Stock market price prediction with Cascading and Ensemble classifier methods

Ritesh Kumar Yadav, Research Scholar, SRK University, Bhopal
Dr. M. Sivakkumar, Professor, SRK University, Bhopal
Dr.P.Gomathi, Professor, N.S.N. College of Engineering and Technology, Karur, Tamilnadu

Abstract :
The stock market is the backbone of the financial status of any country. The unpredictable behaviors of the stock market change the mental group of investors and buyers. If the stocks traders predict right trend in stock price, they can realize profits. Therefore, prediction of stock price is very important factor for buyers and seller in stock market. The accuracy and behavior is accurate in limited data of stocks, if the range of data are increases the impact of prediction is decrease. The recent research in the area of stock price prediction with machine learning methodologies makes use mainly of SVM architectures and takes into account several predictors. The aim of this proposed work is to increase the capacity of sample size of classifier and increase the accuracy of prediction with the derivation of cascading and ensemble classifier.
Keywords: Prediction, Support Vector machine, Ensemble, Stock market.

1. INTRODUCTION:

Stock market prices and its trends form the most volatile sector of finance. It tends to be the center point of some important researches that aim at capturing its unpredictability and anticipating its next moves. The historical data produced by stock markets is immense. Since the amount of data to be analyzed is so huge, it is very difficult for any individual to consider all the data values while predicting future stock trends. The fundamental and technical analyses are two main methodologies prevalent in finance that are used for stock market predictions. Fundamental analysis takes in account reasons and factors that affect companies or economies usual business and its future prospects. On other hand, technical analysis attempts to study variations in prices of a stock and uses this historical data to predict its future prices [1]. Thus, it considers past trends in the stock price to forecast stock price changes in future. There is a prominence of virtual stock market games in recent times, where news and events are created using online platforms and scored manually [3]. These events in turn have a big impact on the value of future stock prices. The objective of this paper is to find the relationship between the historical stock data of any company and news/events related to the same company, and to build a model which can forecast the stock prices accurately considering the impact of news and events on stock price. Finally, a regression model is built to understand the impact of news/events on changes in stock market prices.

2. LITERATURE SURVEY

Ican, Ozgur Et al(2017) previous studies featuring an artificial neural networks-based prediction model have been reviewed. The main purpose of this review is to examine studies which use directional prediction accuracy or profitability of the model as a benchmark. For the argument that they are not able to actually show how useful the prediction model is, in terms of financial gains (i.e. for practical usage).

Bao, Wei Et al. (2017) authors apply long-short term memory networks to a large-scale financial market prediction task on the S&P 500, from December 1992 until October 2015. With their work, authors make three key contributions to the literature: The first contribution focuses on the large-scale empirical application of LSTM networks to financial time series prediction tasks. Authors provide an in-depth guide, closely following the entire data science value chain. Specifically, authors frame a proper prediction task, derive sensible features in the form of 240- day return sequences, standardize the features during pre-processing to facilitate model training, discuss a suitable LSTM architecture and training algorithm, and derive a trading strategy based on the predictions.

Vargas, Manuel R. Et al. (2017) this work uses deep learning methods for intraday directional movements prediction of Standard & Poor's 500 index using financial news titles and a set of technical indicators as input. Deep learning methods can detect and analyze complex patterns and interactions in the data automatically allowing speed up the trading process. This work focuses on architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which have had good results in traditional NLP tasks. Results has shown that CNN can be better than RNN on catching semantic from texts and RNN is better on catching the context information and modeling complex temporal characteristics for stock market forecasting.

Pradeepkumar, Dadabada Et al.(2017) Accurate forecasting of volatility from financial time series is paramount in financial decision making. This work presents a novel, Particle Swarm Optimization (PSO)-trained Quantile Regression Neural Network namely PSO-QRNN, to forecast volatility from financial time series. authors compared the effectiveness of PSOQRNN with that of the traditional volatility forecasting models. Pyo, Sujin Et al.(2017) Method of Data Handling (GMDH), Random Forest (RF) and two Quantile Regression (QR)-based hybrids including Quantile Regression Neural Net-work (QRNN) and Quantile Regression Random Forest (QRRF).

Di Persio, Luca Et al(2017) A huge quantity of learning tasks have to deal with sequential data, where either input or out-put data can have sequential nature. This is the case, e.g., of time series forecasting, speech recognition, video analysis, music generation, etc., since they all require algorithms able to model sequences.

