Utilizing Machine Learning to Reassess the Predictability of Bank Stocks

Hera Antonopoulou [1], Leonidas Theodorakopoulos [1], Constantinos Halkiopoulos [1*],
Vicky Mamalougkou [1]

[1] Department of Management Science and Technology, University of Patras, Greece.

Abstract

Objectives: Accurate prediction of stock market returns is a very challenging task due to the volatile and non-linear nature of the financial stock markets. In this work, we consider conventional time series analysis techniques with additional information from the Google Trend website to predict stock price returns. We further utilize a machine learning algorithm, namely Random Forest, to predict the next day closing price of four Greek systemic banks. *Methods/Analysis*: The financial data considered in this work comprise Open, Close prices of stocks and Trading Volume. In the context of our analysis, these data are further used to create new variables that serve as additional inputs to the proposed machine learning based model. Specifically, we consider variables for each of the banks in the dataset, such as 7 DAYS MA,14 DAYS MA, 21 DAYS MA, 7 DAYS STD DEV and Volume. One step ahead out of sample prediction following the rolling window approach has been applied. Performance evaluation of the proposed model has been done using standard strategic indicators: RMSE and MAPE. *Findings*: Our results depict that the proposed models effectively predict the stock market prices, providing insight about the applicability of the proposed methodology scheme to various stock market price predictions. *Novelty /Improvement*: The originality of this study is that Machine Learning Methods highlighted by the Random Forest Technique were used to forecast the closing price of each stock in the Banking Sector for the following trading session.

1- Introduction

Financial markets are known for their volatile, dynamic, and non-linear nature [1]. Accurate prediction of stock prices is an extremely challenging task because of factors such as political conditions, the global economy, unexpected events, market anomalies, relevant companies' features, etc. [2]. Considering that financial markets are linked to a country's economic conditions, attracting major funds from investors and issuing equities in the public interest, it is, therefore, crucial to forecast the movement of the stock prices in order to prevent excessive losses and make relevant investment decisions [3]. Traditionally, two main approaches have been widely employed in stock price predictions [4]. These approaches include technical analysis methods that utilize historical stock prices in order to predict future prices [5].

Recently, Google Trends have been considered a useful and effective tool for forecasting economic indicators as well as further predicting political outcomes and, more recently, cryptocurrencies returns and volatility. The Google Trends Service is a free, cross-platform service from Google that analyzes a huge volume of searches and provides data on user search terms based on geographic and temporal patterns, that is, data about the volume of the user's search terms (search volume). Investors tend to gather information mainly through the internet (social networks, specialized forums, etc.). As

---

* **CONTACT**: halkion@upatras.gr

a result, Google searches tend to form a sign of investors' attention. Specifically, Google Trends gives us access to the number of actual search requests in the Google search engine, allowing us to measure people's interest in a particular topic in several languages and regions worldwide.* In our study, we focus on Greek systemic banks since, in recent years, there has been an increasing interest in the performance of the Greek and European banking sectors. The four banks considered are Alpha Bank, Eurobank, National Bank, and Piraeus Bank. In Figure 1, we present Google Trends indices for the National Bank and Piraeus Bank to further motivate the importance of the inclusion of such information to predict stock market returns of the four banks in our analysis, as it is evident that the online interest in the two leading banks in Greece has dramatically increased.
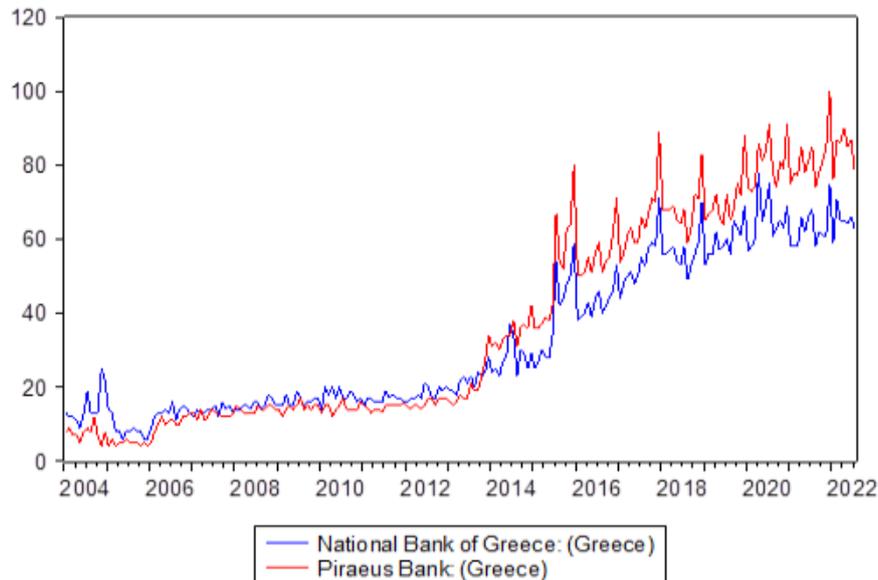
Figure 1. Online interest's evolution for the Greek banks

