Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES-2021)
IEEE Xplore Part Number: CFP21AWO-ART; ISBN: 978-0-7381-1405-7

Comparative Study of Stock Price Prediction using Machine Learning

1st Parag P. Kadu
Dept. Computer Science &Engg
P.R.M.I.T & R, Badnera
Amravati, India
ppkadu@mitra.ac.in

2nd Dr. G. R. Bamnote
Dept. Computer Science &Engg
P.R.M.I.T & R, Badnera
Amravati, India
grbamnote@mitra.ac.in

2021 6th International Conference on Communication and Electronics Systems (ICCES) | 978-1-6654-3587-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCES51350.2021.9489170

***Abstract*—The stock market is mainly an aggregation of different sellers and buyers of stock. "A stock (also known as shares more commonly) in general represents ownership claims on business by a particular individual or a group of people". The effort to find out the upcoming stock market value is recognized as a stock market prediction. The forecast is anticipated to be efficient, accurate and robust. The system should function based on the real-life circumstances and it have to be well-matched to realistic surroundings. The system is also expected to consider the entire constraints, which may have an effect on the stock's value and performance. In stock market the selection of stock is very important for trading and investment, if fundamentally the stock is good but if the sector of that stock is down then ultimately the one will lose its trade for buy position. That's why there is need to continuously know the updated knowledge of current situation, News that has direct or indirect impact on stock or sector. The recent example is due to covid-19 pandemic the auto sector was underperforming as the sell-value of vehicles are dropped down and the pharmacy sector is performing well due to covid-19 medicine. During past years many researchers have given contribution in this field but there is still a need to do the research for stock selection based on fundamental and technical analysis, this provides strong motivation for a system that can efficiently extract data from different web sources can be done using web crawling and capable of prediction of stock price, which would be helpful for individual for selection of stock for trading and investment. Research gaps and challenges between existing techniques are listed and detailed, which helps researchers to upgrade future works. Here, this paper provides an overview of the applications of machine learning in stock market forecasting to determine what can be done in the future.**

Keywords*—Stock Price Prediction, Machine Learning, Sentiment Analysis, Web Crawling, Convolutional Neural Network, Feature Optimization, Gradient Descent Boosting, Random Forest, Support Vector Machine, Artificial Neural Network.

I. INTRODUCTION

In recent times, predicting the stock market return is attaining much consideration, due to the fact that if the market direction is predicted successfully, the shareholders might be guided in a better way [1]. The productivity of trading and investing in the stock market to a huge extent is based on the predictability. If a technique is established that constantly predicts the movements of the dynamic stock market, it could make the proprietor wealthier. In addition, the predicted movements of the share market will assist the supervisory body of the market in formulating counteractive measures [3].

A further inspiration for analysis in this domain is that it holds numerous experimental and theoretical challenges. EMH (Efficient Market Hypothesis) is the most commonly accepted investment theory in economics. It is stated that all the related data about the stock will be reflects in the stock's prices in the market and the current value of stock cannot be affected by the past value [6]. As a result, the perfect mediation is maintained for the stock market and having no possibility for the cost on research, enthusiastic profits and trading are considered as the consumption of energy and resource [10]. Many individuals pay out certain time daily, obtaining the data, analysing that data and deciding to invest based on the obtained data in the investment management industry. Therefore, the stock market price can be analysed by two methods such as sentiment analysis and by the mathematical indicator [16, 17].

Sentiment analysis is the process of analysing the financial data which is available in the social media like news articles and tweets on the basis of stock prices [28]. The sentiments from news article and twitter over a period are extracted using the feature extraction process. The sentiments can deserve short-term market fluctuation that may cause disconnection between the true value and the price of the company shares. The extracted features are compared with the change in stock price for the identical period for a particular firm [29].

In the sentiment dictionary, there are various positive and negative words are available. These positive and negative words are formed the features of the feature vector in the sentiment dictionary. In machine learning technique, these feature vectors are useful to predict accurately the sentiment of the news feed and in each news feed, the feature has a particular value which is based on the positive and negative word as well as the frequency of the word in the article [31]. Another process for predicting the stock market price is the technical indicators. While considering Indian stock market, it has mainly two indicators or indexes such as Sensex and Nifty. The performance of all the stock markets are evaluated by the index. In Sensex, there are totally thirty stocks that return the whole sentiment of big companies and in Nifty contains fifty stocks. Several other indexes also available such as automobile index, real estate, bank index, media index, IT index, etc [33]. In general, the technical indicators are utilized by the professional traders to analyse stocks value and decisions making. However, it's always to find that buying and selling are created among various selected indicators. At a particular time, traders have to make a decision according to their personal experience or decision models [34, 35].