Feuerriegel, Stefan Et al. (2018) Share valuations are known to adjust to new information entering the market, such as regulatory disclosures. Authors study whether the language of such news items can improve short-term and especially long-term forecasts of stock indices. For this purpose, this work utilizes predictive models suited to high-dimensional data and specifically compares techniques for data-driven and knowledge-driven dimensionality reduction in order to avoid overfitting. Their research provides implications to business applications of decision-support in financial markets, especially given the growing prevalence of index ETFs.

Cao, Jian Et al. (2019) In order to improve the accuracy of the stock market prices forecasting, two hybrid forecasting models are proposed in this work which combines the two kinds of empirical mode decomposition (EMD) with the long short-term memory (LSTM).

3. PROPOSED WORK :

3.1 Cascaded classifier

The proposed method of stock price prediction describes the process of cascaded classifier and ensemble methods for the final prediction of stock price.

The variation of attribute degraded the prediction of stock price. For the optimization of attribute variance applied particle swam optimization. The particle swarm optimization is iterative based optimization algorithms. In particle of swarm optimization two main derivates of data processing one is acceleration factor and other is constant factor c1 and c2. The process of particle swarm optimization [11] mapped with NSE stock data and the derivation describe as

$$Pk = \frac{FSk - FSb}{FSmax - FSmin} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Define the process of data movement (Velocity)

$$V_{k,} = \beta \times V_{k,} + c \times r1 \times (x_{best,} - x_{k,}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

The set of variance data is

D=1, 2, M ....(3)

Update the position of new data

$$x_{\text{k},} = x_{\text{k},} + V_{\text{k}.d} \text{ .... (4)}$$

Where FS are fitness constraints and P is value of data iteration of probability of each data point.

B is the weight; c is acceleration constant r1 is random value between 0 to 1.Process of algorithms

1. Initialization of particle of data mapped in terms of swarm
2. Define the random position of particle
3. Define random velocity of particle
4. Iteration process stated and move to new data set
5. Terminate the process of iteration
6. Final set of optimal attribute Finals(d1,d2, ...,dn)
7. Repeat step 2 to 4

Cascading of machine learning is the process of design of accurate algorithms for the prediction of stock price. The process of cascading used three level of support vector machine. The three level of support vector machine define as C0, C1, Cn. The processing of data of NSEis normalized and optimized with particle of swarm optimization. The processing of cascading of support vector machine describe here.

A dataset describes by a raw attribute vector of dimension (f*1) we define t level with t≥ 3 in each level one or more class predict the dataset and process the result of prediction in nextlevel of class.

The chain of cascaded classifier defines as

$$cacdaded\ class\text{ifier} = (C0)\nabla(C1)\nabla(C2) \quad t \geq 3 \text{......................(5)}$$

Algorithm for cascaded machine learning

Begin

Set of vector SV1= (C0, SVM) Set of Vector SV2= (C1, SVM)Set of vector SV3= (C2, SVM)

M= level of model for derived class
C! = {SV1, SV2, SV3}
End
Wher
e

1. C0 is class of data space for the processing of vector
2. C1 is training level of class for the process of learning factor
3. C2 is final class of prediction of stock

price**Process of cascade of ML**
Proceed data (D, C2,
L)Begin
M =mapped model of L from
classNew D=Φ
For each prediction =$(x, y) \in D'$(D' is trained sample of data)
For j= 1 to final C
Prediction ={x, C2 class of M: map of
final_s}newD =newD V (predication)
return
predictionend

3.2 Ensemble method

Ensemble method is predominated process of machine learning and mathematical derivation. Ensemble algorithms drive to merge multiple classifiers to single classifier to achieve maximum accuracy and minimum error. For the designing of ensemble classifier various methods such as maximum voting, average voting, bagging, boosting and many more. In these ensemble methods of bagging we used two classifier one is decision tree algorithm and other is support vector machine [2, 4] . The support vector machine is base classifier and decision tree is variable classier. It describe of ensemble classifier describe here.

