Stock Market Forecasting using Machine Learning: Today and Tomorrow

Sukhman Singh
Department of Computer Applications
National Institute of Technology
Kurukshetra, Haryana, India
sukhman.07singh@gmail.com

Tarun Kumar Madan
Department of Computer Applications
National Institute of Technology
Kurukshetra, Haryana, India
tarunmadan9@gmail.com

Jitendra Kumar
Department of Computer Applications
National Institute of Technology
Kurukshetra, Haryana, India
jitendrakumar@ieee.org

Ashutosh Kumar Singh
Department of Computer Applications
National Institute of Technology
Kurukshetra, Haryana, India
ashutosh@nitkkr.ac.in

***Abstract*—Stock market prediction is a major exertion in the field of finance and establishing businesses. Stock market is totally uncertain as the prices of stocks keep fluctuating on a daily basis because of numerous factors that influence it. One of the traditional ways of predicting stock prices was by using only historical data. But with time it was observed that other factors such as peoples' sentiments and other news events occurring in and around the country affect the stock market, for e.g. national elections, natural calamity etc. Investors in the stock market seek to maximize their profits for which they require tools to analyze the prices and trend of various stocks. Machine learning algorithms have been used to devise new techniques to build prediction models that can forecast the prices of stock and tell about the market trend with good accuracy. Many prediction models have been proposed to incorporate all the major factors affecting the price of stocks. This paper focuses on portraying distinct machine learning algorithms such as support vector machine, deep learning, random forest, boosted decision trees, ensemble methods and a few hybrid methods which have been used to build prediction model and predict the stock prices for different stock exchanges. This paper also covers the various challenges that are encountered while building prediction models.**

Keywords— Stock market prediction, machine learning, evolutionary algorithms, regression, classification, deep learning, ensemble methods, hybrid methods.

I. INTRODUCTION

Over the years, stock market has played a vital role in the prosperity of many businesses and also in the GDP of a country. As stock market is too uncertain, there is no surety that the investments made in the market would bear some profits rather it may incur some losses as well. As a part of the economic liberalization, the stock markets have been given the most important place in the financial schemes of the global corporate sector. Many factors have been found out that affect the stock prices out of which the historical data has been the most prominent one. However, it was observed that solely historical data does not give the predictions accurately. Hence more factors were identified which came out to be affecting the stock prices significantly, these were people's sentiments and news events [6]. Thus, in addition to historical data, financial news and people's reviews become major sources of such information that help in designing good prediction models that can predict the stock market prices with improved accuracy than their predecessors.

However, devising such models is not an easy task. Prediction using solely historical data was also not an easy task as it involved selecting the most essentials features from the large datasets and then using pre-processing techniques on them to filter out the required data as per the specifications of the devised model. Now with the inclusion of other sentiment data, the task becomes even more difficult but the results have improved significantly.

With the evolution of computer science, various new disciplines came into existence which provided better prediction models. One such discipline of computer science is Machine Learning. Over the years, machine learning has played a vital role in predictions. Predictions like workload management in cloud [1] [41-42], heart disease prediction [37], house rent price prediction [38], stock market price prediction [18] etc. were now possible with various techniques of machine learning. It helped in building new and improvised prediction models, which gave better results with lesser complexity. In context with stock market prediction, many researchers have been able to devise models for stock market prediction which uses various techniques of machine learning such as SVM (Support Vector Machine) [7][19][23], Linear Regression [27][31], Random forest [10][13][32], K-Nearest Neighbour (KNN) [12], ANN [39], deep learning [2], LSTM[23], MLP[33][5], Boosted Decision Tree [8], Evolutionary algorithms [28][30][36] and many more hybrid techniques which would be further discussed in this paper. This paper also discusses the challenges that are faced or can be faced by researchers while devising prediction models.

Hybrid techniques use multiple algorithms to use the impacting features of those individual algorithms to predict

more refined outputs. The data used in the papers have been extracted from various sites like Chinese Microblogging site, Chinese Stock Market, US Stock market, Twitter and Indian stock market.

In Section 2 of the paper, various techniques of machine learning have been discussed along with the research done using those techniques in stock market prediction. Section 3, summarizes all the techniques that have been proposed with their corresponding results. Section 4, discusses the various challenges that were faced by researchers while devising prediction models and the Section 5 concludes the paper.

II. STOCK MARKET PREDICTION TECHNIQUES

Many models of prediction have been proposed till date to forecast the stock prices and stock market trends. Some of the machine learning techniques have been discussed in this paper. Table 1 displays the brief summary of all the techniques proposed by various researchers. All the techniques have been classified into various sub categories like classification techniques, regression techniques, ensemble algorithms, evolutionary techniques, deep learning, hybrid models and some other additional techniques.