II. LITERATURE SURVEY

In 2019, Ehsan and Saman [1] have established a novel "Convolutional Neural Network (CNN)-oriented approach, which could be deployed on a set of data attained from diverse markets for mining the features for future market prediction. The presented scheme was exploited for predicting the movement of the indices of NASDAQ, DJI, RUSSELL, S&P 500, and NYSE depending on a variety of initial parameters. Experimentations have demonstrated that the outcomes attained by the presented approach were better over the other schemes in terms of prediction".

In 2019, Zhao [2] have suggested "a framework that brought out the variations among Baidu Index and Baidu News as an alternative for private information. Here, private information has displayed negative association with volatility and turnover. In addition, online searches and news has offered a solid predictive power for indicators. On carrying out evaluations, the private information has revealed better volatility with consistent predictability. Finally, the performance of the presented scheme was examined, and the stock performance of the established scheme was found to be better than the other conventional schemes".

In 2018, Zhang et al. [3] have "established an approach that extracted the events from user sentiments and Web news for investigating the mutual impacts on the movement of stock price by means of a tensor factorization and coupled matrix model. Particularly, a tensor was constructed initially for fusing the diverse data and the basic associations amongst the investors' sentiments and events were captured. Moreover, for assisting the decomposition of tensors, the correlation matrix and feature matrix were formulated and incorporated. Further, the analysis outcomes have revealed the betterment of the presented prediction model in a precise way".

In 2019, Samui and Kar [4] have presented "a hybridized model for predicting the stock price time series based on fuzzistics, in which cumulative probability distribution approach (CPDA) was exploited for attaining the linguistic value intervals. In addition, fuzzy rule generation was exploited along with rough set theory for reducing the sets of rules. Subsequently, time series data was predicted from defuzzification by means of minimized rule base and the corresponding historical facts. Computations have revealed the capable efficacy and utility of the established framework in predicting the stock markets".

In 2019, Idrees et al. [5] have presented a novel scheme for predicting the stocks in the future. "The stock market is considered to be one of the most highly complex financial systems which consist of various components or stocks, the price of which fluctuates greatly with respect to time". The aim of every investor was to reduce the related risks and to increase the returns on their investments. Here, the adopted scheme has examined the "Indian stock market (ISM)" and a statistical model was formulated for predicting the upcoming stocks efficiently.

In 2019, Liu and Wang [6] have suggested a numerical-based attention (NBA) technique for predicting the stock market using dual sources. As the initial step, an attention-oriented technique was proposed for effectively exploiting the complementarities among numerical data in forecasting the stock prices. Accordingly, the stock trend information concealed in the news was converted into mathematical data and was subjected to encoding. The adopted scheme has thus filtered the noise efficiently and it has exploited the entire trend information in news.

In 2019, Lee et al. [7] have established a novel method, where "stock chart images were considered as input for making predictions on global stock market. Consequently, Deep Q-Network was applied with a CNN function for carrying out the intended operations. The implemented approach has not only attained better profit in the stock markets of the country, but it has also attained increased profits in global stock markets. Experimentation outcomes demonstrate that the prediction of stock prices was better even while analysing the data from diverse countries".

In 2016, Ramiro and Agustín [8] have presented an approach that authenticated a series of predictive approaches using traditional topic discovery and machine learning techniques. The experimentations were simulated and trained with diverse mixtures of extracted features from the previous stock price values. In addition, it was discovered that incorporating this information could enhance the classification performance that were trained only on technological indicators. Furthermore, high analytical features obtained from the message board data were found to have relevant and significant semantic contents.

