2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI) | 978-1-6654-0898-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICTAI52525.2021.00160

Stock Market Trend Forecasting Based on Multiple Textual Features: A Deep Learning Method

Zhenda Hu[1†], Zhaoxia Wang[2*], Seng-Beng Ho[3], Ah-Hwee Tan[2]
School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai, China[1]
School of Computing and Information Systems, Singapore Management University, Singapore, Singapore[2]
Social and Cognitive Computing Department, Institute of High Performance Computing (IHPC),
Agency for Science, Technology and Research (A*STAR), Singapore, Singapore[3]
huzhenda2020@gmail.com, {zhendahu, zxwang, ahtan}@smu.edu.sg, hosb@ihpc.a-star.edu.sg

***Abstract*—Stock market trend forecasting is a valuable and challenging research task for both industry and academia. In order to explore the influence of stock news information on the stock market trend, a textual embedding construction method is proposed to encode multiple textual features, including topic features, sentiment features, and semantic features extracted from stock news textual content. In addition, a deep learning method is designed by using financial data and multiple textual features obtained from multiple news textual embeddings for short-term stock market trend prediction. For evaluation, extensive experiments on real stock market data are conducted. The experimental results illustrate that the proposed method can enhance the performance of predicting stock market trend by obtaining effective information from stock news.**

***Index Terms*—Stock Market Trend Forecasting, Textual Features, Deep Learning, Sentiment Analysis**

I. Introduction

Stock market prediction is an interesting and valuable research task for both industry and academia [1]. With the development of big data, AI techniques and machine learning methods, advanced methods using machine learning even deep learning models have been widely applied to stock market prediction tasks recently [2]. For example, researchers have utilized various machine learning algorithms, such as Support Vector Machine (SVM) and Artificial Neural Network (ANN), for stock market trending analysis and prediction [3]–[7]. Furthermore, advancement of deep learning technology has resulted in most stock forecasting tasks being carried out using deep learning frameworks [2]. Various studies have shown the importance of Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) as the most representative models [8]–[10].

On the other hand, inspired by behavioral finance, researchers began to incorporate information that reflects investors' behaviors to enhance stock forecasting models. As a result, NLP techniques for performing text mining from financial news for stock market trend prediction have also been widely explored [11]–[13]. For example, financial news aggregator and market insights providers, such as Bloomberg, can potentially further value-add to their clients by providing data-driven market trending indicators through text mining on the financial news pipelines. Furthermore, a massive amount of data is generated each day as a result of social networking. Consequently, using social media data to improve prediction performance is becoming more and more popular [14]–[18].

In this paper, in order to explore the influence of stock news information on the stock market trend, two types of input features, namely the financial data features and textual data features are taken into consideration. The textual features are derived from the news data, consisting of topic features, sentiment features and semantic features to produce the most effective information representation. Specifically, topic features are obtained from the news article titles represented by their topic distribution to describe what the articles mainly talk about, sentiment features include the sentiment polarity of news title and body to represent their sentiment polarity, and semantic features are constructed by the word embedding of news title and body through word2vec, a popular pre-trained language model.

In summary, we make the following contributions: 1) We propose a news textual embedding construction method which can reflect the emotion, semantic and topics information in stock news textual content by encoding the news texts into embedding vectors; 2) We design a deep learning method integrating financial data and news textual embeddings for short-term stock market trend prediction; 3) We conduct extensive empirical experiments on real stock market data. Based on our experimental results, we find that our model is capable of encoding relevant stock market news information and predicting market trends in an efficient manner.

The rest of this paper is organized as follows: In Section 2, we discuss related work on stock prediction. Section 3 presents the proposed stock trending prediction approach. Section 4 presents the experiment details of the experiments and discusses the results. Finally, Section 5 presents the conclusion and future work.

[†]This work was done while the first author was an Visiting Postgraduate Research Student (PhD) at Singapore Management University

[*]Corresponding Author

978-1-6654-0898-1/21/$31.00 ©2021 IEEE
DOI 10.1109/ICTAI52525.2021.00160

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on September 20,2022 at 06:10:59 UTC from IEEE Xplore. Restrictions apply.