Below Fig. 1 shows the classification of these technique, which have been explained in the paper for better understanding.

A. Classification Techniques

1) Support Vector Machine(SVM): One of machine learning algorithm that possesses the desired features such as the decision function, usage of kernel method and also the sparsity of the solution is known as the Support Vector Machine (SVM) technique.

Yan et al [7] in their paper portrayed that public mood and stock market price have some relation. They tried to devise a relation between Chinese Stock Market and Chinese local Microblogs. In their research they tested SVM and Probabilistic Neural Network to make predictions. The experiments showed that Support Vector Machine gives better accuracy than Probabilistic Neural Network. It was also observed that using the public mood as a feature increased the accuracy by 20% compared to only using historical data for the purpose of prediction. ROST Content Mining is used as a tool to analyse sentiments. It is a widely used emotion analysis tool and it is believed to be one of the best tools to analyse Chinese emotions. It is a typical three-class mood analyser. For keyword spotting method, C-POMS is proposed as it chooses Brief Chinese POMS as seeds to create a 7 moods' dictionary. After this a Granger Causality Test is carried out to check which factors of historical data are aligned with the 7 features identified earlier. The output is then taken as input for the two testing methods named SVM and Probabilistic Neural Network, where the former performed better than the latter.

Shen et al [19] in their paper proposed the use of an improvised version of SVM. They experimented on the data collected from the US NASDAQ. In the market trend prediction, SVM was tested by passing different number of features as input, but it was observed that SVM performed with more accuracy when there was a smaller number of features, the prediction accuracy started reducing once the number of features increased. In the regression experiments, based on the RMSE values SVM again performed well when it was tested with other techniques like Linear Regression, generalized linear model etc. Hence, SVM was considered to be better as compared to other algorithms.

Batra et al [23] have proposed a different way of using the SVM algorithm. In their paper, they have extracted the sentiment from twitter using StockTwits [23] via a pipeline API of python. The data extracted using the former was pre-processed for sentiment analysis and Natural Language Processing (NLP). SVM is used to predict the sentiment of each data and then classifying the tweets into positive and negative tweets for easier and better prediction models. The output of this is then combined with the available historical data and used for stock price prediction. It was observed that the accuracy was around 76.65%.

2) Multi-Source Multiple Instance Learning(MMI): Zhang et al [15] proposed an improvement of the Multiple Instance Learning model that could integrate the features from multiple sources to get accurate predictions related to the stock prices. They developed a multi-source multiple instance model that could combine sentiments, events and the quantitative data into a comprehensive framework to predict better results. Essentially the structured events were acquired from the news events and then they were used as the input data for the Restricted Boltzmann Machines for the training of the algorithm. After this, the output of RBM is used as input to the sentence2vec framework. This model was compared with SVM, TeSia, n-MIL, WoR-MI, O-MI and WoH-MI. It was observed that MMI performed the best with 60.10% accuracy.

B. Regression

1) Linear Regression: Linear regression is a technique used to predict the relationship between the dependent and independent variable. Relationship between the two variables is said to be deterministic if one variable can be accurately expressed by the other.

Roy et al [31] proposed a modification of the least square method, the LASSO (Least Absolute Shrinkage and Selection Operator) which was based on a linear regression model. This method was able to produce sparse solutions and performed well when the numbers of features were less as compared to the number of observations. It was used to predict the future price of the chosen stock. The dataset used was Goldman Sachs