The present paper aims to explore to what extent investors' attention to the Greek banking sector is captured by a set of keywords measured by Google Trends. To this end, we further utilize in our analysis external factors regarding information based on Google Trends data concerning overall market conditions and macroeconomic issues. As highlighted in the related literature, such information is expected to be crucial to better predicting the next day's closing price of the bank series as well.

Additionally, we contribute to the literature by proposing, in the context of our analysis, the use of a machine learning algorithm, namely Random Forests, to estimate out-of-sample predictions of stock returns for each of the four Banks of the Greek Banking sector. From a financial point of view, we make use of a set of financial variables based on intraday data with (i) Open stock price, (ii) High stock price, (iii) Low stock price, and (iv) Close stock price of a particular Greek systemic bank. Finally, we evaluate the predictive performance of the proposed models considering widely used evaluation indicators in the existing literature: (i) RMSE and (ii) MAPE. Our results indicate that the proposed methodology can capture investors' attention in the Greek banking sector. Furthermore, in line with previous results presented in the related literature, Google Trends variables in our forecasting models have additional predictive value for the out-of-sample prediction of the stock returns of each of the four banks considered in the analysis.

2- Literature Review

Forecasting stock market prices has been extensively studied in the financial related literature. Earlier studies in this area mainly explored classic algorithms such as linear and non-linear regressions, random walk methods, moving average convergence/divergence, and several traditional linear econometrics models such as Autoregressive Integrated Moving Average (ARIMA) family models to predict stock prices [6]. Recently, many of the abovementioned traditional econometric methodologies have been modified following approaches that consider several machine learning, and deep learning techniques. The usage of machine learning as opposed to linear regression methodologies relies on the fact that machine learning algorithms offer the ability to flexibly incorporate many features as additional predictors in the regression models while at the same time not imposing assumptions on the functional form of how signals indicate market movements.

---

* We gathered Google Trends daily data for specific keywords concerning the Greek Bank sector. A search has been conducted in the Greek language. To conclude on keyword selection, the online interest in the below variations is examined, and the variations are further compared, i.e., "Pireaus Bank"; "National Bank (search term)"; "Alpha Bank (search term)"; "Eurobank (search term)"; "Greek Dept (search term)". Only "Pireaus Bank" and "National Bank (search term)" yield, considerably increasing online interest. Among the above, i.e., the topic (Greek banks) and the search term, "economic uncertainty" is selected for further analysis. Finally, Google Trends data are retrieved in .csv format and normalized over the examined period.

Additionally, the use of machine learning algorithms has been found to produce more reliable and accurate results when compared to standard econometric methods, as they can capture complex and nonlinear interactions between the data. To this end, more sophisticated machine learning methods have been utilized to forecast stock prices that follow either technical or fundamental analysis, considering the nature of the data in terms of data size and nonlinearities in the dataset [7]. In order to account for such data characteristics, machine learning and deep learning algorithms can effectively identify complex relations and hidden patterns in such large data sets [8]. The most frequently used in the literature on forecasting stock market price returns, machine learning techniques include SVM, Neural Networks, KNN, and Random Forests.