In 2019, Das et al. [9] have proposed "the feature optimization capacity of firefly with an evolutionary structure that evaluated the social factors of Firefly algorithm combined with the selection process of unbiased value. OSELM and firefly method has been used particularly in stock market prediction and it is totally focused on forecasting of the future value of the stock market price. OSELM is the application oriented and the technical indicators as well as statistical measures are used to regenerate the dataset based on the obtainable features such as maximum price, minimum price, open price, change of price and close price. The firefly and OSELM approach provided better prediction ability".

In 2019, Nam, K. and Seong, N.[18] depicted "a novel machine learning model called Multiple Kernel Learning (MKL) for the prediction of stock price movement which depends on the financial newsfeed considered mainly on causality. The normal relationship among the companies is analyzed using transfer entropy and the Global Industry Classification Standard sectors. MKL is utilized for the combination of features of target companies and normal companies. Identification of causal relationship is significant in prediction problems, and it is also important to find out the connections with the difficult model theory. Context-aware text mining based on the company-specific financial news is utilized for the better performance".

In 2019, Derakhshan, A. and Beigy.[27] H suggested a novel "predictive modeling structure for analyzing the future values of the stock markets. The fluctuation of the financial markets is chaotic, volatile and also not linear in nature so it is very difficult for predicting the stock market. A three-stage method is proposed to overcome the drawbacks of the stock markets. Fractal modeling and recurrence analysis are utilized for the first step and EMH is tested to perceive the temporal behavior of the financial market to investigate the properties of autoregressive. Granger causality tests are used in the next step in a vector auto regression environment for the investigation of the causal interaction structures between the

indexes. Finally, the maximal overlap discrete wavelet transformation is performed to perish the stock indexes into linear and nonlinear subcomponents".

In 2019, Ramos-Pérez et al[30]proposed "a model based on a set of Machine Learning approaches namely GDB (Gradient Descent Boosting), RF (Random Forest), SVM (Support Vector Machine) and ANN (Artificial Neural Network) and these algorithms are stacked to forecast S&P500 volatility. The proposed model depended wholly on the forecasting made by market data and the machine learning algorithms. Moreover, in particular of the Stacked-artificial neural network the final prediction done by the high-level algorithm that are directly utilized as inputs within the artificial neural network".

In 2019, Bernabé-Moreno, et al [32] proposed "a novel approach termed as binned corpus for the extraction of a polarity dictionary automatically from a specified domain or a stock market without individual intervention and the issues of scaling and thresholding are addressed. Binned corpus traces the value, which is changed of specified stocks over time, using it as a guiding polarity price. The enormity of the price variation for a particular stock is then assigned to the economic newsfeed regarding this stock in a particular period and that is called as the working corpus. The TF–IDF data retrieval approach is used to measure the value of TF–IDF for every term. These prices are distributed within the nearby of every term based on the embeddings-enabled cosine distance".

III. Recent Advances

The Indian stock exchanges (ISE) have attained much significance globally. One among the older exchange across the world is "Bombay Stock Exchange (BSE)", whereas the National Stock Exchange (NSE) is said to be the most excellent one with respect to advancement and sophistication of technology. Stock market investment involves high gains and high risks and therefore, it attracts huge quantity of economists and investors. On the other hand, information concerning a stock is usually, complex, vague, incomplete and uncertain, thus formulating it a challenge to forecast the upcoming financial performance. Money is invested in the stock market depending on some investigation. Prior to the computer era, individuals carry out trading in commodities and stocks depending on their gut opinions. As the degree of trading and investing developed, individuals searched for methods and tools, which would raise their gains while reducing their threat. Globally, trade in the stock market has attained huge fame and it turn out to be the part of everyday practice for several people to harvest fine profits. Nevertheless, the movement of stock price prediction becomes challenging owing to the complexity of data in the stock market. Although examining stock movement performance is a demanding task, the robust predictive modeling could lead a shareholder in segmenting and identifying higher performance securities and as a result, better investment resolutions can be taken. Linear regression, fundamental, technical and statistics analysis are exploited for predicting the market's direction. Anyhow, these methods fail to be the consistent accurate prediction model. Current developments in soft computing methods present constructive tool for examining the movement of stock markets and stock prices for retrieving information, which might direct investors on when to sell and buy.