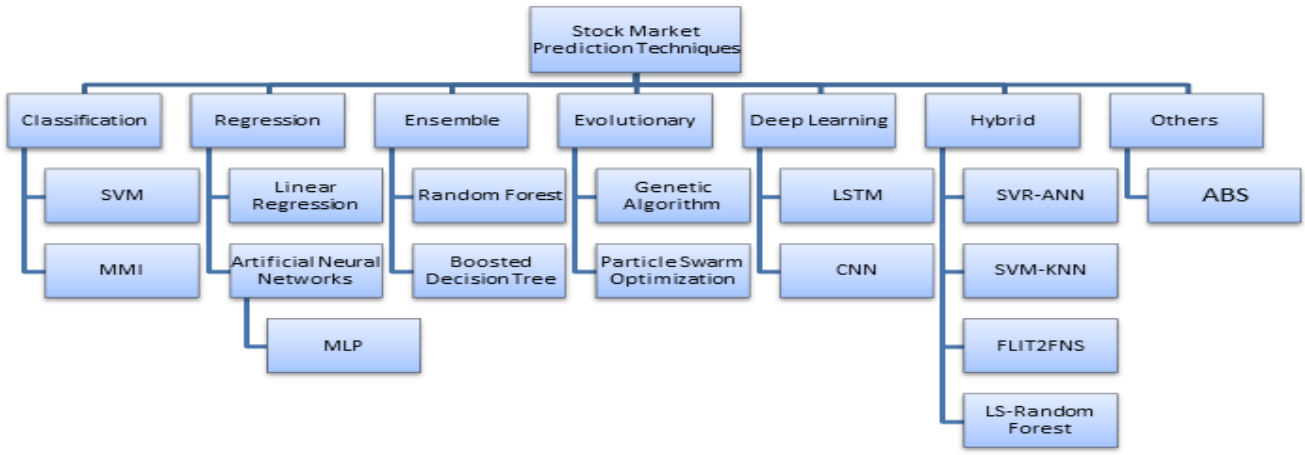

Fig. 1 Classification of Stock Market Prediction Techniques

Group. The model's performance was compared to ridge regression and it was found out the MAPE of LASSO (1.4726) was less, compared to ridge regression (1.8065).

2) Artificial Neural Networks(ANN): Artificial Neural Networks (ANN) [29] is a computing system which uses the idea of using the structure of brain as a model to develop such methods that assist in building better prediction models. These system models have the tendency to learn based on their training datasets. ANN is built from a collection of connected nodes called neurons.

Moghaddama et al [39] considered the short-term historical prices as well as the day of the week for prediction. They used BPNN for stock prediction. The model was trained on various parameters by LM, GDA and OSS.A network with 20-40-20 neurons in hidden layers OSS training method and LOGSIG transfer function was found out to be the optimized network with R2 of 0.9622.

a) Multi Layer Perceptron(MLP): Multi-Layer Perceptron (MLP) is a sub domain of artificial neural networks. It is a logistic regression classifier technique in which the input data is converted with the help of previously learnt non-linear transformation methods to create intermediate hidden layers. These hidden layers further make the MLP to a universal approximator.

Turchenko et al. [33] in their research used a neural network technique, Multi-Layer Perceptron to build a short-term prediction model. This prediction model focused on the use of historical data to predict the stock prices for a company. Stock of Fiat company was tested using the MLP technique to test the prediction model. In their research, it was observed that MLP performed well and gave results with very less error. The average relative prediction error calculated was around 2.2%. Most of the experiment results showed prediction error less than or equal to 5%, in which 33% of the experimental data had less than or equal to 1%.

Whereas, Shi et al. [5] in their research used Multi-Layer Perceptron to build a prediction model for the stock market in China. In their research work they found out that the sentiments play a more significant role in prediction than the historical data. They gathered their dataset using a Chinese social networking site - Xueqiu. From this they extracted the news feeds and user posts related to stocks. Using MLP and Naive Bayes, tweets were then divided into three classifications: neutral, positive and negative. Their research work was based on 3 features namely sentiment feature, stock specific features (SSF) and stock relatedness feature. The MLP technique was compared to SVM and using the accuracy parameters and area under the curve parameter it was observed that MLP using stock specific features and sentiment data proved to be better than the same being used in association with SVM.

C. Ensemble Techniques

In machine learning, Ensemble learning is a technique that uses multiple prediction models. As this model uses more than one prediction technique, it becomes more flexible and less data sensitive. Following are some ensemble techniques used for prediction.

1) Random Forest: Random forest technique is a supervised type of machine learning algorithm. It produces multiple decision trees and later combines them together to obtain more stable and more accurate predictions. Random forest can be used for both regression and classification problems in machine learning. While growing the tree, random forest adds functionality of randomness.

Gryp et al. [10] compared ensemble techniques like Random Forest, AdaBoost and Kernel Factory against classifier models such as Logistic Regression, Neural Networks-Nearest Neighbor, and Support Vector Machines. They fetched data from over 5767 European companies from the Amadeus Database and used the Area under the curve (AUC) as a performance measure. In the results it was observed that the Random forest algorithm outperformed other algorithms

followed by Support Vector Machines, Kernel Factory, AdaBoost.

Stajduhar et al. [13] in their research, used historical data to predict the future stock prices. In this approach, the machine learning technique used was Random Forest algorithm. This algorithm was used as a classifier in association with the WEKA (Waikato Environment for Knowledge Analysis) toolkit based on the CROBEX index. The model was tested on a 5-days and a 10-days ahead forecasting experiment. The results obtained were really good, having 76.5% accuracy for 5-day ahead and 80.8% for 10-day ahead. It was observed that random forest algorithm predicted the market trend with good accuracy, hence this model of approach was recommended by the authors.

Basaka et al. [32] compared the performance of Random Forest and XGBoost.Only the closing price of the stocks was considered for the prediction and the following indicators were calculated from it :Relative strength index (RSI),Stochastic oscillator (SO), Williams percentage range (W%R),Moving average convergence divergence (MACD),Price rate of change (PROC), On balance volume (OBV).It was found out that Random Forest performed better than XGBoost.

