Forecasting Stock Market Volume Price Using Sentimental and Technical Analysis

Siddesh G. M., Ramaiah Institute of Technology, India

S. R. Mani Sekhar, Ramaiah Institute of Technology, India*

Srinidhi H., Ramaiah Institute of Technology, India

K. G. Srinivasa, National Institute of Technical Teachers Training and Research, India

ABSTRACT

The stock market volume and price are active areas of research. Behind every dollar of investment, the customer will be hoping for profit in one or the other way. There is a positive correlation between investor sentiment and stock volume. Predicting the stock market is the most difficult task due to the dynamic fluctuation of volume and price. The traditional analysis methods carried out lead to satisfactory results. In this paper, the proposed system uses real-time data from Twitter to detect the user opinion about the product along with the stock volume for prediction. The stock volume data and the Twitter data are collected first, and then the classification of the polarity is carried out using the SentiWordnet dictionary. The algorithm for the prediction of the stock prices uses long short-term memory, a neural network, as the prices are sequentially evolving in nature. The results of the proposed system are correlated between the stock market and Twitter data to obtain better insights that are positive.

KEYWORDS

LSTM, NB, Real-Time Analysis, Stock Market With LSTM, Stock Prices, Twitter, Twitter Analysis

1. INTRODUCTION

The stock volume and price are the most important objective in the financial world. The stock market is a complex system which considers many components into pictures such as price, volume, trends, stock direction and etc. The accurate prediction will provide the significant profit and helps in yield of the market risk. Predicting the stock volume and price is not an easy task (Loke, 2017). The stock price is considered to be more fluctuating. Billions of dollars are treaded every day and investor's hopes for profit in one way or other.

In order to achieve the performance for establishing the predictive model various methods should be taken care. The people should easily understand the financial field to take a proper decision based on the predicted result. The customer should be able to get more knowledge and better insights should be provided. In order to invest in the stock market, the investor should be able to make a proper decision.

*Corresponding Author

The stock price prediction can be done for the short term which considers even seconds, minutes, days and week. The medium-term prediction can be done for years and long-term prediction can be done for 2 years beyond. There are mainly three types of prediction methodologies (Fan et al., 2014). Fundamental analysis includes company stock details, performance ratios, and validity of stock. Technical analysis is concerned with the chart. It helps in determining the future analysis by gathering historical data.

The volume represents the trade record for the specific period. The period can range from 1min, 1 day, 1 month and 1 year. Mainly there is two key to predicting the stock market they are price and volume. The selection of input plays an important role for prediction of the volume and price. The traditional analysis methods yield in the satisfactory results. The stock market contains various parameters which would help in decision making to invest or not. The data generated is of nonlinear type and are dynamic (Fan et al., 2014). The essential information from the raw data set should be extracted. It mainly consists of five steps they are: collecting the source information, the collected information must be pre-processed to extract the relevant information from the collected data, building the model from the pre-processed information, the key value should be identified, result's should be validated.

The social media sites allow the user to share the individual opinion or sentiments about the particular product or event etc. The fast and effective way of monitoring the public feelings towards the business, brand etc is using sentimental analysis using Twitter data. User shares both positive and negative experience on social media. Twitter helps in determining what is happening at any movement of time, anywhere and anytime in the world.

The model should be built in such a way they identify the hidden patterns. The Long Short Term Memory (LSTM) is kind of recurrent neural network it contains a new structure which is known as a memory cell. The memory cell mainly contains the input, forget gate, output, and neuron. The gates help in providing the interaction with the cell and environment. The purpose of the input is to allow the signal which are incoming or block them. The signals which are incoming will change the memory cell state. The output affects the neurons or it will block it. The forget gate is used to connect the self-recurrent. It helps in predicting the interval and delays in time series. It is used to capture the patterns and it keeps the error for the reverse passes. As LSTM works well with raw time series data and also due to its dynamic nature, the author has incorporated this method in the proposed work.

The rest of the paper is organized as follows: Section II discuss a survey on different methods used in stock price prediction with their drawbacks. Section III illustrates the proposed model. Furthermore, Section IV provides a brief discussion on the achieved results. Finally, section V discusses the conclusion and future work.

2. LITERATURE REVIEW

(Liu., 2018) considers the stock market by correlating the stock trading volume. The feature value is extracted and analysis is done on the stock data prediction model based on the transactions. The advantages and disadvantages of the network have to be understood properly to improve the accuracy, in the proposed work author has considered the real-time twitter data which can improve the accuracy further.