Algorithm 1. Ensemble of SVM and
DTInput: input dataXi = {, i = 1, .. ,
n}
Output: O – predication of price
Begin
Step 1. Define the classifier ratio M*N

Step 2. Measure entropy of $price(DT;\ x$i)for each input, i = 1,........., n

Step 3. $Em = SVM$; Features assigns$stock\ price$ i

Then, set$em \leftarrow \{x\mathrm{i}\}; SVM \leftarrow MN \cup \{DT\mathrm{i}\};\ n\ EM_s = EM_s - 1$

Step 4. While$EM \neq \emptyset$ do

Measure value of voting to $x_i$wherei $\in \{1,2,...., n\}$;

em = em − 1;

if (SVM > 0) then
em ← em ∪
{mMNi}end
end

Step 5 Find the prediction ratio of stock price
return $price$

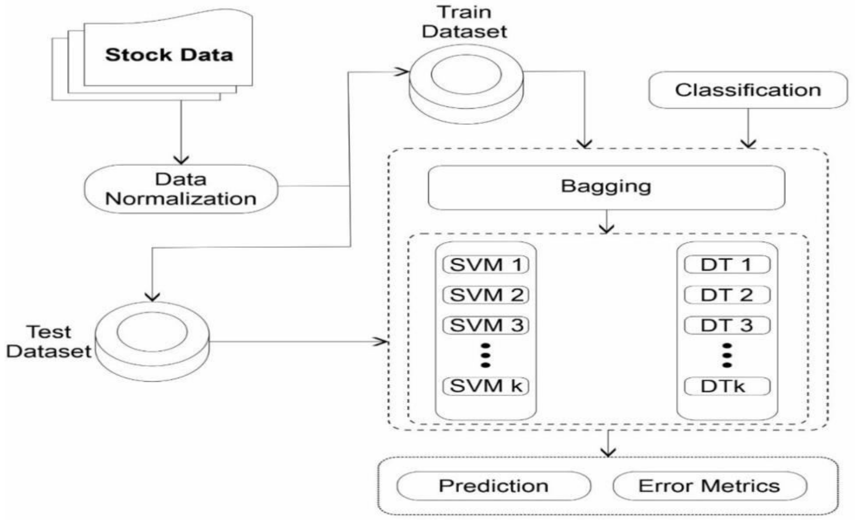

Figure 1: process block of ensemble-based classifier using support vector machine and decision tree algorithm

4. DATASET

A stock's performance can, to some extent, be analyzed based on financial indicators presented in the company's annual report. The annual report contains a vast amount of information that can be transformed into various ratios. Previous literature suggests that financial ratios are important tools for assessing future stock performance. Analysts, investors, and researchers use financial ratios to project future stock price trends. For the experimental analysis used NSE dataset of 6000 thousand of data of different selection point. Source of dataset given below: https://www.quandl.com/