2) Boosted Decision Tree: Boosted Decision tree is a machine learning algorithm which is an implementation of gradient boosted trees. In this stepwise regression trees are used to make a prediction model which comprises of multiple weaker prediction models.

Pai et al. [8] in their research work observed that they got around 70% accuracy with the implementation of supervised learning algorithms on a day to day prediction model. Their work was divided into two phases, first one predicted the stock prices on a daily basis and the other one predicted stock prices on a monthly basis. Data for experiment was extracted from Yahoo Finance, through which the closing and opening prices of stocks were considered. The model built considered both sentiment data and historical data. It was observed that on the considered dataset, Decision Boosted Tree (DBT) gave more accurate results as compared to SVM and Linear Regression.

D. Evolutionary

1) Genetic Algorithm: It's a searching technique which is often used in computing an approximate solution to search and optimize the problems. They are a particular class of evolutionary computation that uses evolutionary biology such as inheritance, mutation, selection, and crossover.

Sable et al. [28] in their research work used the genetic algorithm to predict the stock market. A total of six attributes were used for the prediction. They compared the performance of genetic algorithms with other evolutionary algorithms and found out that the accuracy of genetic algorithm was much better (73.87%).

2) Particle Swarm Optimization(PSO): It's a computational technique which uses multiple iterations in lieu of finding the best candidate solution. A problem using this technique is solved by having a population of candidate solutions and moving these in the entire search space based on simple mathematical operations. The main aim behind this is to escort a particle to the best-known point in the search space.

Majhi et al. [30] in their paper proposed the use of Particle Swarm Optimization to predict the stock prices of S&P 500 and DJIA stocks. Initially for the experiment, k patterns were defined which were passed through a linear combiner, multiplied by their partial weights and the output of this was compared with the actual stock prices to calculate the MSE. Also, the mean absolute percentage (MAPE) was also calculated for comparison with MLP. The comparisons were done on the basis of 1 to 7 days. The results portrayed that PSO had a lower MAPE value as well as lower computational time compared to the latter. Hence PSO was proved to be the better one among the two.

Nenortaite et al. [36] used Particle Swarm Optimization to build a model that could help in predicting the best stock trading system, so that the investors could invest right and gain good profits. The model has been compared using the historical data of SP500, with other techniques and it was found out that when commission rates were less, the PSO model was better and suggested other trading stocks but it was not the case when the commission was on a higher side. In this, data was preprocessed and ranks were given to all the indices, later using PSO the "global best" particle was found. 5-7 times better results were obtained using this technique.

E. Deep Learning

Deep Learning in stock market prediction [14] is a technique which is used to acquire abstract features and recognize the hidden features in financial markets without making any assumptions. It is a technology that extracts features mechanistically. In its working, at first it extracts low level properties from original signals and after that it extracts high level properties.

Chen et al. [2] in their research work used the deep learning approach to forecast the stock market behavior and trend, and then compared the performance of this method with the performance of back propagation (BP) method, radial basis function (RBF) and extreme learning machines (ELM). In their work, they made use of auto-encoder & RBM methods for deep learning. It was observed that the RMSE value of this approach was around (1.0437), and the MAPE value of DL was (0.0002). The deep learning approach performed the best as it had the lowest error on the available dataset.

Deep learning further was implemented using few of its classification like, Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) which are discussed below.

1) Convolutional Neural Network(CNN): CNN is a technique which automatically detects the important features without any human interference and it's computationally

efficient as well. It uses a variation of MLP as it consists of an input layer, multiple hidden layers and an output layer.

Hiransha et al. [34] in their paper used 4 types of deep learning models, CNN, LSTM, MLP, RNN to predict the stock price using the historical data. The training of the network was done using the data from NSE and the prediction was done for both NSE and NYSE. This was possible because both the markets share some common dynamics. Normalization of data was done to bring the data to a common range. Only the closing price was extracted from the dataset for prediction. It was observed that CNN outperformed all other models.

2) Long Short Term Memory(LSTM): According to [35], LSTM is a special type of RNN (Recurrent Neural Network) which overcomes the limitations of RNN. One major limitation of RNN is that it lags learning properly when the extent between the point of requirement and required previous information increases. LSTM is designed to avoid the long term dependency problem. It proves to be skilful at retaining values for either short or long durations of time.

Gao et al. [22] proposed LSTM. In their model, there were three layers in neural network including the input layer, the output layer and the hidden layer and every unit in a layer was connected with all the units in adjacent layers. They used MA, EMA, SVM to compare the outputs. It was observed that the MAPE of LSTM was the least (0.7240).