In their work, Atsalakis et al. (2019) explore the sensitivity of stock prices to external conditions, including daily quotes of commodity prices such as gold, crude oil, natural gas, corn, and cotton, with the results indicating that logistic regression performed better [9]. Zhang (2013) analyzed stock data containing daily stock information ranging from 2008 to 2013 [10]. Various algorithms were explored in a one-step ahead and multi-step ahead forecasting exercise of stock prices. Results revealed that the SVM machine learning technique produced better results in the long-term, considering it had the highest accuracy score among the three models explored (Logistic Regression, Quadratic Discriminant Analysis, and SVM). Di (2014) uses various technical indicators such as RSI, on balance volume, and Williams %R, among others, as features in the extremely randomized tree algorithm implemented [11]. The most relevant features were then selected, which were further used as inputs to a Kernelized SVM model. In their work, Devi et al., (2015) propose a hybrid model that utilizes cuckoo search to optimize the parameters of SVM. The given model explored several technical indicators such as RSI, Money Flow Index, EMA, Stochastic Oscillator, and MACD [12]. Liaw & Wiener (2012) proposed the use of a neural network ensemble, to predict the direction of the stock price [13]. Khaiden et al. (2016) also used ensemble learning algorithms and specifically Random Forest to predict the direction of stock prices with technical indicators such as the Relative Strength Index (RSI) and stochastic oscillators used as inputs to train the model [14]. The results indicate that the Random Forest algorithm outperforms existing algorithms explored in their study. [15]   Manish & Thenmozhi (2011) apply five machine learning techniques i.e., (i) Support Vector Machine, (ii) Random Forest, (iii) K-Nearest Neighbor (KNN), (iv) Naive Bayes, and finally, (v) SoftMax, to predict stock market trends, with experimental results showing that the Random Forest algorithm performs best for large datasets while Naïve Bayesian Classifier performs best for small datasets [15]. Misra et al., (2018) analyze several machine learning classification models, concluding that the effectiveness and accuracy of the algorithm mainly depend on the type and volume of data on which predictions are analyzed [16].

Singh et al. (2019) explored distinct machine learning methods (including support vector machines, random forests, and boosted decision trees, among others) in order to build prediction models and forecast the prices of stocks for various exchange markets [17]. Pabuccu et al. (2020) follow a combination of technical and fundamental analysis and further apply machine learning models, i.e., Random Forest, Support Vector Machines, and a feed-forward Neural Network to predict the market using time series prediction and sentiment analysis [18]. Shen & Shafiq (2020) focus on the Chinese market and propose comprehensive customization of feature engineering and deep learning-based models in order to predict the price trend of stocks in the examined market [19]. The proposed system achieves high accuracy in terms of stock market trend prediction.

Subasi et al. (2021) compare different machine learning classification algorithms against the National Association of Securities Dealers Automated Quotations System (NASDAQ), New York Stock Exchange (NYSE), Nikkei, and Financial Times Stock Exchange (FTSE), concluding that Random Forest and Bagging with leaked datasets provide satisfactory performance [20]. Additionally, Srinu Vasarao & Chakkaravarthy (2022) use a random forest model to predict stock price movements [21]. Singh (2022) focuses on predicting the Nifty 50 Index by examining eight Supervised Machine Learning Models (Random Forest, Linear Regression, SVM, and KNN, among others) [22]. The evaluation results showed that SVM performed better than the other machine algorithms explored, but with an increase in the dataset size, Stochastic Gradient Descent gave better results. Aliyeva (2022) employs Random Forest and Logistic regression on a five-year-long dataset of volume and price of the Tesla company stocks traded on the New York Stock Exchange (NYSE) to estimate the closing prices of these stocks [23].

Most of the methods above can be used for stock market forecasting prediction but they may also be employed broadly in other economic and financial series and, more recently, in the cryptocurrency market. Examples include, among many others, Demirer et al. (2021) [24]. The Random Forest machine learning technique has been further employed by Bouri et al. (2021) and Gkillas et al. (2021) to predict bitcoin volatility [25, 26]. In their work, Görgen et al. (2022) employ Generalized Random Forests to estimate and predict the risk measure "value at risk" for cryptocurrencies [27].