4.1. IMPLEMENTATION PROCESS

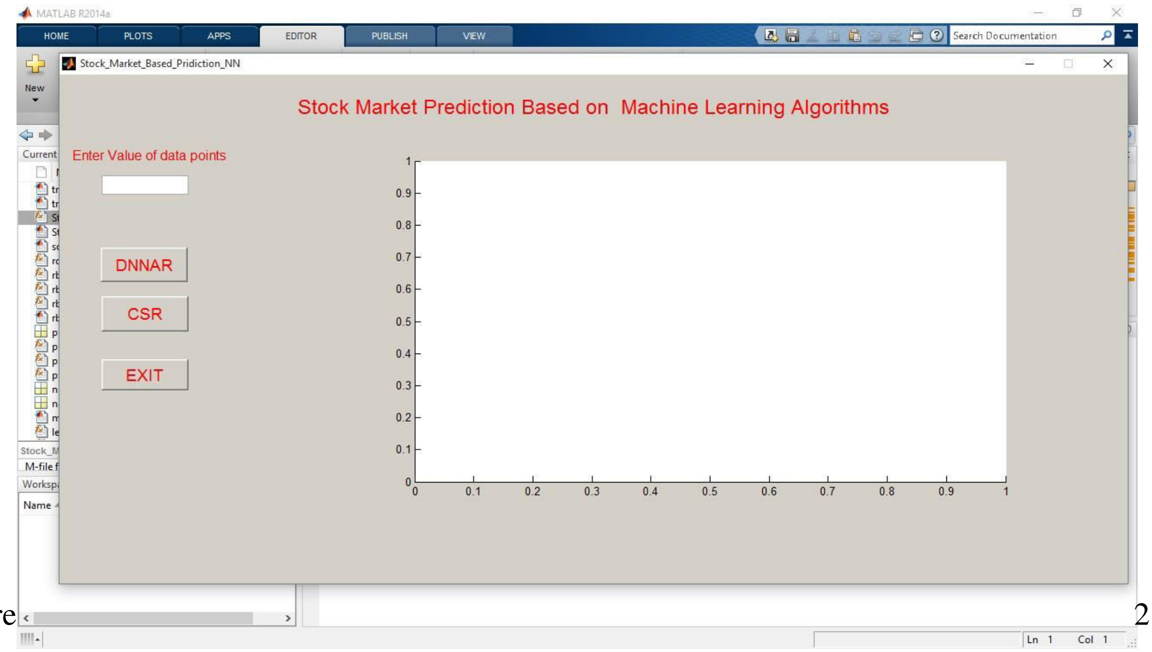

Figure 2: Stock market prediction based on machine leaning algorithms with input field ofdata points and three buttons DNNAR, CSR and EXIT.

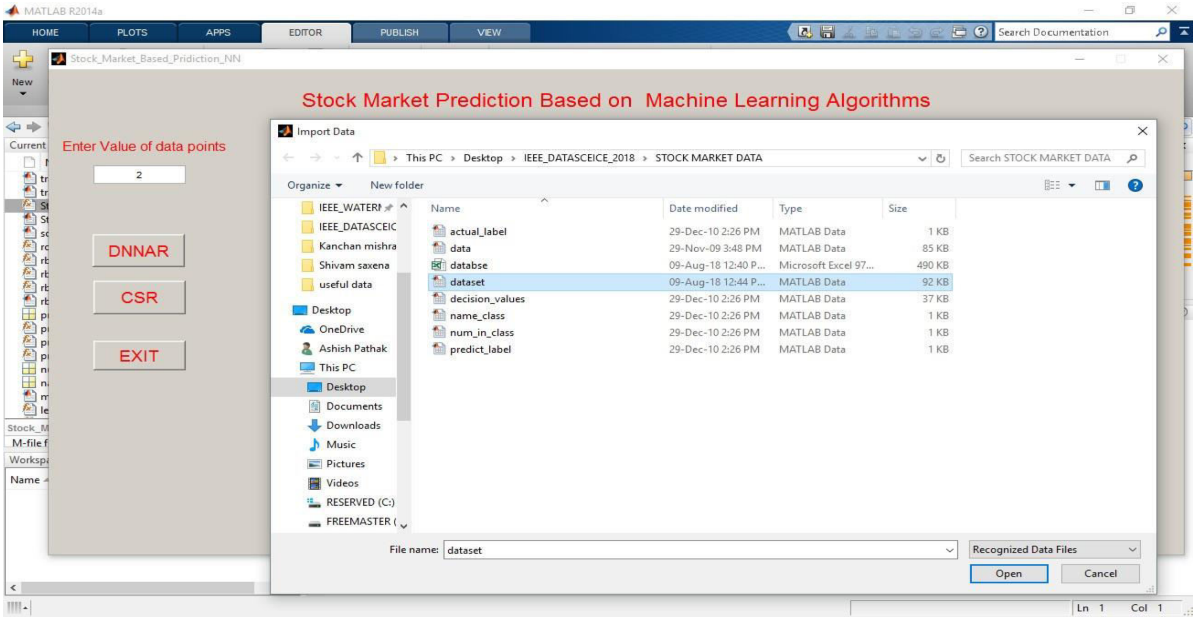

Figure 3: After the DNNAR method button browse the dataset file in stock market prediction based on machine learning algorithms with input field of data points value is 2.