Nelson et al. [40] also used LSTM to build their prediction model. They used historical data from the Brazilian Stock Exchange to predict the stock market price movement. Initially in their model, the acquired data was preprocessed and then later used for experiments. It was found that the use of LSTM gave better results than Random Forest and MLP. Average accuracy for this prediction model was around 55.9%.

F. Hybrid Methods

To get better prediction results, many researchers proposed the idea of combining two or more prediction algorithms. Few such models have been discussed below:

1) Support Vector Regression and Artificial Neural Network (SVR-ANN): Patel et al. [9] proposed a two-stage fusion approach involving Support Vector Regression (SVR) in the first stage and Artificial Neural Network in the second stage of the fusion. The output of the first stage was treated as an input for the second stage. The SVR-ANN technique was compared with 5 other models SVR-SVR, SVR-Random Forest, Random forest, ANN and SVR. Predictions were made on the basis of 1-10 days, 15 days and 30 days. As a result, it was observed SVR–ANN performed better than the other models and gave better predictions.

2) Support Vector Machine and K-Nearest Neighbour (SVM-KNN): Nayak et al. [11] in their paper proposed the fusion of two machine learning algorithms SVM and KNN to build their prediction model. In this hybrid model, the output of prediction model SVM assists in computing the nearest neighbours as the input for KNN algorithm. The dataset for experiments has been extracted from BSE India. SVM is used to train the model and the outcome of SVM is then compared with class labels to compute accuracy, later the mean of K-nearest neighbour is computed. In their research work, SVM-KNN model was compared with two different hybrid models FLIT2FNS [20] and CEFLANN [21]. The evaluation parameters to compare these models were MSFE, RMSFE, MAFE based on the different time horizons. It was observed that SVM-KNN performed 5 to 15 times better than FLIT2FNS and 10 to 15 times better than CEFLANN.

3) FLIT2FNS: Functional Link Interval Type-2 Fuzzy Neural System (FLIT2FNS) is a combination of FLANN (Functional Link Artificial Neural Network) and interval type-2 fuzzy logic system. Chakravarty et al. [20], in their paper proposed the use of this model as it was proved to give better prediction accuracy. This model was compared with various other models like Functional Link Artificial Neural Network (FLANN) [25] & Type-1 and Local Linear Wavelet Neural Network (LLWNN) [26]. It was compared on the evaluating parameters of MAPE and RMSE. It was observed that after conducting multiple experiments on the available dataset, the MAPE result for FLIT2FNS was 0.32% whereas for FLANN & Type-1 it was 1.1% and for LLWNN it was around 1.4%. Hence, FLIT2FNS was proved to be the working better on the given dataset.

4) LSboost-RF: N. Sharma et al. [3] proposed to collaborate Random Forest using LSboost with the predictions of ensemble trees. LSboost (Least Square Boost) technique was used as a training loss function. For combining the estimates of ensemble of multiple trees in random forest, LSboost was used. The idea behind this model was that the Mean Squared Error (MSE) used in Random Forest produced poor estimations in certain scenarios [24]. Therefore, to improvise upon the error estimates, LSboost was used as a training loss function. Its performance was compared to the SVR technique and the MAPE value of the former was found out to be 0.5745.

G. Others

1) Aspect Based Sentiments(ABS): Thien Hai Nguyena and Kiyoaki Shirai [6] built a model to use the sentiment data extracted from social media sites to predict the stock price trend. In this approach, the sentiments related to certain companies were integrated in this prediction model. They discovered that along with the people's sentiment, stock rates are also affected by more factors like macroeconomic and microeconomic factors. The paper proposed a new feature which was topic-sentiment which improved the performance of the above-mentioned approach. For this, two different datasets were used for conducting experiments. First one being the historical stock data and the other one being the people's mood

dataset. Mood information dataset was based on the fact that there were two kinds of messages. The first type of message is the one in which a new post is generated that initiates a fresh topic of discussion and the second type of message was a response message to the pre-existing ones. Many users' posts are response messages. This paper displayed the effectiveness of incorporating the sentiments by investigating on a large dataset. After experiments, it was observed that this method gave 2.07% times more accuracy than historical data prediction models.

III. Literature Ananlysis

From the above discussed techniques, a brief summary in Table 1 is derived displaying the techniques used or proposed in the various research papers and their corresponding results. The table also tells about whether preprocessing is done on the datasets obtained from relative sources or not. It was also observed that not all approaches used the sentiment data to build their prediction model, though in each of the cases, historical data analysis was a must to forecast the stock prices.