In line with the above-mentioned research in this field, we propose the use of random forests for stock price forecasting purposes. Random Forest is an ensemble technique, basically operating by building several decision trees at training time [28]. Predictions using random forests are generated by averaging the predictions of the decision trees, which allows us to reduce the variance and improve the efficiency of test set decision trees, and avoid overfitting. The proposed methodology (Figure 2) is presented in Figure 2.
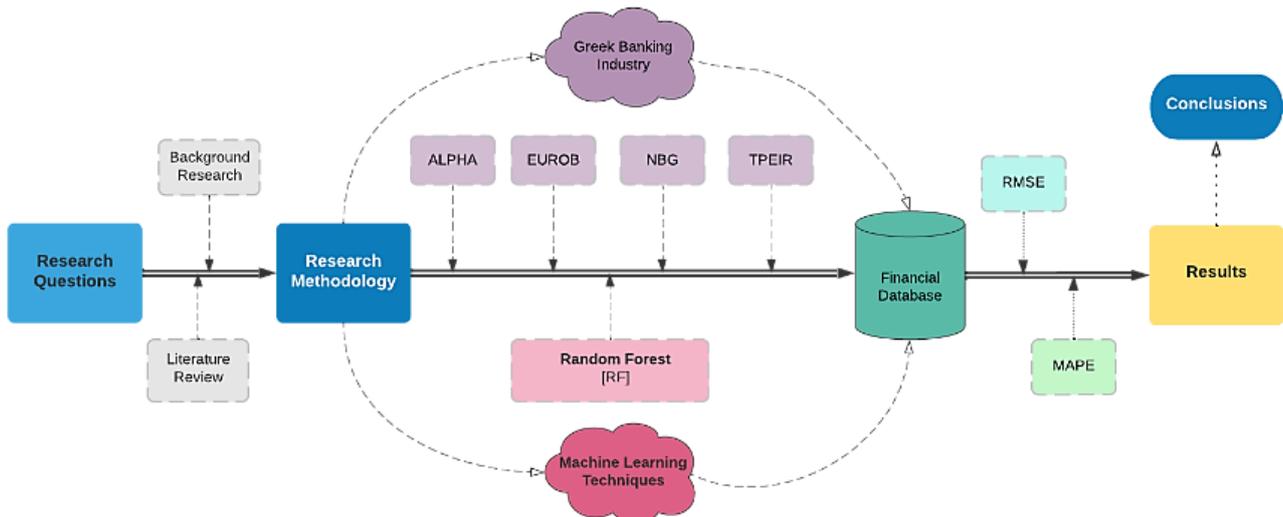
Figure 2. Flowchart of Proposed Methodology

3- Methodology

3-1- Description of Data

The data used here for the four Greek systemic banks has been collected since 2001. The dataset includes 20 years of data from 1/1/2001 to 30/12/2020 from Alpha Bank (ALPHA), the National Bank of Greece (NBG), Eurobank (EUROB), and Piraeus Bank (TREIR), incorporating various market phases such as booms and crashes. The data used here includes intraday data information (i.e., (i) open stock price, (ii) close stock price, and (iii) volume). Only the day-wise closing systemic bank stock price has been considered. The rolling window approach has been followed for model validation as a more adequate approach for time series forecasting.

For the purposes of our analysis, we used the Python programming language. Specifically, we relied on Python's Pandas and Sklearn libraries for data preparation, machine learning model selection, and training of the model to obtain the model predictions for each stock examined. We considered as explanatory variables the variables presented in the following section, with the further inclusion of Google Trends data that served as inputs to the random forest algorithm to predict daily stock price returns for each bank.