II. Related work

A. Stock Forecasting Based on Deep Learning

Compared with the traditional statistical methods, machine learning algorithms have the capability of dealing with complex and multi-dimensional data. Due to the better prediction performance, various machine learning algorithms, such as SVM and ANN have been utilized for stock market trending analysis and prediction [3]–[7].

In particular, deep learning (DL), a modern trend in ML, utilizes a deep model structure with high-level representation of attribute features, and is capable of extracting relevant information from the financial time series [19]. With the development of deep learning technology, most stock forecasting tasks have been carried out using deep learning frameworks recently [8]. LSTM and CNN are the most typical models which have been commonly used in the related research studies [8]–[10].

For example, a hybrid model which combined LSTM and Genetic Algorithm was developed by Chung et al. [20] to predict stock market. The final results indicated that the hybrid model produced superior results when compared with the baseline model. Long et al. [21] designed a novel deep learning model named multi-filters neural network (MFNN) for feature extraction and price prediction task on financial data. The results showed that MFNN outperformed statistical models, traditional machine learning and deep learning models such as CNN, RNN, and LSTM in terms of the experimental performances. Also, Zhang et al. [22] proposed an innovative prediction method based on deep learning, and under the decomposing and integrating framework, they constructed a hybrid prediction model, called CEEMD-PCA-LSTM, for stock markets.

Furthermore, many research studies on stock forecasting utilized attention mechanism to enhance the performance [23]–[26]. For instance, Hierarchical Attention Networks (HAN) [24] applied to stock market prediction [25] and Self-attention Networks [26] designed by Zheng et al. employed the idea of Transformer for stock volatility forecasting. Inspired by three principles including sequential content dependency, diverse influence, and efficient learning, Hu et al. [25] also used HAN to forecast the stock trend by analyzing the sequence of related news and implemented a self-paced learning mechanism to emulate efficient learning.

Compared with the above work, in this paper, we proposed a relatively flexible and elegant framework in which different deep learning models can be selectively applied according to the features of specific stock indices.

B. Stock Forecasting with Social Media Sentiment Analysis

A popular research area in Natural Language Understanding (NLU) is social media sentiment analysis, which identifies and categorizes opinions expressed in news, articles, tweets, blog posts, or any other texts [27], [28]. In the field of stock market prediction, many researchers enhance stock movement prediction by incorporating public sentiment based on events and scenarios.

Generally, the usage of public information such as Twitter data or New York Times news data for stock price forecasting comprises two essential steps. The first is to transform text into an understandable form for machines, and the next is to utilize statistical models or machine learning models to make predictions with information from diverse sources [29]. For converting news text into a machine-recognizable language, many effective methods have been proposed [29]–[31]. For example, in [31], a stock prediction model was developed based on the sentiments of the company's specific topics. The proposed approach as well as existing topic models are used to automatically extract topics and associated sentiments from a message board. In a different approach, Ni et al. [29] proposed a hybrid stock market forecasting method with tweets embedding and historical prices. In contrast to the classic text embedding methods, the method considered internal semantic features as well as external structural features of Twitter data to obtain more effective information from the generated tweet vectors.

Additionally, there are numerous studies that integrate sentiment analysis from social media with stock market prediction in an effective manner [32]–[35]. Bouktif et al. [32] conducted a finer-grained analysis experiment based on stock price data, sentiment polarity, and customized text-based features with various feature lags, which provided constructive contributions by empirically investigating the predictability of stock market trends. Wang et al. [34] proposed a novel framework which included time-sensitive and target-aware investment stance detection, expert-based dynamic stance aggregation, and stock movement prediction. In their proposed stance detection model named Multi-view Fusion Network (MFN), the representation of each review derived from integrating multi-view textual features and extended knowledge in financial domain to distill bullish/bearish investment opinions could be learned.

Different from the work reviewed above, we utilize multiple news textual features containing topic, sentiment and semantic information for stock market trend prediction. The research work and findings of this research not only demonstrate the merits of the proposed method, but also point out the correct direction for future work in this area.