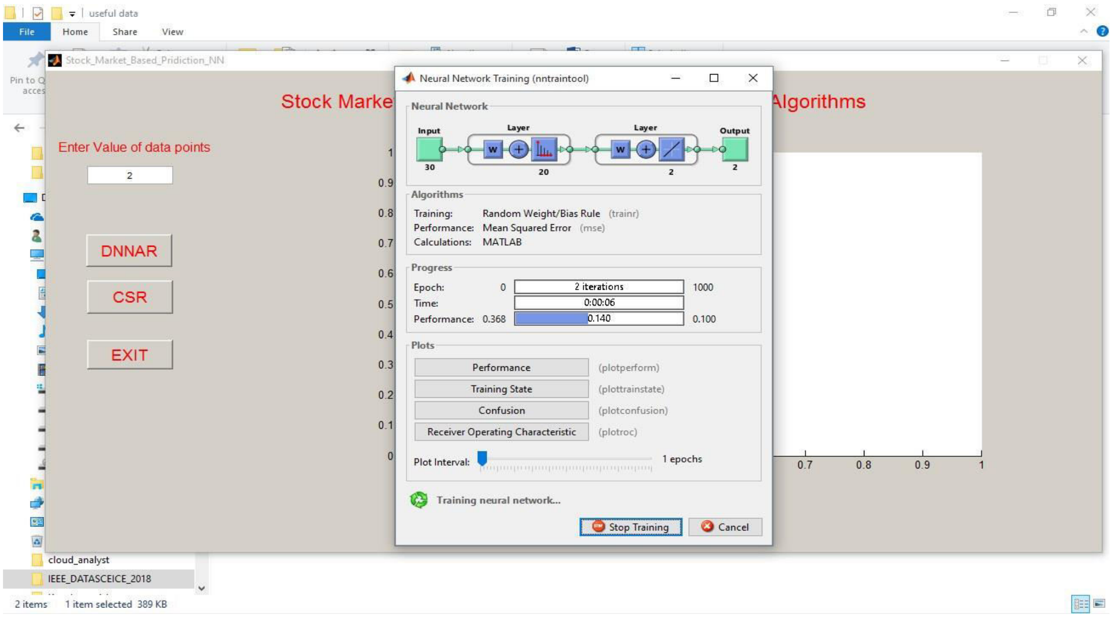

Figure 4: Neural network training wizard with complete process performance in stock market prediction based on machine leaning algorithms using input field of data point's value is 2.

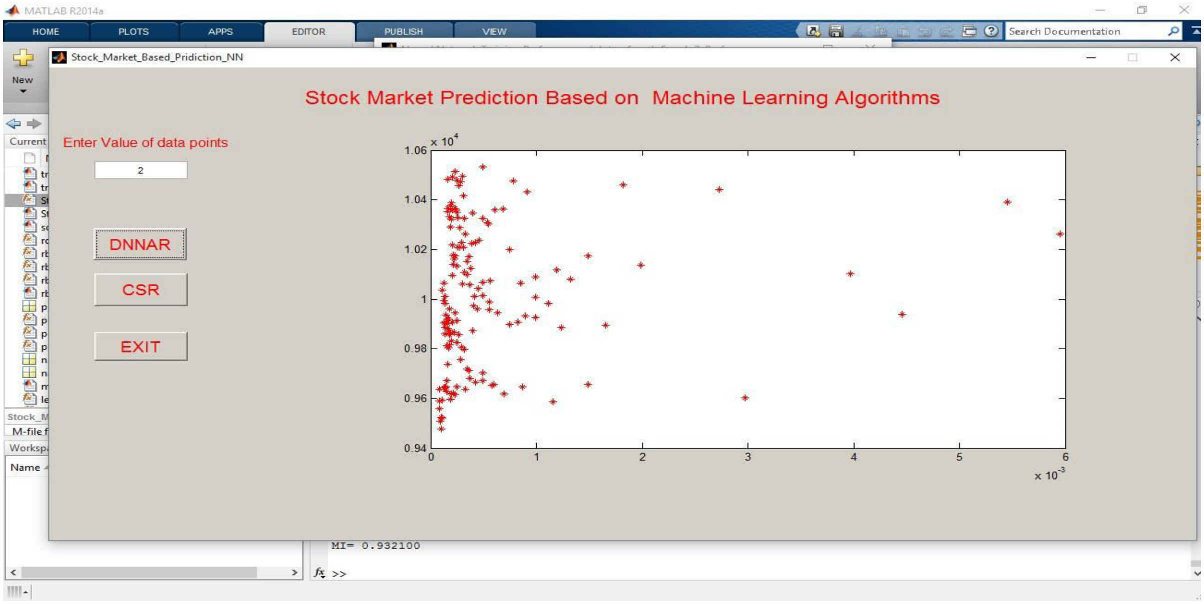

Figure 5: Output view with complete process performance in stock market prediction basedon machine learning algorithms using input field of data point's value is 2.