In stock market the situation may changes at an instant period so the individuals carefully observe this news feed from the articles, twitter, etc and invest at a right time. The information collected from the public are extracting and analysing by the investors to predict the present situation of the stock market. Several research are undertaken for the extraction and analysis of data automatically so that the time and power of the individual can minimize and for ease of predicting the current stock market trends. For analysing, various learning methods are utilized such as deep learning, time series prediction and machine learning. In machine learning, the sentiment analysis and mathematical indicators are the two powerful methods helpful for analysing the data accurately. In various papers, the researchers compared their results with some other analytical tools and these findings give motivation for research in this field for predicting the market value. Other reason for this research is to reduce the risk low and to enhance the maximum profit for the investors. This research domain is mainly focus on the sentiment analysis and the mathematical indicator using the machine learning method.

IV. Features and Challenges

Table 1 show the reviews on the stock market prediction systems. At first, CNN model was introduced in [1], which presents improved F-measure, and it also extracts high level features. However, it has to focus more on the trading systems. Regression model was exploited in [2] that provide improved volatility and it also offers positive predictive capacities, but it needs deliberation on portfolio management. Pearson Correlation coefficient was deployed in [3] that make better stock movement prediction and it also offers improved accuracy. Anyhow, it needs consideration on the psychological biases. Likewise, CDPA model was exploited in [4], which offers reduced MSE, and it contains minimal number of rules. However, it has to focus more on the single centroid mean. Also, Box-Jenkins model was employed in [5], which proffers better volatility and it also minimal prediction error; however, it needs consideration on time series data. NBA framework was exploited in [6] that filter the noise effectively and it also minimizes the noise, anyhow, it needs consideration on index level data. CNN model was introduced in [7], which offers optimal robustness, and it also offers minimal transaction costs. However, it has to focus more on the "Efficient Capital Markets". At last, Random Forest (RF) model was deployed in [6] that provide improved accuracy and it also offers better true positive rate (TPR), anyhow, it needs consideration on online events.

TABLE 1: REVIEW ON CONVENTIONAL STOCK MARKET PREDICTION SYSTEMS

Author	Adopted models	Features	Challenges
Ehsan and Saman [1]	CNN	❖ Extracts high level features. ❖ Enhanced F-measure.	❖ Have to focus more on the trading systems.
Zhao [2]	Regression model	❖ Positive predictive capacities. ❖ Improved volatility.	❖ Requires consideration on portfolio management.
Zhang *et al.* [3]	Pearson Correlation coefficient	❖ Better stock movement prediction. ❖ Improved accuracy.	❖ Needs consideration on the psychological biases.
Samui and Kar [4]	CDPA	❖ Reduced MSE. ❖ Minimal number of rules.	❖ Have to focus more on the single centroid mean.
Idrees *et al.* [5]	Box-Jenkins model	❖ Better volatility. ❖ Reduced prediction error.	❖ Time series data has to be concerned more.
Liu and Wang [6]	NBA framework	❖ Filters the noise effectively ❖ Minimal noise.	❖ Requires more consideration on index level data.
Lee *et al.* [7]	CNN	❖ Optimal robustness. ❖ Reduced transaction costs.	❖ Have to focus more on the "Efficient Capital Markets"
Ramiro and Agustín [8]	Random forest (RF)	❖ Improved accuracy ❖ Better true positive rate	❖ Requires more consideration on the online events

V. CONCLUSION

"Predicting stock market is basically defined as trying to determine the stock value and offer a robust idea for the people to know and predict the market and the stock prices". It is usually offered by means of the periodical financial proportion using the dataset. Therefore, one dataset is not enough for prediction, which may lead to imprecise results. The estimation of stock price continues to be an issue if an enhanced stock market prediction mode is not introduced. The stock market movement is generally portrayed by the sentiments of several investors. Predicting the stock market, calls for a capability to forecast the consequence of current actions on the investors. These actions could be political measures (like a proclamation by a political leader), a piece of reports on scam etc. All these actions have an effect on the corporate earnings that may affect the attitude of investors. These entire aspects make the prediction of stock price very complex. Once the exact data is gathered, it can be deployed for training a machine and a better predictive outcome can be generated.