In the table 1 below, the results have been deduced based on either one these: their accuracy (as per the testing dataset), mean absolute percentage error (MAPE), root mean square error (RMSE), Area Under the Curve (AUC), Determination coefficient (R^2) or Inter Quartile Range (IQR).

TABLE I. Summary of all the techniques used/proposed with their corresponding Results

Reference	Historical Datasets	Sentiment Analysis	Pre processing	Techniques Proposed/ used	Results
[6]	Yahoo Finance	✓	✓	Aspect Based Sentiment	2.07% more accurate than historical data models
[8]	Yahoo Finance	✓	✓	Boosted Decision Tree	76% (Accuracy)
[2]	Chinese securities index 10 (CSI 300)	X	✓	Deep Learning	1.0437(RMSE) 0.0002(MAPE)
[20]	S&P Global, BSE India, DJIA	X	✓	FLIT2FNS	0.32%(MAPE)
[28]	NASDAQ	X	X	Genetic Algorithm	73.87%(Accuracy)
[27]	NSE India	X	✓	Linear Regression	90%(Variance)
[3]	CNX Nifty S&P 500 BSE	X	✓	LS-Random Forest	0.5745(MAPE)
[22]	S&P 500	X	X	LSTM	0.7240(MAPE)
[5]	TuShare API	✓	✓	MLP	0.56(AUC).
[15]	Wind Financial Terminal	✓	X	MMI	60.10% (Accuracy)
[30]	DJIA, S&P 500	X	X	Particle Swarm Optimization	0.6558(MAPE)
[10]	Amadeus Database	X	✓	Random Forest	0.0061 (Least IQR)
[13]	Zagreb Stock Exchange	X	✓	Random Forest	80.8%(Accuracy)
[7]	Tongdaxin	✓	✓	SVM	71.429%(Accuracy)
[19]	NASDAQ	X	✓	SVM	74.4%(Accuracy)
[23]	Yahoo Finance	✓	✓	SVM	76.65%(Accuracy)
[11]	BSE India	X	✓	SVM-KNN FLIT2FNS [20] CEFLANN [21]	0.1123 (MAPE) 0.61 (MAPE) 1.8 (MAPE)
[9]	BSE India, NSE India	X	✓	SVR-ANN	2.66 (MAPE)
[31]	Goldman Sachs	X	X	Regression	1.4726 (MAPE)
[32]	AAPL	X	✓	Random Forest	78% (Accuracy)

[33]	Fiat Company	X	X	MLP	<= 5 (MAPE)
[34]	NSE, NYSE	X	✓	CNN	5.31 (MAPE)
[39]	NASDAQ	X	X	BPNN	0.962(R^2)
[40]	Brazilian Stock Exchange	X	✓	LSTM	55.9% (Accuracy)
[36]	SP500	X	✓	PSO	5-7 times better results

IV. Challenges and Future Scope

Predicting the stock market prices and trend is not an easy task, as it involves incorporating many distinct factors that affect the stock market immensely. No stock market prediction model till now has predicted the accurate prices/trend. While going through the literature, it was found that there are many challenges/hinderances that the researchers had to face, which prohibited them from developing a perfect model. One of the major challenges faced in stock market prediction is that solely the historical data of the stock market isn't sufficient to predict the stock market. The news events and the sentiment data play a major role in addition to historical data as they also act as a factor in affecting the stock prices on daily basis. Another challenge that was faced is the extraction of the sentiment data or the financial news related to certain stocks from microblogging websites such as Twitter or Xueqiu [5] and then to preprocess them as per the model. There are chances that people may even post fake good/bad reviews related to a particular stock so that they can either uplift or downgrade their reputation. One more challenge that was observed was that, some models used predefined values of number of topics and sentiments, which is not efficient. In order to overcome this, the model should be developed in which the two factors can be automatically acquired. Apart from these, a company's balance sheet or the cash flow [6] can be considered to be important factors to forecast stocks.

In future, it is advised to devise such models that accommodate more features and can also overcome the above-mentioned challenges. As it is observed from the literature analysis, the hybrid/fusion models should be preferred to build new prediction models as they have the potential to use the capabilities of their individual components and help in forecasting more accurate results. These hybrid methods are also capable of overcoming the challenges as it can use different techniques to acquire and process the sentiment data, news events and historical data.

