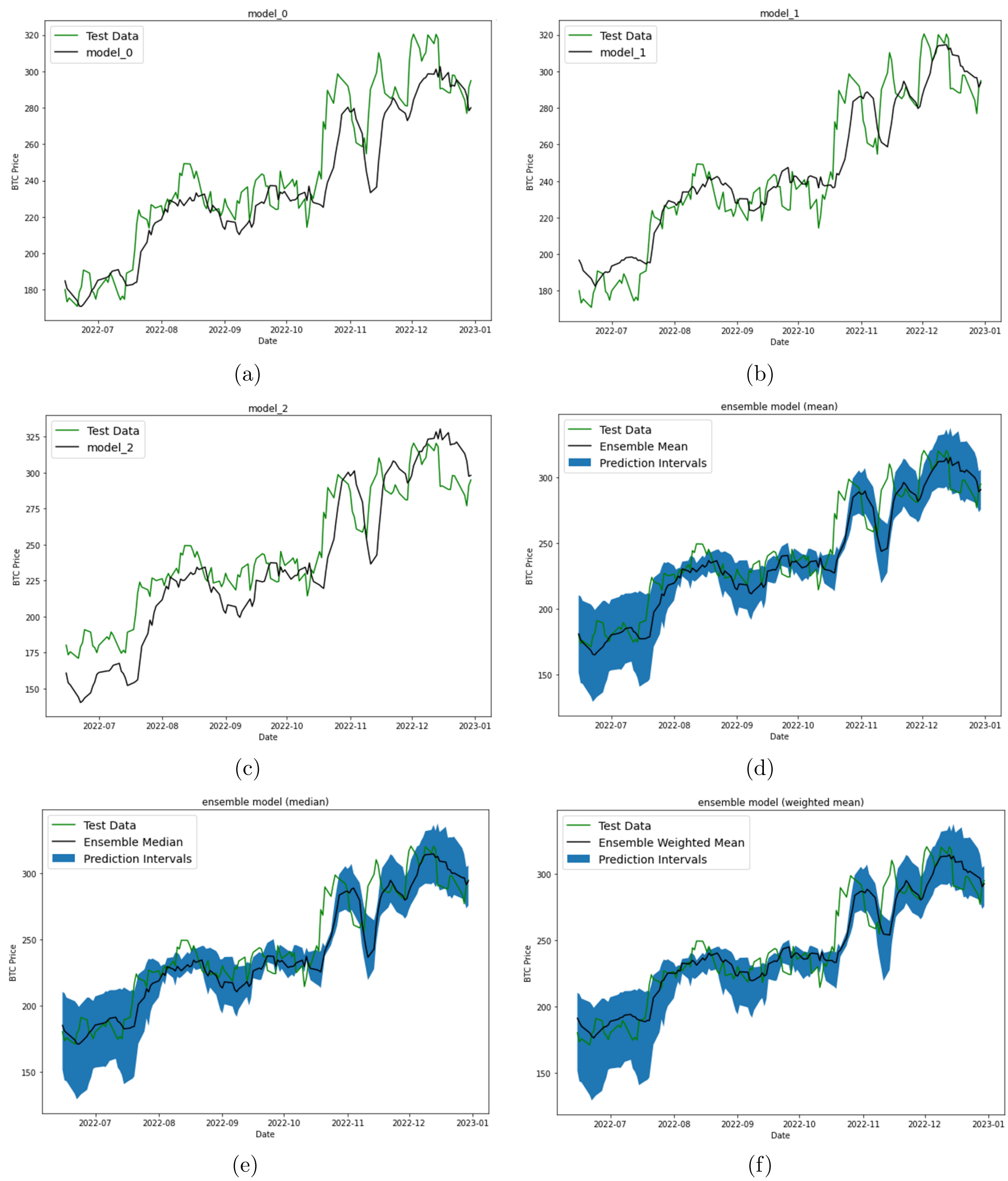

FIGURE 2. (a) Prediction results of Model_0; (b) prediction results of Model_1; (c) prediction results of Model_2; (d) prediction results of the ensemble model using Mean; (e) prediction results of the ensemble model using Median; (f) prediction results of the ensemble model using Weighted Mean

predictions are off by a moderate amount on average. However, it is important to keep in mind that the scale of the stock prices being predicted could influence the interpretation of these values. Lastly, the Median ensemble had the highest MAPE value, which suggests that it might not be the best choice for predicting stock prices for long-term investments.