III. Proposed Method

A. Problem Formulation

In this research, the stock market forecasting problem is formulated as a three-class classification task of predicting the stock price movement: Up, Hold and Down (i.e., 1, 0, -1). Specifically, given the stock prices sequences $X_T(T = 1, 2, 3, \cdots, n)$ and multiple textual features $M_T(T = 1, 2, 3, \cdots, n)$, the task is to predict the stock price trend of the next day Y_{T+1}. It can be formulated as equation (1):

$$\begin{aligned} Y_{T+1} = F(&X_T, X_{T-1}, X_{T-2}, \cdots, X_{T-K}, \\ &M_T, M_{T-1}, M_{T-2}, \cdots, M_{T-K}) \end{aligned} \tag{1}$$

where $F()$ represents the mapping function from input to output; K represents the size of the sliding window; multiple

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on September 20,2022 at 06:10:59 UTC from IEEE Xplore. Restrictions apply.

textual features M_T consist of topic features M_T^T, sentiment features M_T^S, and semantic features M_T^W, as shown in equation (2):

$$M_T = (M_T^T \odot M_T^S \odot M_T^W) \quad (2)$$

where $\odot$ represents an operation to combine the different textual features. In this paper, the simple but useful way of direct concatenation is used for the integration of different textual features.

B. Proposed Method

In this section, we describe our method in details. Our proposed method aims to leverage real stock market time series data and stock news texts data to predict the stock price trend. We introduce a novel stock market trend forecasting method to investigate and compare the performance of different deep learning models which utilize multiple textual features from social media news texts analysis. Our method consists of three stages, each with multiple steps. Figure 1 illustrates the overview of our method.

Stage 1: In stage 1, we decide on the optimal number of topics for stock news texts and generate the topic features M_T^T to represent the topic distribution for each news item.

Intuitively, certain news items may gain more traction because of the underlying topics they discuss; such topics or coverage discuss impactful events on the market development, so those news items stand out as more important features in predicting market activities. Additionally, a piece of text document can also be represented by its topic distribution using the Latent Dirichlet Allocation (LDA) [36] framework. Through the following two steps, we can obtain the topic features.

- Step 1: We use coherence score to evaluate the LDA models with a varying number of topics, and then determine the optimal number of topics.
- Step 2: After deciding the optimal number of topics n discovered within the entire corpus, we implement topic model on the news title by representing each document as an n-dimensional vector.

Stage 2: In stage 2, we construct the news textual embeddings M_T consisting of topic features M_T^T, sentiment features M_T^S and semantic features M_T^W. Also, we construct financial features X_T from stock price data.

Besides topic features, we also need to construct sentiment features and semantic features to explore more information contained in the news. The sentiment expressed in mainstream media on market outlook can deeply influence how market participants make their buy / sell decisions, as they would be likely to short their positions on equities of which the prices will fall due to the negative outlook and conversely, long their positions on equities of which the prices will rise in view of the positive outlook. In addition, we choose word embeddings as input features to capture the semantic meaning of the textual data. Furthermore, we need to decide the optimal size of sliding window K for the financial features. As a result, we conduct the following three steps:

- Step 1: Two sentiment features are formed, one each for the news title and news body to represent its sentiment polarity. The feature is the compound score obtained using the NLTK Valence Awareness Dictionary for Sentiment Reasoning (VADER) toolkit when performing sentiment analysis on the subject text, which is normalized between the lower limit -1 and the upper limit 1.
- Step 2: Instead of using the bag-of-words or TF-IDF set up, we use word embedding to represent words from the corpus to counter high dimensionality. Specifically, we use the pre-trained embeddings from a subset of the Google News, which generate two 300-dimensional vectors, one each for the news title and news body.
- Step 3: To decide the optimal size of sliding window K for stock price data.

Stage 3: To predict the final stock trend results with deep learning model using the stock price features X_T and news textual embeddings M_T as the input by the simple but useful way of direct concatenation.

- Step 1: We compare the performance of different deep learning models, including FNN, CNN and LSTM, through the comparative experiments.
- Step 2: After identifying the most suitable model with the highest accuracy when training with the stock price features X_T, we adopt the predicting model for our proposed method integrating both the stock price features X_T and the news textual embedding features M_T constructed in the previous two stages.

IV. Experiments and Results

In this section, the details of the experiments, including the data sets, data preprocessing, and experimental setup are described and then an analysis is provided on whether our method is feasible with the real-world stock news data and stock price data. Lastly, based on the evaluation metrics, we summarize the experiment results of the above mentioned method.