3-2- New Variables

Additional variables are used to forecast each of the systemic bank's closing prices returns. These variables have been employed for model training. The new variables considered are the following:

1. A new variable based on the bank closing price using a five-day moving average (noted by MA-5);

2. A new variable based on the bank closing price using a ten-day moving average (noted by MA-10);

3. A new variable based on the bank closing price using a twenty-one-day moving average (noted by MA-21);

4. A new variable based on the bank closing price using standard deviation for the past five days (noted by STD DEV-5).

3-3- Machine Learning Methods

This sub-section aims to briefly discuss and explain the machine learning techniques deployed in preceding research for stock prediction and forecasting [29]. The following section summarizes the various machine learning approaches presented: (i) Artificial Neural Networks, (ii) Support Vector Machine, (iii) Naïve Bayes, and (iv) Deep Neural Network.

i. Artificial Neural Networks (ANN): ANN stands for a network that is based on a simple computer node (neurons) interconnected. It is inspired by the Central Nervous System (CNS), which it attempts to simulate. Their architecture is based on the architecture of Biological Neural Networks and so they are able to perform massive calculations. The purpose of ANNs is to be able to perform the calculations performed by the human brain, that is, to be able to transmit information about the stimuli they receive. ANNs are trained to be able to solve the tasks assigned to them or to be able to perform certain processes on their own, e.g., recognize images. First, however, it is essential that they are properly trained.

ii. Support Vector Machine (SVM): SVMs stand for a set of supervised learning models with algorithms used for classification and regression analysis. The fundamental principles of SVM, developed by Vapnik, are based on the theory of statistical learning. They have been used in a wide range of real-world applications, such as text categorization, image recognition, audio, data sorting, and data detection.

iii. Naïve Bayes (NB): The Naive Bayes algorithm is a family of probabilistic machine learning models used for classification problems and their central idea is based on Bayes's theorem.

iv. Deep Neural Network (DNN): DNN is a new approach to data learning. Uses a specific family of models: sequences of simple functions interconnected. These function chains are the neural networks. These sequences of functions can analyze a complex idea in a hierarchy of simpler ones. Each layer organizes the previous layer into more advanced and abstract concepts. Below, we present the Random Forest procedure, which was used for the empirical part of the paper, as it gives the best results across the previous methods presented.

3-4- *Random Forest*

Numerous machine learning and deep learning forecasting algorithms have been developed in recent years to address the increasing diversity and complexity of forecasting challenges. The selection of a machine learning algorithm is contingent on numerous variables, including the business question you are attempting to answer, the availability and relevance of historical data, the accuracy and success metrics you must achieve, the horizon, and the amount of time your team must develop a forecasting solution [30]. These limits must be regularly and on multiple levels weighed. In the case of demand forecasting, load data must be continuously and regularly projected, and a robust and dependable data intake procedure must be in place to assure the flow of raw data [31].

Random Forest (RF) is an ensemble machine learning technique widely used for both classification and regression problems. Random forests are comprised of many individual arbitrary generated decision trees [28]. In regression, the Random Forest algorithm is basically described by the following: Consider a regression tree, which is consisted of branches and internal and terminal nodes. By constructing the regression tree, at each step, we select features in a random way that come from the set of predictor variables, and the best split- based on a pre-chosen cost function- is used to split the node. This random selection of predictor variables helps mitigate the influence of influential predictors on each Tree building. The building of the decision tree continues until the set of terminal nodes for the tree stopping criterion, or the number of observations on every terminal node, is met. The final prediction using random forests is produced by averaging the predictions of the individual ensemble trees. This allows us to reduce the variance and improve the efficiency of test set decision trees, thereby avoiding overfitting and delivering higher accuracy predictions [28] In this study, newly created technical analysis tools are provided as variables for training.

4- Results and Discussion

In order to assess the forecasting performance of the models used on four different banks, namely Alpha Bank, Eurobank, the National Bank of Greece, and Piraeus Bank with the use of the Random Forest machine learning technique, we employ two widely used evaluation criteria: (i) the Root Mean Square Error (RMSE), (ii) and the Mean Absolute Percentage Error (MAPE).

The RMSE metric which has been utilized to evaluate the performance of the model is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_i^n(y_i - \hat{y}_i)} \tag{1}$$

where 'y_i' is the actual closing price series for each bank considered, '$\hat{y}_i$' is the estimation closing price value, while the symbol 'n' stands for the total size of the window considered.