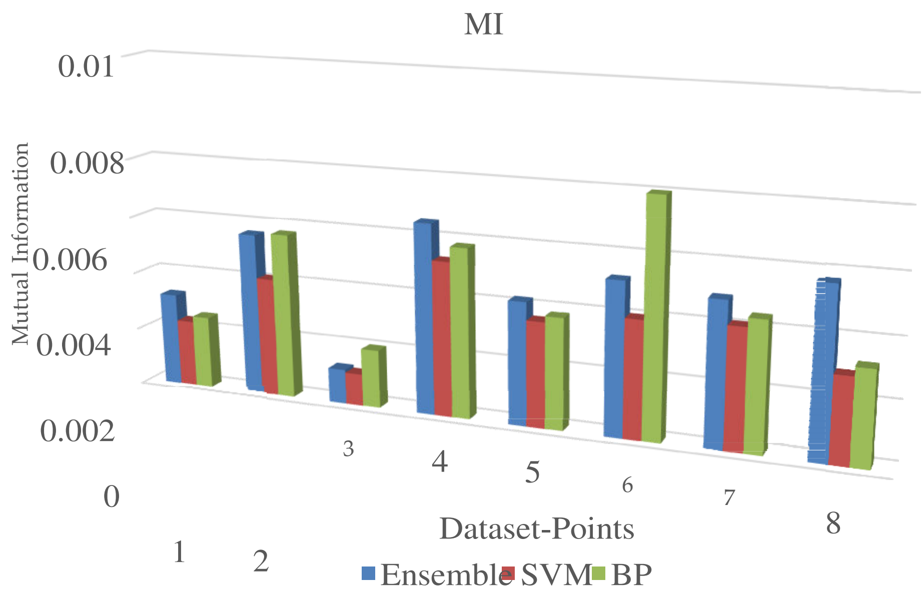

Figure 6: Comparative performance analysis of MI (Mutual Information) using Ensemble, SVM, BP techniques.

In Fig we observe the performance of ensemble techniques is better compare to other two SVM and BP techniques with all datasets.

CONCLUSION & FUTURE WORK

The trends of stock price play a significant role in decision-making to either sell orbuy stocks. The price trends of the stock market decide the investment future of the stock market. The stock market's future trends decide the behaviors of the opening and closing of the stock market. The non-parametric models of stock price prediction applied linear and non-linear regression and artificial neural network models. The proposed Cascaded classifier based on a machine learning algorithm improves the prediction ratio. For the ensemble processing, bagging methods are applied in the bag of two classifier support vector machines and decision trees. The decision tree algorithm measures the price entropy factor of risk interest of data and mapped these data with strike price for support vector machine. The ensemble of two classifiers performs well results in compression of the support vector machine.

REFERENCES

1. Ican, Ozgur, and BugraCelik Taha. "Stock market prediction performance of neural networks: A literature review." *International Journal of Economics and Finance* 9, Vol no.11, pp.100-108, (2017).
2. Bao, Wei, Yue Jun, and Rao Yulei. "A deep learning framework for financial time series using stacked autoencoders and long-short term memory." *PloS one* 12, Vol no.7, (2017).
3. Feuerriegel, Stefan, and Gordon Julius. "Long-term stock index forecasting based on text mining of regulatory disclosures." *Decision Support Systems,* Vol no.112, pp.88-97, (2018).
4. Cao, Jian, Li Zhi, and Li Jian. "Financial time series forecasting model based on CEEMDAN and LSTM." *Physica A: Statistical Mechanics and its Applications,* Vol no.519, pp. 127-139, (2019).
5. Oliveira, Nuno, Cortez Paulo and Areal Nelson. "The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices." *Expert Systems with Applications*, Vol no. 73, pp. 125-144, (2017).
6. Vargas, Manuel R., Lima Beatriz SLP De, and Evsukoff Alexandre G. "Deep learning for stock market prediction from financial news articles." In *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pp. 60-65, (2017).

7. Pradeepkumar, Dadabada, and Vadlamani Ravi. "Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network." *Applied Soft Computing,* Vol no. 58, pp. 35-52, (2017).
8. Pyo, Sujin, Jaewook Lee, Mincheol Cha, and Huisu Jang. "Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets." *PloS one* 12, Vol no. 11, (2017).
9. Di Persio, Luca, and Honchar Oleksandr. "Recurrent neural networks approach to the financial forecast of Google assets." *International journal of Mathematics and Computers in simulation,* Vol no.11, (2017).
10. Zhong, Xiao, and Enke David. "Predicting the daily return direction of the stock market using hybrid machine learning algorithms." Financial Innovation 5, Vol no. 1, pp.4, (2019).
11. Pang, Xiongwen, Zhou Yanqiang, Wang Pan, Lin Weiwei, and Chang Victor. "An innovative neural network approach for stock market prediction." *The Journal of Supercomputing*, pp.1-21, (2018).