(Bordino et al., 2014) considers the stock market, the volume which has a correlation with stock trading volume. The author considers 2600 stocks which are traded on the New York Stock Exchange (NYSE). The stock volume trading of two-three days was considered which result in higher predictive trading volume. The browsing information is mainly considered by the intuition of theoretical and with the help of test statistical. It is considered for daily granularity. In the proposed work LSTM is used for predicting the events and delay in time series, here a positive correlation was mainly obtained for hourly. It is observed, the prediction decreases power in the strength of correlation when it is moved from daily to hourly.

(Chuang et al., 2009) Chuang the trading volume is used to predicting the stock prices. By analyzing the high and low volumes level is used in signaling the pattern return. The dynamic return is used and helps in decision making and to reduce the risk. First the return volume is considered for future returns and second the high and low volume level mainly contains the returns of future. The model has to be improved which support the other factor those are investigate in the future by considering the index of all individual stock. The relationship between the high and low stock cycles can signal the market crash. The construction should be done in such a way that there should be a relation between the volumes. Proposed approach supports over confidence preposition which will majority related the sentiment of the investors. The dynamic relation of return volume should be observed properly to make indicators easier for investors.

Thus, an extensive research and survey have been done in the field of social media like twitter sentiment analysis and stock market data how to predict the stock market movements and social media by using machine learning algorithms. (Saif et al., 2016) The twitter data using sentiment analysis and stock market next day volume prediction is carried out. These predictive models will further play a major role in recommending the most suitable machine learning algorithms for the given dataset. One of the models studied as a part of the predicting the trading is model factorization and components (Xu & Cohen, 2018). The model is based on the latent factors involved in trading such as volume and price. However, the approach is not based on the sentiments based on the stocks. In the proposed system, this is addressed by using the LSTM and twitter analytics for sentimental analysis. Hence, the prediction of the stock price and the volume can be improved.

One of the work has been carried out by (Rai et al., 2018) where the relationship between the news, tweets and the market behavior is analyzed using the daily news available. In this approach, the daily news is collected using the corpus provided by News API (https://newsapi.org/). The analysis was done using the correlation between the news counts, stock prices and tweet counts with the transaction volume. This approach has motivated in this paper for gathering the data of tweets for price prediction of stocks. However, the data gathering and analysis would not be sufficient to establish the correlation among the entities for stock price prediction. Thus, in the proposed system real time twitter data is consider for computation.

The earlier research has been carried out using the time series approach based on the historical data of trading price and volume. The complex patterns of the market trend are analyzed using the Autoregressive model for time series analysis. However, the applicability of the models based on the time series analysis is limited because of the non-linear nature of the stock prices. In this regard, deep learning methods help in understanding the complex relationships among the stock prices. Some of the works that exploit the deep learning methods for stock price prediction are based on the Recurrent neural networks (RNN) and Long short term Memory (LSTM) networks (Akita et al., 2016) (Rather et al., 2015). Thus, in the proposed system LSTM is used for the stock price prediction.

News data has been explored in analyzing the stock price prediction using the multi-layer dimensionality reduction method (Nassirtoussi et al., 2015). Event embeddings are used as the external source of knowledge and analyze the directional movements and the semantics behind the stocks (Ding et al., 2016). Support vector machines (SVM) are also explored to address the non- linearity of the stock prices by building the tree representation of the predictive models (Xie et al., 2013). A generic stock prediction framework has been developed in (Li et al., 2014) (Nguyen et al., 2013) using the sentimental analysis. The framework is designed for only Chinese stocks. The topics that are related to the stocks are extracted from the social media and sentimental analysis is performed to understand the semantics of the stocks and its prices.

There have been many efforts in determining the stock prices and prediction as discussed above. However, the model should pay attention to quality, comprehensiveness as the analysis of the text affects the model. In the proposed work, the stock price prediction is carried out with the sentimental analysis using the real time tweets with the LSTM model for prediction.
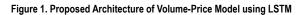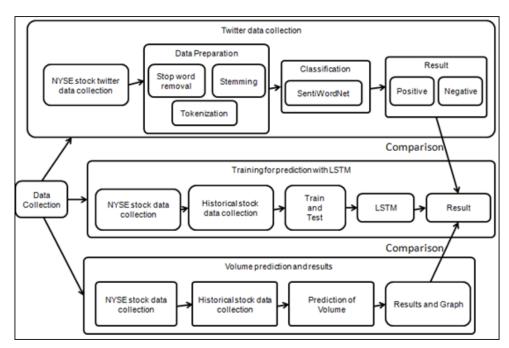
3. PROPOSED SOLUTION

In this section, the proposed system for the volume price model and prediction is discussed. The modules considered for the proposed system are data collection, data pre-processing, classification and prediction as shown in the figure 1. Initially, the data collection of the NYSE is done by using stock market API and twitter data is collected using twitter API. The twitter and stock market data are extracted by using the keyword. It can be apple, Google, Microsoft and etc.