The criterion MAPE is also employed for model evaluation performance. This criterion is given by the following:

$$MAPE = \frac{100}{n}\sum_i^n\frac{y_i - \hat{y}_i}{y_i} \tag{2}$$

where 'y_i' again is the actual closing price series for each bank considered, '$\hat{y}_i$' is the estimation closing price value, while the symbol 'n' is the total window size considered.

In our analysis, the prediction of the returns of each of the four stocks has been made considering the rolling window approach for time series data. Specifically, the expanding window where the window in each forecasting step is expanded with the addition of a new daily observation in the training set [32]. To assess the actual performance of the proposed approach, we consider one step ahead (the following day) out-of-sample prediction. Each of the following figures depicts the actual price of each of the four systemic Greek banks considered in regard to the forecast value obtain for the RF model employed. Further results for the RMSE and MAPE evaluation criteria are given in the following table (Table 1). In particular, the figure above (Figure 3) and below figures (Figures 4 to 6) shows a visual representation of the performance of the proposed machine learning models for each of the four banks considered. We observe in all four cases that the random forest-based forecasting models predicted values follow for the most part similar patterns as the observed values.

Table 1. RMSE and MAPE values obtained using RF models

Banks	RMSE	MAPE
Alpha Bank	12.9450	4.7247
Eurobank	414.4279	5.1782
National Bank of Greece	1201.3288	10.6036
Piraeus Bank	185.7749	5.4684

Figure 3. Alpha Bank closing price predicted vs. actual values

Figure 4. Eurobank closing price predicted vs. actual values

Figure 5. National Bank of Greece closing price predicted vs. actual values

Figure 6. Piraeus Bank closing price predicted vs. actual values

Additionally, we observe that in most cases, the predictions are very close to the actual values. Moreover, as it is evident from the plots in each case, the model has captured well upward or downward moves in the corresponding bank return series. Furthermore, according to the results presented in Table 1, when considering the MAPE values obtained among the four banks examined, the proposed models based on a one-step-ahead out-of-sample forecasting approach performed better when predicting Alpha Bank returns.

Moreover, we notice low-magnitude differences between the MAPE values of three of the four cases (Alpha Bank, Eurobank, and Piraeus Bank), meaning that the proposed models have similar predictive performance for the specific stocks. Specifically, the MAPE values obtained from our models in these cases are very close to the performance of alternative machine learning techniques for predicting stock prices, as observed in the detailed summary of various algorithms and their performance presented in the work of Soni et al. (2022) [33]. For the case of the National Bank of Greece, the corresponding MAPE value is 10.60%. Overall, we conclude that our proposed forecasting approach is able to accurately predict Greek bank sector returns. Such a result is in line with the related literature on stock market price prediction that highlights the superiority of machine learning methods to accurately predict stock market returns (for a more detailed overview of the subject of stock market prediction with machine learning algorithms, see Rouf et al. (2021) [34].

5- Conclusion

Accurate forecasting of bank prices is a challenging task that captures the interest of academics and investors. The benefits of using machine learning algorithms for forecasting stock market returns relative to traditional econometric strategies have been highlighted in the related literature. Indeed, several machine learning algorithms better suited for time series forecasting have been proposed, considering their effectiveness in identifying complex relations in stock market data. Furthermore, the predictive capabilities of internet search data on stock returns have also been examined as additional predictors of interest in the forecasting models.

In this work, the Random Forest machine learning algorithm has been employed to predict stock prices in a one-step-ahead, out-of-sample forecasting exercise. Additional technical analysis tools have been used as inputs to the machine learning model for each of the four cases of Greek banks considered in the analysis, as in Theodorakopoulos et al. (2022) [35], along with Google Trends data as an effective proxy for investor attention regarding the Greek banking sector. Random Forest has been effectively used in predicting stock market prices. We show that in all four cases, the random forest-based forecasting models predicted values follow, for the most part, similar patterns as the observed values. We also observe that in most cases, the predictions are very close to the actual values. Overall, we conclude that our proposed forecasting approach is able to accurately predict Greek bank sector returns. However, further research should be conducted considering more sophisticated machine learning and deep learning models. Future research could consider alternative sources indicative of investment attention and/or sentiment related to financial news, for instance, users' sentiment from Twitter posts, as variables to improve the effectiveness of predicting Greek banks' stock prices.