V. Conclusion

As the stock market is too uncertain, the investors must invest their money after assessing the affecting factors such as public reviews, historical data and news events. Many researchers have tried to devise prediction models using machine learning algorithms to predict the accurate prices of stocks using various tools and techniques, but have yet not been able to come up with the best possible solution. Our paper overall summarizes a few of the machine learning techniques that have been used by the research scholars to forecast the stock market trend and prices using machine learning and artificial intelligence algorithms, keeping in mind the extensive detailing, features and parameters involved. Even though after analyzing the major affecting factors and incorporating the social reviews related to stock, the accurate prediction of stock price is not possible. There are some techniques that have been able to get a really close approximation. The major techniques which have been reviewed in this paper are Support Vector Machine, Random Forest, Boosted Decision Tree, Evolutionary methods, Deep Learning methods, Regression etc. Also, some of the hybrid techniques [3][9][11] [20] have also been used to build prediction models for the same, such techniques are a combination of two or more algorithms used in correspondence to each other like SVM & KNN, SVR & ANN etc. It is possible to utilize any of the techniques and develop a hybrid system model for the stock market price prediction, but on the other hand, it is important to design the system in which the accuracy and performance can be increased with less computational complexity.

Though each model has its own advantage and disadvantage over others based on their evaluating factors and the datasets used for their experiments. Some models work better with historical data and others with sentiment data. Based on the literature analysis, among all the models discussed, the fusion algorithms predicted results with higher precision. They incorporate the important features of the individual techniques that they are comprised of and a result the computational time was also improved as compared to other prediction models.

References

[1] Jitendra Kumar, Ashutosh Kumar Singh, Workload prediction in cloud using artificial neural network and adaptive differential evolution, Future Generation Computer Systems, Volume 81, 2018, Pages 41-52, ISSN 0167-739X

[2] L. Chen, Z. Qiao, M. Wang, C. Wang , R. Du and H. E. Stanley: "Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?," in IEEE Access, vol. 6, pp. 48625-48633, 2018.

[3] N. Sharma and A. Juneja: "Combining of random forest estimates using LSboost for stock market index prediction," 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, 2017, pp. 1199-1202.

[4] Q. LI, L. JIANG,P. LI, H. CHEN : Tensor-Based Learning for Predicting Stock Movements. AAAI Conference on Artificial Intelligence, North America, feb. 2015.

[5] Xi Zhang,Jiawei Shi,Di Wang, BinxingFang : Exploiting investors social network for stock prediction in China's market, Journal of Computational Science, Volume 28, 2018

[6] Thien Hai Nguyen,Kiyoaki Shirai, Julien Velcin : Sentiment analysis on social media for stock movement prediction, Expert Systems with Applications, Volume 42, Issue 24, 2015, Pages 9603-9611, ISSN 0957-4174

[7] D. YanG. Zhou, X. Zhao, Y. Tian and F. Yang: "Predicting stock using microblog moods," in China Communications, vol. 13, no. 8, pp. 244-257, Aug. 2016.

[8] Aparna Nayak, M. M. ManoharaPai, Radhika M. Pai : Prediction Models for Indian Stock Market, Procedia Computer Science, Volume 89, 2016.

[9] Jigar Patel, Sahil Shah, Priyank Thakkar, K Kotecha: Predicting stock market index using fusion of machine learning techniques, Expert Systems with Applications, Volume 42, Issue 4, 2015

[10] Michel Ballings, Dirk Van den Poel, Nathalie Hespeels, Ruben Gryp: Evaluating multiple classifiers for stock price direction prediction, Expert Systems with Applications, Volume 42, Issue 20, 2015

[11] Rudra Kalyan Nayak,Debahuti Mishra, Amiya Kumar Rath: A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices, Applied Soft Computing, Volume 35, 2015

[12] Khalid Alkhatib, NajadatHassan, Ismail Hmeidi&K Ali Mohammed Shatnawi: Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm(2013).

[13] T. Manojlović and I. Štajduhar: "Predicting stock market trends using random forests: A sample of the Zagreb stock exchange," 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2015, pp. 1189-1193.

[14] Dang, L. Minh & Sadeghi-Niaraki, Abolghasem& Huynh, Huy& Min, Kyungbok& Moon, Hyeonjoon. (2018). Deep Learning Approach for Short-Term Stock Trends Prediction based on Two-stream Gated Recurrent Unit Network. IEEE Access

[15] X. Zhang,S. Qu, J. Huang, B. Fang and P. Yu : "Stock Market Prediction via Multi-Source Multiple Instance Learning," in IEEE Access, vol. 6, pp. 50720-50728, 2018

[16] T. Mankar, T. Hotchandani ,M. Madhwani , A. Chidrawar and C. S. Lifna: "Stock Market Prediction based on Social Sentiments using Machine Learning," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-3

[17] Lucas Nunno : "Stock Market Price Prediction Using Linear and Polynomial Regression Models." (2014).

[18] S.P Pimpalkar., JenishKaria, Muskaan Khan, Satyam Anand, Tushar Mukherjee: Stock Market Prediction using Machine Learning. International Journal of Advance Engineering and Research Development (IJAERD) Special Issue on Recent Trends in Data Engineering, Volume 4, Special Issue 5, Dec 2017

[19] ShunrongShen, HaomiaoJiang and TongdaZhang : "Stock Market Forecasting Using Machine Learning Algorithms"(2012).