The most challenging task is preprocessing the collected data by removing the hashtags, emoticons, stop words and URLs. Also due to the dynamic nature of price fluctuate and its volume going up and down in nature. The feature extraction of data is performed after processing which is subjected to a much different aspect such as verbs, nouns, and an adjective. The classification is performed by applying polarity and sentiwordnet (CCA-2017) because the real-time twitter data and stock values are dynamically in nature. The historical stock market volume data is collected and the prediction is carried out for the next one day. The artificial neural network algorithm Long Short-Term Memory (LSTM) is applied to stock market data for predicting the price (ANN-P). The results of the classification performed using the twitter and the LSTM are compared to see the accuracy of prediction using the LSTM & Naïve Bayes Classifier (NB).

3.1 Data Collection

In this module, the data is collected from the twitter for the classification purpose. (Twitter API, 2017) (Twitter4j API, 2016) The twitter streaming API is used for the data collection using the twitter. The stock market data is collected using stock market API. The source of tweets is NYSE stock exchange related tweets. The tweets related to apple, Google, Microsoft and etc stock exchange tweets are extracted. The retrieved tweets are converted into JSON file format. The data collected need to pre-processed for the classification purposes further. In the next section, the data pre-processing is discussed.

Figure 1. Proposed Architecture of Volume-Price Model using LSTM

3.2 Data Pre-Processing and Feature Extraction

The data collected from the twitter feeds needs to be pre-processed to remove the unnecessary data. For example, stop words such as is, an, a etc. doesn't add relevance to the polarity of the classification. The other information such as the URL, account id, retweets need to be removed from the tweets. For example, in the tweet "Stock price goes up!! Wonderfullllll!!" the wornderfullllll needs to be cleaned as wonderful. The features in the tweet need to be extracted for determining the polarity. The process is carried out using the sentiwordnet dictionary where the tokenized words in the tweet are compared with sentiwordnet dictionary. The tokens compared are given the label positive or negative based on the polarity in the sentiwordnet dictionary. The classification of the twitter data collected and the score calculation is discussed in the next section.

3.3 Classification

The classification is carried out using the knowledge based approach (Yadav, 2012). Knowledge based is also known as lexicon based. It mainly focuses on the lexicons opinion based from the text. It identifies the polarity of the text. The precompiled sentiment terms are collected by the lexicons. The lexicon approaches are further classified into dictionary based and corpus based. The polarity and setiwordnet are used for classification purpose. The twitter tweets are classified as neutral tweets (NT), positive tweets (PT) and negative tweet (NT) as represented in equation 1 & 2. If tweet score is greater than zero that is one, then the tweet is considered as a positive tweet. If tweet score is less than one it is considered as a negative tweet. The positive percentage of tweets is calculated by positive tweet divided by tweets total number as represented in equation 3. The negative percentage of tweets is calculated by negative tweets divided by tweets total number as represented in equation 4.

The total tweets are calculated by adding all tweets. The terms used in the equation (1), (2), (3) and (4) are defined as follows:

- TP denotes the true positive tweets.
- FP denotes the negative positive tweets (false positive).
- TT denotes the total tweets.

$$s(w) = \begin{cases} 1, if\, w > 0 \\ -1, if\, w < 0 \\ 0, if\, w = 0 \end{cases} \tag{1}$$

$$s(w) = \begin{cases} +\ ,\ t > 0 \\ \\ -,\ t < 0 \end{cases} \tag{2}$$

$$TP = TP/TT \tag{3}$$

$$FP = FP/TT \tag{4}$$

3.4 Long Short Term Memory for Volume Price Prediction

The stock price model should be built correctly in order to attain profit. The prediction can be performed using machine learning models. LSTM is special kind of RNN that has the capability to learn the long dependencies (Hochreiter, & Schmidhuber, 1997; kumar, toorang & bali 2020; bali, kumar, & gangwar, 2020). In a LSTM a cell state is added along with the hidden state to store the long

term memory. They act as the memory blocks in the hidden layer of RNN. The temporal state of the network is maintained in the memory blocks along with gates to control the state of the information. A memory block has an input gate and output gate. The input gate controls the flow of input activations to memory blocks and the output gate handles the flow from memory block to rest of the network.