6- Declarations

6-1-Author Contributions

Conceptualization, H.A., L.T., C.H., and V.M.; methodology, H.A., L. T., C.H., and V.M.; investigation, H.A., L.T., C.H., and V.M.; data curation, H.A., L.T., C.H., and V.M.; writing—original draft preparation, H.A., L.T., C.H., and V.M.; writing—review and editing, H.A., L.T., C.H., and V.M. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6-3- Funding

The publication of this article has been financed by the Research Committee of the University of Patras, Greece.

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Participants gave their written consent to use their anonymous data for statistical purposes. All of them were over 18 years old and voluntarily collaborated without receiving any financial compensation.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Masoud, N. M. H. (2013). The impact of stock market performance upon economic growth. International Journal of Economics and Financial Issues, 3(4), 788–798.

[2] Tan, Z., Yan, Z., & Zhu, G. (2019). Stock selection with random forest: An exploitation of excess return in the Chinese stock market. Heliyon, 5(8), 2310. doi:10.1016/j.heliyon.2019.e02310.

[3] Jiang, W. (2021). Applications of deep learning in stock market prediction: recent progress. Expert Systems with Applications, 184, 115537. doi:10.1016/j.eswa.2021.115537.

[4] Murkute, A., & Sarode, T. (2015). Forecasting Market Price of Stock using Artificial Neural Network. International Journal of Computer Applications, 124(12), 11–15. doi:10.5120/ijca2015905681.

[5] Hur, J., Raj, M., & Riyanto, Y. E. (2006). Finance and trade: A cross-country empirical analysis on the impact of financial development and asset tangibility on international trade. World Development, 34(10), 1728–1741. doi:10.1016/j.worlddev.2006.02.003.

[6] Dai, Y., & Zhang, Y. (2013). Machine learning in stock price trend forecasting. Stanford University, California, United States.

[7] Li, L., Wu, Y., Ou, Y., Li, Q., Zhou, Y., & Chen, D. (2017). Research on machine learning algorithms and feature extraction for time series. 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). doi:10.1109/pimrc.2017.8292668.

[8] Seber, G. A., & Lee, A. J. (2003). Linear regression analysis. John Wiley & Sons, New York, United States. doi:10.1002/9780471722199.

[9] Atsalakis, G. S., Atsalaki, I. G., Pasiouras, F., & Zopounidis, C. (2019). Bitcoin price forecasting with neuro-fuzzy techniques. European Journal of Operational Research, 276(2), 770–780. doi:10.1016/j.ejor.2019.01.040.

[10] Zhang, P. G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175. doi:10.1016/S0925-2312(01)00702-0.

[11] Di, X. (2014). Stock trend prediction with technical indicators using SVM. Independent Work Report, Leland Stanford Junior University, Stanford, United States.

[12] Devi, K. N., Bhaskaran, V. M., & Kumar, G. P. (2015). Cuckoo optimized SVM for stock market prediction. 2015 International Conference on Innovations in Information, Embedded and Communication Systems. doi:10.1109/iciiecs.2015.7192906.

[13] Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. R News, 2(3), 18-22.

[14] Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. arXiv Preprint arXiv:1605.00003. doi:10.48550/arXiv.1605.00003.

[15] Manish, K., & Thenmozhi, M. (2011). Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest. SSRN Electronic Journal. doi:10.2139/ssrn.876544.

[16] Misra, M., Yadav, A. P., & Kaur, H. (2018). Stock Market Prediction using Machine Learning Algorithms: A Classification Study. 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE). doi:10.1109/icrieece44171.2018.9009178.

[17] Singh, S., Madan, T. K., Kumar, J., & Singh, A. K. (2019). Stock Market Forecasting using Machine Learning: Today and Tomorrow. 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). doi:10.1109/icicict46008.2019.8993160.

[18] Pabuçcu, H., Ongan, S., & Ongan, A. (2020). Forecasting the movements of Bitcoin prices: an application of machine learning algorithms. Quantitative Finance and Economics, 4(4), 679–692. doi:10.3934/qfe.2020031.