[20] S. Chakravarty, P.K. Dash: A PSO based integrated functional link net and interval type-2 fuzzy logic system for predicting stock market indices, Applied Soft Computing, Volume 12, Issue 2, 2012

[21] RajashreeDash, P.K. Dash, RanjeetaBisoi : A self-adaptive differential harmony search based optimized extreme learning machine for financial time series prediction, Swarm and Evolutionary Computation, Volume 19, 2014

[22] T. Gao, Y. Chai and Y. Liu :"Applying long short term momory neural networks for predicting stock closing price," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578.

[23] R. Batra and S. M. Daudpota: "Integrating StockTwits with sentiment analysis for better prediction of stock price movement," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp. 1-5.

[24] Carolin Strobland Z. Achim: "Danger: high power!–exploring the statistical properties of a test for random forest variable importance,"2008.

[25] R. Chodhury, K. Garg: A hybrid machine learning system for stock market forecasting, in: Proceeding of World Academy of Science, Engineering and Technology, vol. 29, 2008, ISSN 1307-6884.

[26] Y. Chen, B. Yang, J. Dong : Time series prediction using a local linear wavelet neuralnetwork, Neurocomputing, Elsevier 69 (2006) 449–465.

[27] M.A. Shah,C.D. Bhavsar : Predicting Stock Market using Regression Technique(2015).

[28] S. Sable,A.Porwal and U. Sing : "Stock price prediction using genetic algorithms And evolution strategies," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore,2017, pp. 549-553.

[29] Egeli Birgul, OzturanMeltem, BadurBertan. (2019). Stock Market Prediction Using Artificial Neural Networks.

[30] R. Majhi, G. Panda, G. Sahoo,A. Panda and A. Choubey : "Prediction of S&P 500 and DJIA stock indices using Particle Swarm Optimization technique," 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp. 1276-1282.

[31] Roy S.S., Mittal D., Basu A., Abraham A. (2015) Stock Market Forecasting Using LASSO Linear Regression Model. In: Abraham A., Krömer P., Snasel V. (eds) Afro-European Conference for Industrial Advancement. Advances in Intelligent Systems and Computing, vol 334. Springer, Cham

[32] Suryoday Basak, Saibal Kar, SnehanshuSaha, LuckysonKhaidem, Sudeepa Roy Dey, Predicting the direction of stock market prices using tree-based classifiers, The North American Journal of Economics and Finance, Volume 47, 2019, Pages 552-567, ISSN 1062-9408V.

[33] Turchenko, P.Beraldi ,F. De Simone and L. Grandinetti , "Short-term stock price prediction using MLP in moving simulation mode," Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems, Prague, 2011, pp. 666-671.

[34] Hiransha M, Gopalakrishnan E.A., Vijay Krishna Menon, Soman K.P., NSE Stock Market Prediction Using Deep-Learning Models, Procedia Computer Science, Volume 132, 2018, Pages 1351-1362, ISSN 1877-0509

[35] Jitendra Kumar, Rimsha Goomer, Ashutosh Kumar Singh,Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters,Procedia Computer Science,Volume 125,2018,Pages 676-682,ISSN 1877-0509

[36] Nenortaite J., Simutis R. (2004) Stocks' Trading System Based on the Particle Swarm Optimization Algorithm. In: Bubak M., van Albada G.D., Sloot P.M.A., Dongarra J. (eds) Computational Science - ICCS 2004. ICCS 2004. Lecture Notes in Computer Science, vol 3039. Springer, Berlin, Heidelberg

[37] A. Gavhane, G. Kokkula, I. Pandya and P. K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 1275-1278.

[38] A. Varma, A. Sarma, S. Doshi and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1936-1939.

[39] Amin Hedayati Moghaddam, MoeinHedayati Moghaddam, MortezaEsfandyari, Stock market index prediction using artificial neural network, Journal of Economics, Finance and Administrative Science, Volume 21, Issue 41, 2016, Pages 89-93, ISSN 2077-1886

[40] D. M. Q. Nelson, A. C. M. Pereira and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 1419-1426.

[41] Ashutosh Singh & Jitendra Kumar (2019). Secure and Energy Aware Load Balancing Framework for Cloud Datacenter Networks. Electronics Letters. 10.1049/el.2019.0022.

[42] J. Kumar and A. K. Singh, "Dynamic resource scaling in cloud using neural network and black hole algorithm," 2016 Fifth International Conference on Eco-friendly Computing and Communication Systems (ICECCS), Bhopal, 2016, pp. 63-67.