A. Dataset

In our experiments, we utilize two types of data for our method, namely, news data and financial data. Specifically, the news data was sourced from Reuters. The original dataset was released by Ding et al. [9] in their study carried out in 2015 on deep learning for event-driven stock prediction. The original dataset contains more than 109,000 Reuter's news items, spanning from Oct 2006 to Nov 2013 and is being organized into daily folders with each folder containing news items published by Reuter's reporters throughout the day. Each piece of news includes the news article title and the article body.

In addition, financial data is sourced from S&P historical trading activities retrieved from Yahoo Finance for the period of Oct 2006 to Nov 2013 to overlay with the duration of news data. The original dataset contains Open, High, Low, Close price of the index and total trading volume over a period of 1,782 trading days.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on September 20,2022 at 06:10:59 UTC from IEEE Xplore. Restrictions apply.

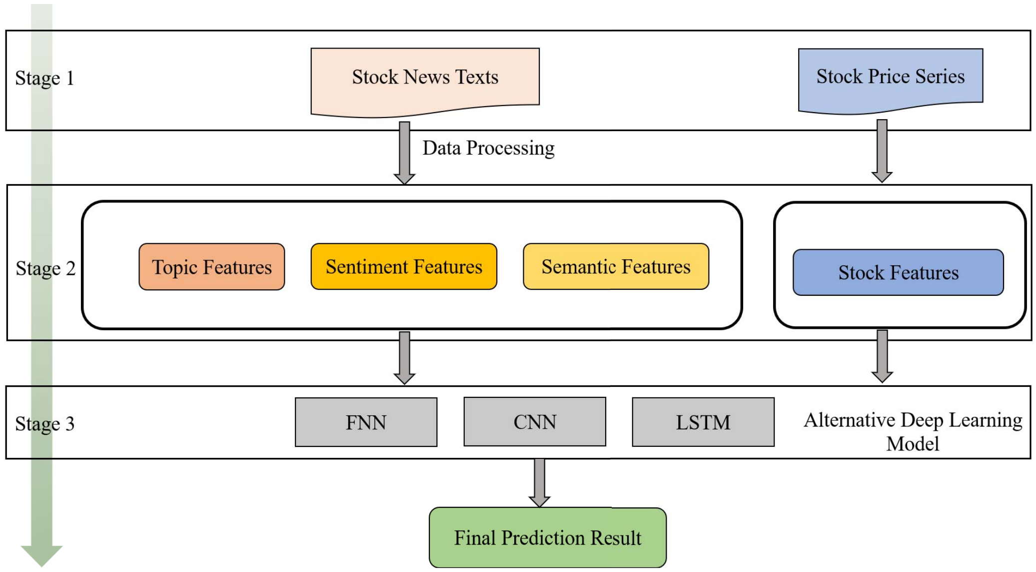

Fig. 1. The proposed framework

B. Data Processing

For news data, assuming the news is published on time T, we use the information gathered on T to predict market trend at $T+1$. For weekends, the news published on Friday, Saturday and Sunday are concatenated to predict the market trend for the following Monday. Using standard preprocessing, we conduct tokenization, punctuation removal, stopwords removal, stemming and other operations for the news texts.

For financial data, we define the movement/trend based on the percentage change against the previous day rather than in absolute terms. We define the percentage change as the difference between $Open_{T+1}$ and $Open_T$ as a percentage of $Open_T$ and the formula is represented as equation (3):

$$open\_chg\_\% = \frac{Open_{T+1} - Open_T}{Open_T} * 100\% \tag{3}$$

Based on transaction rules, trading is associated with costs such as transaction fees. Hence, it would be meaningful to undertake a buy or sell decision only if the profits resulted from the trading activity offset the costs. After inspecting the data, we found that the distribution across the three classes (1, 0, -1) would be close to even when the threshold was set at 0.3%. Thus, for better model training, a threshold of 0.3% was chosen.

C. Experimental Setup

The whole dataset was separated into training and test set. Specifically, 70% of the dataset was selected as the training set, for training the deep learning model, and the remaining 30% was used as the testing set, for verifying the performance of the models.