Stock market speculation is an old and important issue. With an effective stock forecasting model, it can gain an understanding of market behavior over time, to see trends that would not otherwise be seen. Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in prediction - physical, psychological, rational and irrational factors, etc. All of these factors combine to create fixed prices and make it extremely difficult to predict with the highest accuracy. Stock markets are affected by a lot of uncertainty and related economic and political factors at local and global levels. The key to successful stock market predictions is getting the best results with the minimum required input data. Finding the right set of factors to make accurate forecasts is a daunting task and therefore a regular stock market analysis is very important.

REFERENCES

[1] Ehsan Hoseinzade, Saman Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables", Expert Systems with Applications, vol. 129, pp. 273-285, 1 September 2019.

[2] Ruwei Zhao, "Inferring private information from online news and searches: Correlation and prediction in Chinese stock market", Physica A: Statistical Mechanics and its Applications, vol. 528, 15 August 2019, Article 121450.

[3] Xi Zhang, Yunjia Zhang, Senzhang Wang, Yuntao Yao, Philip S. Yu, "Improving stock market prediction via heterogeneous information fusion", Knowledge-Based Systems, vol. 143, pp. 236-247, 1 March 2018.

[4] Shanoli Samui Pal, Samarjit Kar, "Time series forecasting for stock market prediction through data discretization by fuzzistics and rule generation by rough set theory", Mathematics and Computers in Simulation, vol. 162, pp. 18-30, August 2019.

[5] S. M. Idrees, M. A. Alam and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," IEEE Access, vol. 7, pp. 17287-17298, 2019.

[6] G. Liu and X. Wang, "A Numerical-Based Attention Method for Stock Market Prediction With Dual Information," IEEE Access, vol. 7, pp. 7357-7367, 2019.

[7] J. Lee, R. Kim, Y. Koh and J. Kang, "Global Stock Market Prediction Based on Stock Chart Images Using Deep Q-Network," in IEEE Access, vol. 7, pp. 167260-167277, 2019.

[8] Ramiro H. Gálvez, Agustín Gravano, "Assessing the usefulness of online message board mining in automatic stock prediction systems", Journal of Computational Science, vol. 19, pp. 43-56, March 2017.

[9] Smruti Rekha Das, Debahuti Mishra, Minakhi Rout, "Stock market prediction using Firefly algorithm with evolutionary framework optimized feature reduction for OSELM method", Expert Systems with Applications: X, vol. 4, November 2019, Article 100016

[10] Kang Zhang, Guoqiang Zhong, Junyu Dong, Shengke Wang, Yong Wang, "Stock Market Prediction Based on Generative Adversarial Network", Procedia Computer Science, vol. 147, pp. 400-406, 2019.

[11] Eunsuk Chong, Chulwoo Han, Frank C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies", Expert Systems with Applications, vol. 83, pp. 187-205, 15 October 2017.

[12] Feng Zhou, Hao-min Zhou, Zhihua Yang, Lihua Yang, "EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction", Expert Systems with Applications, vol. 115, pp. 136-151, January 2019.

[13] Huiwen Wang, Shan Lu, Jichang Zhao, "Aggregating multiple types of complex data in stock market prediction: A model-independent framework", Knowledge-Based Systems, vol. 164, pp. 193-204, 15 January 2019.

[14] Mu-Yen Chen, Chien-Hsiang Liao, Ren-Pao Hsieh, "Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach", Computers in Human Behavior, vol. 101, pp. 402-408, December 2019.

[15] Ruwei Zhao, "Inferring private information from online news and searches: Correlation and prediction in Chinese stock market", Physica A: Statistical Mechanics and its Applications, vol. 528, 15 August 2019, Article 121450.

[16] Xi Zhang, Jiawei Shi, Di Wang, Binxing Fang, "Exploiting investors social network for stock prediction in China's market", Journal of Computational Science, vol.1 28, pp. 294-303, September 2018.

[17] Yujin Baek, Ha Young Kim, "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module", Expert Systems with Applications, vol. 113, pp. 457-480, 15 December 2018.

[18] KiHwan Nam, NohYoon Seong, "Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market", Decision Support Systems, vol. 117, pp. 100-112, February 2019.

[19] Hiransha M, Gopalakrishnan E.A., Vijay Krishna Menon, Soman K.P., "NSE Stock Market Prediction Using Deep-Learning Models", Procedia Computer Science, vol. 132, pp. 1351-1362, 2018.