The various models' results used in this research to predict the closing stock price of Netflix for a period of time can be seen in Figure 2.

From the figures, it can be concluded that the ensembled model can achieve better accuracy in predicting stock prices compared to the standalone models. This indicates that the desired effect of "decision of the crowds" is achieved using the ensembled model, which helps in achieving better accuracy.

While the results of this study are promising, there are several potential limitations to consider. For example, the models used in this study might be overfitting to the training data, which could lead to poor performance when making predictions on new data. Additionally, the models might not be taking account of all relevant information when making predictions, which could also lead to suboptimal performance. It is worth considering whether additional feature engineering or model tuning could help improve the models' performance.

Finally, it is important to consider the context in which these models are being used. For example, if the goal is to make short-term trades based on predicted stock prices, then models that perform well in terms of absolute error (MAE) may be more important than models that perform well in terms of percentage error (MAPE). On the other hand, if the goal is to make long-term investments, then models that can accurately predict the direction of the stock prices (i.e., models with low MAPE values) may be more useful.

5. **Conclusion.** Using our proposed method, we are able to achieve an outstanding MASE (Mean Absolute Scaled Error) score of 1.7153773 on predicting stock prices using news as a source of sentiment. The combined ensemble model has better results compared to using each model individually. For future studies, researchers may be able to use other loss functions and sentiment models to achieve better results. Sentiment is also a good source of information. By adding new sentiment data sources, such as tweets from Twitter, for further analysis, models may be able to achieve greater predictive power. It is also possible to use another method of implementing sentiment, such as error correction.

REFERENCES

[1] IG Markets Limited, How to pick stocks: The complete guide, *IG*, https://www.ig.com/en/trading-strategies/how-to-pick-stocks-190926, Accessed on Jan. 02, 2023.
[2] S. Mehtab and J. Sen, Stock price prediction using convolutional neural networks on a multivariate timeseries, *arXiv Preprint*, arXiv: 200109769, 2020.
[3] Y. Chen et al., Stock price forecast based on CNN-BiLSTM-ECA model, *Sci. Program.*, vol.2021, pp.1-20, 2021.
[4] W. Lu, J. Li, J. Wang and L. Qin, A CNN-BiLSTM-AM method for stock price prediction, *Neural Comput. Appl.*, vol.33, pp.4741-4753, 2021.
[5] B. Beers, How the news affects stock prices, *Investopedia*, https://www.investopedia.com/ask/answers/155.asp, Accessed on Sep. 30, 2021.
[6] S. Halder, FinBERT-LSTM: Deep learning based stock price prediction using news sentiment analysis, *arXiv Preprint*, arXiv: 221107392, 2022.
[7] X. Li, P. Wu and W. Wang, Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong, *Inf. Process. Manag.*, vol.57, no.5, 102212, 2020.
[8] O. E. Orsel and S. S. Yamada, Comparative study of machine learning models for stock price prediction, *arXiv Preprint*, arXiv: 220203156, 2022.
[9] H. Wang, J. Wang, L. Cao, Y. Li, Q. Sun and J. Wang, A stock closing price prediction model based on CNN-BiSLSTM, *Complexity*, vol.2021, pp.1-12, 2021.
[10] S. Rikukawa, H. Mori and T. Harada, Recurrent neural network based stock price prediction using multiple stock brands, *International Journal of Innovative Computing, Information and Control*, vol.16, no.3, pp.1093-1099, 2020.

[11] F. S. Syeda, *Sentiment Analysis of Financial News with Supervised Learning*, https://www.diva-portal.org/smash/get/diva2:1636643/FULLTEXT01.pdf, 2022.

[12] J. Deviln, M.-W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *arXiv Preprint*, arXiv: 181004805, 2018.

[13] P. Malo et al., Good debt or bad debt: Detecting semantic orientations in economic texts, *Journal of the Association for Information Science and Technology*, vol.65, no.4, pp.782-796, 2014.