[19] Shen, J., & Shafiq, M. O. (2020). Short-term stock market price trend prediction using a comprehensive deep learning system. Journal of Big Data, 7(1), 66. doi:10.1186/s40537-020-00333-6.

[20] Subasi, A., Amir, F., Bagedo, K., Shams, A., & Sarirete, A. (2021). Stock Market Prediction Using Machine Learning. Procedia Computer Science, 194, 173–179. doi:10.1016/j.procs.2021.10.071.

[21] Srinu Vasarao, P., Chakkaravarthy, M. (2022). Time Series Analysis Using Random Forest for Predicting Stock Variances Efficiency. Intelligent Systems and Sustainable Computing. Smart Innovation, Systems and Technologies, 289, Springer, Singapore. doi:10.1007/978-981-19-0011-2_6.

[22] Singh, D. G. (2022). Machine Learning Models in Stock Market Prediction. International Journal of Innovative Technology and Exploring Engineering, 11(3), 18–28. doi:10.35940/ijitee.c9733.0111322.

[23] Aliyeva, A. (2022). Predicting Stock Prices Using Random Forest and Logistic Regression Algorithms. 11th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions and Artificial Intelligence, Lecture Notes in Networks and Systems, 362, Springer, Cham, Switzerland. doi:10.1007/978-3-030-92127-9_16.

[24] Demirer, R., Gkillas, K., Gupta, R., & Pierdzioch, C. (2022). Risk aversion and the predictability of crude oil market volatility: A forecasting experiment with random forests. Journal of the Operational Research Society, 73(8), 1755–1767. doi:10.1080/01605682.2021.1936668.

[25] Bouri, E., Gkillas, K., Gupta, R., & Pierdzioch, C. (2021). Forecasting Realized Volatility of Bitcoin: The Role of the Trade War. Computational Economics, 57(1), 29–53. doi:10.1007/s10614-020-10022-4.

[26] Gkillas, K., Tantoula, M., & Tzagarakis, M. (2021). Transaction activity and bitcoin realized volatility. Operations Research Letters, 49(5), 715–719. doi:10.1016/j.orl.2021.06.016.

[27] Görgen, K., Meirer, J., & Schienle, M. (2022). Predicting Value at Risk for Cryptocurrencies Using Generalized Random Forests. SSRN Electronic Journal. doi:10.2139/ssrn.4053537.

[28] Suykens, J. A. K., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural Processing Letters, 9(3), 293–300. doi:10.1023/A:1018628609742.

[29] Theodorakopoulos, L. (2022). Big data analysis in humanities and economics with machine learning techniques and use of cloud computing technologies. PhD Thesis, Department of Business Administration, University of Patras, Patras, Greece.

[30] Antonopoulou, H., Mamalougou, V., & Theodorakopoulos, L. (2022). The Role of Economic Policy Uncertainty in Predicting Stock Return Volatility in the Banking Industry: A Big Data Analysis. Emerging Science Journal, 6(3), 569–577. doi:10.28991/ESJ-2022-06-03-011.

[31] Lazzeri, F. (2019). "Data Science Mindset: Six Principles to Build Healthy Data-Driven Organizations. InfoQ. Available online: https://www.infoq.com/articles/data-science-organization-framework (accessed on January 2023).

[32] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts, Melbourne, Australia.

[33] Soni, P., Tewari, Y., & Krishnan, D. (2022). Machine Learning Approaches in Stock Price Prediction: A Systematic Review. Journal of Physics: Conference Series, 2161(1). doi:10.1088/1742-6596/2161/1/012065.

[34] Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. Electronics (Switzerland), 10(21). doi:10.3390/electronics10212717.

[35] Theodorakopoulos, L., Antonopoulou, H., Mamalougou, V., & C. Giotopoulos, K. (2022). The drivers of volume volatility: A big data analysis based on economic uncertainty measures for the Greek banking system. Banks and Bank Systems, 17(3), 49–57. doi:10.21511/bbs.17(3).2022.05.