The LSTM addition in the stock price prediction improves the algorithm to perform better than other standard averaging. The data like open, close, high, low and volume is considered for prediction. The data is split into training and testing set. The training data set is pointed to the data collected from the pipeline of data collection module and the rest is considered as test data. The normalization is performed because the different time periods data will have different value ranging. The model takes previous stock details as input and prediction are carried out. The back propagation is used to optimize the model. The placeholder is set for training the input. The mean square error is calculated, when it goes down it represents the good sign of how well the model is trained.

The historical data of the stock market is collected using the stock market API. The average volume (A[V]) is calculated by adding all volume and dividing it by a total number of volume represented in the equation 5, where V1, V2, V3 are different tweets with various volumes. The data is extracted in real time and the average is calculated. The total tweets (TT) are calculated by adding all tweets which are extracted for the keyword search as represented in the equation 6. The positive tweet percentage (TP) is calculated by equation 6 and Tweet Negative (TN) percentage is calculated by equation 7 & 8. The twitter tweets and stock volume is correlated for prediction purpose. The correlation helps in determining the stock value and identifies the trends as follows:

- If the volume is high there will be movement in stock price.
- Volume is an indication of price movements.
- When volume increases the current trend will increase.
- When the volume is high the buyers interested will improve.
- The low volume will reflect in price trading range.
- The volume decrease will reflect in decrease of trends.
- Volume is typically coupled with the price.

$$A[V]= (V1+V2+V3+…+Vn)/ \text{Total number of tweets} \tag{5}$$

$$TT=t1,t2,t3…..,tn \tag{6}$$

$$TP=Tweetpositive/TT \tag{7}$$

$$TN=Tweet\ negative/TT \tag{8}$$

If the tweet is positive then the average is calculated by the average of the volume with the positive tweet count plus total tweet as represented in the equation 9. If tweets are negative then the average is calculated by the average of the volume with the negative tweet count minus total tweets by the equation 10:

$$S(t)>0=A[V]+TT \tag{9}$$

$$S(T)<0 = A[V]-TT \tag{10}$$

Proposed Algorithm for Stock price prediction

```
Input: Twitter data and stock market volume
Output: Volume prediction
Begin
   Search k in Key list
   if k found then
   for each tweet ∈ T do
    s(t) = classification using sentiwordnet and polarity
    if s(t) > 0 then
      tweet_pos ← s(t)
    else tweet_neg ← s(t)
    Total tweets = t1 + t2 +t3 +…tn
    for each v ∈ Volume
```

$$Avg[v] = \frac{v1 + v2 + v3 + ...vn}{total\ number\ of\ volume}$$

```
    if tweet_pos > tweet_neg
```

$$positive\ tweet_{percentage} = \frac{tweet_{pos}}{Total\ tweets}$$

```
    if positive tweet_percentage > 80%
```

$$do\ A[v] = \frac{Avg[v]}{positive\ tweet_{percentage}}\ then$$

```
     A[v] + Total tweets
     Else A[v] + Total tweets1
    else tweet_pos > tweet_neg then
   Avg[v]-total tweets
End
```

Notations used in the proposed work:

- k is the key search of NYSE stock
- T is the tweets collected for the search {t1, t2, t3….tn}
- S(t) is the score from the classification
- $tweet_{pos}$ is the tweet positive of the twitter data
- $tweet_{neg}$ is the tweet negative of the twitter data
- Avg[v] is the average of volume {v1, v2, v3…….vn}

4. RESULT AND DISCUSSION

Three data sets have been considered for the purpose of implementing stock market volume price prediction. The NYSE are extracted using twitter API, 2800 stocks of NYSE are considered. The tweets are extracted using the keyword. Then tweets are classified using polarity and setiwordnet. The historical stock data is extracted and the data set contains the details such as open which describe the opening price of the stock, closing price of the stock is close, high of the stock is highest price, low represent the low price of stock and volume are total volume traded. The LSTM is used for predicting the stock price, thereafter Naïve Bayes Classifier is used for comparison. The stock volume is extracted from stock market API. The twitter and stock market is correlated for predicting the stock volume.