For topic features, we used the LDA model from thr Genism library, which can automatically extract semantic topics from documents. Specifically, we can examine the joint occurrence patterns of lexical statistics for the document corpus, which can be used to explore the semantic structure of documents. After comparison of coherence scores obtained by using 1 to 15 topics, the optimal number of topics was determined to be 6 for the entire news dataset.

For sentiment features, we used SentimentIntensityAnalyzer from the NLTK library and obtained a total of two sentiment features, one for news title and the other for news body. For news body, we calculated the polarity score of each sentence then took the average.

In addition, for semantic features, we utilized the word2vec model from the Genism library. Specifically, we generated a 300-dimension word embedding for the title and body of each news item. Finally, we obtained a 600-dimension vector.

As for the deep learning method, we deployed three deep learning models, including Convolutional neural network (CNN), Feedforward Neural Network (FNN) and Long Short-Term Memory network (LSTM). Specifically, FNN consists of two fully connected dense layer, with 32 units and 16 units each, and an output layer of three units corresponding to three classes with a softmax activation function. The CNN network consists of three 1-D convolution layers, with 32, 64 and 128 filters each, followed by a dropout layer with a 0.2 dropout rate, and further followed by a fully connected layer before the final output layer. The LSTM model comprises of two LSTM layers with 50 units each, followed by a dropout layer with a 0.2 dropout rate. All those configurations were set through Keras, a famous Python deep learning API.

D. Evaluation Metrics

The experimental performance is evaluated based on two common classification indices, namely Accuracy rate and $F1$-score. Specifically, the measurement of Accuracy rate takes into account the number of correct predictions out of the total

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on September 20,2022 at 06:10:59 UTC from IEEE Xplore. Restrictions apply.

number of predictions. $F1$-score combines the precision and recall rates in a classification task. The formula for $F1$-score is represented as equation (4):

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

E. Experimental Results

(1) Decide on the Optimal Sliding Window Size K

To analyze the impact of the sliding window size, we use SVM classifier to predict the next day trend of the open price on the test dataset with a sliding window size of K in the range of 3–13 and increased K by 2 each time. The results are shown in Table I. The optimal sliding window size K emerges when the best prediction outcome is achieved under the optimal K. It can be seen that the $F1$-score slightly improves when the sliding window size is increased from 3 to 7. After the sliding window size increases beyond 7, the values don't change in an obvious manner or become worse. The reason might be that more noise was introduced into the model. Through the above analysis, we use a sliding window size of 7 for forecasting stock market trending.

TABLE I
THE PERFORMANCE WITH DIFFERENT SLIDING WINDOW SIZE K

Value of K (Sliding window)	$Precision$	$Recall$	$F1$
3	0.91	0.91	0.91
5	0.91	0.90	0.90
7	**0.92**	**0.92**	**0.92**
9	0.89	0.89	0.89
11	0.90	0.90	0.90
13	0.89	0.89	0.89

(2) Results for Baseline Models

As shown in Table II, we compare the three deep learning models with the baseline. It can be seen that LSTM outperforms the other two models both in training accuracy and testing accuracy, which implies that LSTM has a good capability for forecasting time series data.

TABLE II
THE ACCURACY VALUES OF BASELINE

Model	Training Accuracy	Test Accuracy
FNN	0.8441	0.8370
CNN	0.8562	0.8258
LSTM	**0.8667**	**0.8502**

(3) Results for the Proposed Method

To explore how different textual features influence the experiment performances, we conduct various combinations of features from topic distribution, sentiment scores, and word2vec embeddings mentioned in Section III. The results are shown in Table III.

Compared to the Baseline model, the Baseline+topic+sentiment+semantic model which incorporates three textual features achieves slight improvement for both $Precision$ and $F1$-score. Also, we find that the performance doesn't change significantly when only topic features or topic features and sentiment features are added. At the same time,