[20] Andrew Sun, Michael Lachanski, Frank J. Fabozzi, "Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction", International Review of Financial Analysis, vol. 48, pp. 272-281, December 2016.

[21] Xu Gong, Boqiang Lin, "Structural changes and out-of-sample prediction of realized range-based variance in the stock market", Physica A: Statistical Mechanics and its Applications, vol. 494, pp. 27-39, 15 March 2018.

[22] Melek Acar Boyacioglu, Derya Avci, "An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange", Expert Systems with Applications, vol. 37, no. 12, pp. 7908-7912, December 2010.

[23] Lili Wang, Zitian Wang, Shuai Zhao, Shaohua Tan, "Stock market trend prediction using dynamical Bayesian factor graph", Expert Systems with Applications, vol. 42, no. 15–16, pp. 6267-6275, September 2015.

[24] C. M. Anish, Babita Majhi, "Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis", Journal of the Korean Statistical Society, vol. 45, no. 1, pp. 64-76, March 2016.

[25] Rajashree Dash, P. K. Dash, Ranjeeta Bisoi, "A differential harmony search based hybrid interval type2 fuzzy EGARCH model for stock market volatility prediction", International Journal of Approximate Reasoning, vol. 59, pp. 81-104, April 2015.

[26] Seyedali Mirjalili, Andrew Lewis, "The Whale Optimization Algorithm", Advances in Engineering Software, vol. 95, pp. 51-67, May 2016.

[27] A. Derakhshan and H. Beigy, "Sentiment analysis on stock social media for stock price movement prediction", *Engineering Applications of Artificial Intelligence*, *85*, pp.569-578, 2019.

[28] A. Pathak, N.P. Shetty, "Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis", *In Computational Intelligence in Data Mining,* pp. 595-603, Springer, Singapore,2019.

[29] C. Lohrmann, P. Luukka, "Classification of intraday S&P500 returns with a Random Forest", *International Journal of Forecasting*, *35*(1), pp.390-407,2019.

[30] E. Ramos-Pérez, P.J. Alonso-González and J.J. Núñez-Velázquez, "Forecasting volatility with a stacked model based on a hybridized Artificial Neural Network", *Expert Systems with Applications*, *129*, pp.1-9, 2019.

[31] Feuerriegel, S. and Gordon, " News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions", *European Journal of Operational Research*, *272*(1), pp.162-175, 2019.

[32] J. Bernabé-Moreno, A. Tejeda-Lorente, J. Herce-Zelaya, C. Porcel and E. Herrera-Viedma, "A context-aware embeddings supported method to extract a fuzzy sentiment polarity dictionary", *Knowledge-Based Systems*, pp.105236, 2019.

[33] K.K.Kulkarni, A.D. Kalro, D. Sharma and P. Sharma, "A typology of viral ad sharers using sentiment analysis", *Journal of Retailing and Consumer Services,* 2019.

[34] M. Zhang, X. Jiang, Z. Fang, Y. Zeng and K. Xu, "High-order Hidden Markov Model for trend prediction in financial time series", *Physica A: Statistical Mechanics and its Applications*, *517*, pp.1-12, 2019.

[35] Q. Xu. Wang, W. H. Zheng, "Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles", *Neurocomputing*, *299*, pp.51-61, 2018.

[36] S. Basak, S. Kar, S. Saha, L. Khaidem and S.R. Dey, "Predicting the direction of stock market prices using tree-based classifiers", *The North American Journal of Economics and Finance*, *47*, pp.552-567, 2019.

[37] S. Das, R. Demirer, R. Gupta and S. Mangisa, "The effect of global crises on stock market correlations: Evidence from scalar regressions via functional data analysis", *Structural Change and Economic Dynamics*, *50*, pp.132-147,2019.

[38] S. Kelly and K. Ahmad, "Estimating the impact of domain-specific news sentiment on financial assets", *Knowledge-Based Systems*, *150*, pp.116-126, 2018.

[39] S. Symeonidis, D. Effrosynidis and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis", *Expert Systems with Applications*, *110*, pp.298-310, 2018.

[40] T.K.Lee, J.H. Cho, D.S. Kwon and S.Y. Sohn, "Global stock market investment strategies based on financial network indicators using machine learning techniques", *Expert Systems with Applications*, *117*, pp.228-242, 2019.