4.1 Results of Performance Evaluation

Precision is calculated by correct positive prediction observation with respect to total positive prediction observation calculated using precision as shown in Table 1. The formula for computing the precision is in the equation 11. The 0.7238 precision is obtained. It is the ratio of correctly predicted positive observed values with actual observation if the class is yes. The formula for the recall is represented in the equation 12. The recall obtained is around 0.6943. FMeasure is calculated by the equation 13. Accuracy is evaluated by the equation 14:

$$precision = Tp / (Tp+Fp) \tag{11}$$

$$recall = (Tp) / (Tp + Fn) \tag{12}$$

$$FMeasure = 2*(Recall * precision) / (recall + Precision) \tag{13}$$

$$accuracy = Tp+Tn /(Tp+Fp+Fn+Tn) \tag{14}$$

4.2 Comparison of Twitter and Stock Market

Figure 2 shows the Comparison of Twitter and the stock market. The x-axis represents the tweets positive, tweet negative, stock high and stock low. The y-axis represents the percentage. The twitter data consists of tweets which are extracted for the particular keyword. The positive tweets are correlated with stock high and negative tweets are correlated with stock low. When the user opinion is good about the product then tweets are considered to be positive. When the tweets are negative about the stock then it is considered to be negative. The user sentiments play a major role so the comparison is carried out to provide better insight.

Figure 3 historical data graph of volume prediction. The x-axis represents the date information and y-axis displays the volume information. One-month volume of the stock market is extracted to predict volume. The twitter is correlated with stock volume. The volume might go up and down; the line represents the prediction of the stock market. The volume contains the total volume of the trading.
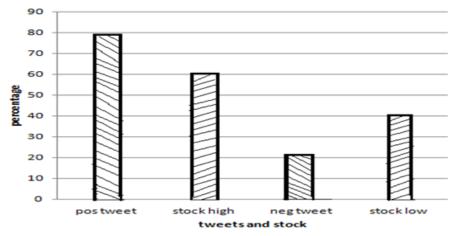
4.3 Stock Price Prediction Using LSTM

In order to predict the stock prices using LSTM, google stock price was collected where the ratio of testing to training set was 80: 20. Figure 4 shows the test samples of the stock price prediction using LSTM. The x-axis represents the time in seconds and y-axis represents the stock price. The stock data is extracted and prediction is carried out. Similarly figure 5, shows the prediction of stock price using Naïve Bayes Classifier. The stock details such as the high, low, date, open, close and volume is considered for price prediction. The red lines indicate the stock price prediction, in some places the stock price is high whereas in some places it goes down, its mainly due to the economics parameter of the market, the same is reflected in the time graph. The graph show that LSTM perform well compared to the NB classifier. Here the proposed model performs very well but still the model

Table 1. Results of performance evaluation metric

Recall	Precision	FMeasure
0.694366	0.7283	0.6825

Figure 2. Comparison of twitter and stock market

Figure 3. Historical data graph of volume prediction

can improve the accuracy further by integrating some of the machine learning methods like CNN, DNN etc. Also here the prediction is carried out for one day which can be further done for weeks.

5. CONCLUSION

The investors rely on the information which is available on the external source. Therefore, the successful prediction will lead to significant profit. Behind every dollar of investment, the customer will be hoping for profit in one or the other way. The prediction is carried out by considering trading data of more than 2800 stocks of NYSE stocks. The classification is carried out using polarity and setiwordnet. The stock market volume is predicted by the twitter and stock volume. The historical volume data is extracted in real time. The long short term algorithm is used for predicting stock price.

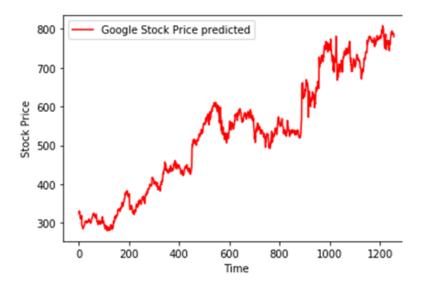**Figure 4. Stock price prediction using LSTM**

Figure 5. Stock price prediction using Naïve Bayes Classifier

The different parameters are correlated. The investor can pick the specific stock of choice from the NYSE list.