TABLE III
THE PRECISION, RECALL AND $F1$-SCORE VALUES OF VARIOUS MODELS

Model	Label	$Precision$	$Recall$	$F1$	$Macro\ F1$
Baseline	1	0.94	0.93	0.93	0.917
	0	0.91	0.88	0.89	
	-1	0.91	**0.96**	0.93	
Baseline+topic	1	0.94	0.93	0.93	0.910
	0	0.89	0.87	0.88	
	-1	0.90	0.94	0.92	
Baseline+topic +sentiment	1	0.94	0.88	0.91	0.900
	0	0.85	0.89	0.87	
	-1	0.91	0.94	0.92	
Baseline +sentiment +semantic	1	0.92	0.94	0.93	0.917
	0	0.90	0.88	0.89	
	-1	0.93	0.93	0.93	
Baseline+topic +sentiment +semantic	1	**0.95**	0.93	**0.94**	**0.923**
	0	0.89	0.90	0.90	
	-1	0.92	0.94	0.93	

we find that the performance of label 0 doesn't improve significantly when compared to label 1 and label -1. This is desirable from a practical standpoint, as a buy or sell decision is triggered by either upward or downward movement. Hence, improving the performance for label 1 and label -1 is more meaningful.

From the results, it can be seen that the effectiveness of our method requires suitable combinations of the textual embeddings generated by multiple textual features with real data. Suitable combinations are thus important to enhance the performance of stock trend forecasting tasks.

V. CONCLUSION AND FUTURE WORKS

In conclusion, to explore the influence of stock news information about the prediction on the stock market trend, we apply two types of input features including financial data features and textual data features to stock market trend prediction. The textual features are derived from the news data, consisting of topic features, sentiment features, and semantic features. Furthermore, to construct a suitable dataset for this research, we merge a news dataset comprising daily news coverage from Reuters with a stock market dataset corresponding to the S&P index historical trading data over the same period of time. Based on the empirical experiments conducted, the baseline model with three multiple textual features outperforms the other models, with the highest $F1$-score of 0.94, which shows the effectiveness of our proposed method.

As a limitation of this research, we obtain textual information only from a single news source - Reuters. It may therefore lack the diversity in opinion expressions as compared to gathering news from multiple news sources. Also, there have been many techniques presented in the fields of information fusion and multi-modal pattern recognition. In the future, we can explore different methods for integrating financial features and textual features. In addition, because the use of partial textual features does not seem to increase the precision of the classifier using price features, it may be worthwhile to explore whether rumours may help to improve the prediction performance, which could be used to devise ad-hoc market strategies.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on September 20,2022 at 06:10:59 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] M. Ananthi and K. Vijayakumar, "Stock market analysis using candlestick regression and market trend prediction (ckrm)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 4819–4826, 2021.

[2] Q. Li, Y. Chen, J. Wang, Y. Chen, and H. Chen, "Web media and stock markets: A survey and future directions from a big data perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 2, pp. 381–399, 2017.

[3] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *The Journal of finance and data science*, vol. 4, no. 3, pp. 183–201, 2018.

[4] Y. Lin, H. Guo, and J. Hu, "An svm-based approach for stock market trend prediction," in *The 2013 international joint conference on neural networks (IJCNN)*. IEEE, 2013, pp. 1–7.

[5] M. Qiu and Y. Song, "Predicting the direction of stock market index movement using an optimized artificial neural network model," *PloS one*, vol. 11, no. 5, p. e0155133, 2016.

[6] K. O. Nti, A. Adekoya, and B. Weyori, "Random forest based feature selection of macroeconomic variables for stock market prediction," *American Journal of Applied Sciences*, vol. 16, no. 7, pp. 200–212, 2019.

[7] R. S. Deepak, S. I. Uday, and D. Malathi, "Machine learning approach in stock market prediction," *International Journal of Pure and Applied Mathematics*, vol. 115, no. 8, pp. 71–77, 2017.

[8] R. Akita, A. Yoshihara, T. Matsubara, and K. Uehara, "Deep learning for stock prediction using numerical and textual information," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*. IEEE, 2016, pp. 1–6.

[9] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[10] A. M. Rather, A. Agarwal, and V. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3234–3241, 2015.

[11] H. T. Thanh and P. Meesad, "Stock market trend prediction based on text mining of corporate web and time series data," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 18, no. 1, pp. 22–31, 2014.

[12] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, "Text mining for market prediction: A systematic review," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7653–7670, 2014.

[13] S. E. Seker, M. Cihan, A.-N. Khaled, N. Ozalp, and A. Ugur, "Time series analysis on stock market for text mining correlation of economy news," *International Journal of Social Sciences and Humanity Studies*, vol. 6, no. 1, pp. 69–91, 2013.