The proposed system helps in obtaining the better accuracy by considering both sentimental and technical analysis. Based on the polarity the different sentimental type is derived. The setiwordnet with additional classification is used for sentimental analysis and LSTM algorithm is used for stock market price prediction. The various performance metrics are evaluated like recall, precision, FMeasure and high accuracy is obtained. The twitter data and stock market data is compared to gain better knowledge about the stock the investor is investing. Later, the model is compared with NB and the result shows that the proposed work performs better then

NB. Stock market prediction seems to be a more challenging task. Since is it more fluctuating and price and volume go up and down dynamically in nature. In future, the accuracy of the model can further be enhanced by using other machine learning algorithm and deep learning approaches can be used such as CNN for prediction. The prediction is carried out for one day it can be done for weeks. The volume is manually calculated and it can be performed by using some machine learning algorithms.

FUNDING AGENCY

REFERENCES

Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K. (2016). Deep learning for stock prediction using numerical and textual information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, IEEE. doi:10.1109/ICIS.2016.7550882

Artificial Neural Network- Perceptron. (2017). http://www. Saedsayad. com/ artificial_ neural_ network_bkp.htm

Bali, V., Kumar, A., & Gangwar, S. (2020). A Novel Approach for Wind Speed Forecasting Using LSTM-ARIMA Deep Learning Models. *International Journal of Agricultural and Environmental Information Systems*, *11*(3), 13–30. doi:10.4018/IJAEIS.2020070102

Bordino, I., Kourtellis, N., Laptev, N., & Billawala, Y. (2014). Stock trade volume prediction with Yahoo Finance user browsing behavior. In *2014 IEEE 30th International Conference on Data Engineering*. IEEE. doi:10.1109/ICDE.2014.6816733

CCA. (2017). *Classification and Clustering Algorithm*. http://dataaspirant.com/2016/09/24/ classification-clustering-algorithms/

Chuang, W. J., Ou-Yang, L. Y., & Lo, W. C. (2009). Dynamic return-volume relation and future returns-implication for reducing investing risk. In *2009 International Association of Computer Science and Information Technology-Spring Conference*. IEEE. doi:10.1109/IACSIT-SC.2009.37

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2016). Knowledge-driven event embedding for stock prediction. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2133-2142.

Fan, M. H., Liao, E. C., & Chen, M. Y. (2014). A TAIEX forecasting model based on changes of keyword search volume on Google Trends. In *2014 IEEE International Symposium on Independent Computing (ISIC)*. IEEE. doi:10.1109/INDCOMP.2014.7011756

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735 PMID:9377276

Kumar, A., Toorang, P., & Bali, V. (2020). User Based Load Visualization of Categorical Forecasted Smart Meter Data using LSTM Network. *International Journal of Multimedia Data Engineering and Management*, *11*(1), 30–50. doi:10.4018/IJMDEM.2020010103

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, *69*, 14–23. doi:10.1016/j.knosys.2014.04.022

Liu, S., Liao, G., & Ding, Y. (2018). Stock transaction prediction modeling and analysis based on LSTM. In *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA).* IEEE.

Loke, K. S. (2017). *Impact of Financial Ratios and Technical Analysis on Stock Price Prediction Using Random Forests.* IEEE. doi:10.1109/ICONDA.2017.8270396

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2015). Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *Expert Systems with Applications*, *42*(1), 306–324. doi:10.1016/j.eswa.2014.08.004

Post, retrieve and engage with Tweets. (2017). https://developer.twitter.com/en/docs/tweets/post-and-engage/ overview

Rai, B., Kasturi, M., & Huang, C. Y. (2018). Analyzing Stock Market Movements Using News, Tweets, Stock Prices and Transactions Volume Data for APPLE (AAPL), GOOGLE (GOOG) and SONY (SNE). In *Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence*. ACM. doi:10.1145/3243250.3243263

Rather, A. M., Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications*, *42*(6), 3234–3241. doi:10.1016/j.eswa.2014.12.003

Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management*, *52*(1), 5–19. doi:10.1016/j.ipm.2015.01.005

Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., & Deng, X. (2013). Exploiting topic based twitter sentiment for stock prediction. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 24-29.

Twitter4j API. (2016). http://twitter4j.org/en/code-examples .html

Xie, B., Passonneau, R., Wu, L., & Creamer, G. G. (2013). Semantic frames to predict stock price movement. *Proceedings of the 51st annual meeting of the association for computational linguistics*, 873-883.

Xu, Y., & Cohen, S. B. (2018). Stock movement prediction from tweets and historical prices. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 1970-1979. doi:10.18653/v1/P18-1183

Yadav, S. K., & Pal, S. (2012). *Data mining: A prediction for performance improvement of engineering students using classification*. arXiv preprint arXiv:1203.3832.