[14] B. Li, K. C. Chan, C. Ou, and S. Ruifeng, "Discovering public sentiment in social media for predicting stock movement of publicly listed companies," *Information Systems*, vol. 69, pp. 81–92, 2017.

[15] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–24, 2020.

[16] M. Javed Awan, M. S. Mohd Rahim, H. Nobanee, A. Munawar, A. Yasin, and A. M. Zain, "Social media and stock market prediction: a big data approach," *MJ Awan, M. Shafry, H. Nobanee, A. Munawar, A. Yasin et al.," Social media and stock market prediction: a big data approach," Computers, Materials & Continua*, vol. 67, no. 2, pp. 2569–2583, 2021.

[17] P. Mehta, S. Pandya, and K. Kotecha, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," *PeerJ Computer Science*, vol. 7, p. e476, 2021.

[18] Z. Wang, S.-B. Ho, and Z. Lin, "Stock market prediction analysis by incorporating social and news opinion and sentiment," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018, pp. 1375–1380.

[19] R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, J. P. Nobrega, and A. L. Oliveira, "Computational intelligence and financial markets: A survey and future directions," *Expert Systems with Applications*, vol. 55, pp. 194–211, 2016.

[20] H. Chung and K.-s. Shin, "Genetic algorithm-optimized long short-term memory network for stock market prediction," *Sustainability*, vol. 10, no. 10, p. 3765, 2018.

[21] W. Long, Z. Lu, and L. Cui, "Deep learning-based feature engineering for stock price movement prediction," *Knowledge-Based Systems*, vol. 164, pp. 163–173, 2019.

[22] B. Yan, M. Aasma *et al.*, "A novel deep learning framework: Prediction and analysis of financial time series using ceemd and lstm," *Expert systems with applications*, vol. 159, p. 113609, 2020.

[23] J. Long, Z. Chen, W. He, T. Wu, and J. Ren, "An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in chinese stock exchange market," *Applied Soft Computing*, vol. 91, p. 106205, 2020.

[24] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480–1489.

[25] Z. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proceedings of the eleventh ACM international conference on web search and data mining*, 2018, pp. 261–269.

[26] J. Zheng, A. Xia, L. Shao, T. Wan, and Z. Qin, "Stock volatility prediction based on self-attention networks with social information," in *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*. IEEE, 2019, pp. 1–7.

[27] Z. Wang, C. S. Chong, L. Lan, Y. Yang, S.-B. Ho, and J. C. Tong, "Fine-grained sentiment analysis of social media with emotion sensing," in *2016 Future Technologies Conference (FTC)*. IEEE, 2016, pp. 1361–1364.

[28] Z. Wang, S.-B. Ho, and E. Cambria, "A review of emotion sensing: categorization models and algorithms," *Multimedia Tools and Applications*, vol. 79, no. 47, pp. 35 553–35 582, 2020.

[29] H. Ni, S. Wang, and P. Cheng, "A hybrid approach for stock trend prediction based on tweets embedding and historical prices," *World Wide Web*, vol. 24, no. 3, pp. 849–868, 2021.

[30] J. Chun, J. Ahn, Y. Kim, and S. Lee, "Using deep learning to develop a stock price prediction model based on individual investor emotions," *Journal of Behavioral Finance*, pp. 1–10, 2020.

[31] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications*, vol. 42, no. 24, pp. 9603–9611, 2015.

[32] S. Bouktif, A. Fiaz, and M. Awad, "Augmented textual features-based stock market prediction," *IEEE Access*, vol. 8, pp. 40 269–40 282, 2020.

[33] M. Jaggi, P. Mandal, S. Narang, U. Naseem, and M. Khushi, "Text mining of stocktwits data for predicting stock prices," *Applied System Innovation*, vol. 4, no. 1, p. 13, 2021.

[34] H. Wang, T. Wang, and Y. Li, "Incorporating expert-based investment opinion signals in stock prediction: A deep learning framework," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 971–978.

[35] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Systems with Applications*, vol. 135, pp. 60–70, 2019.

[36] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on September 20,2022 at 06:10:59 UTC from IEEE Xplore. Restrictions apply.