

Stock Market Forecast Analysis using Machine Learning Mechanism

Kushal Patel
Student,
ME-CE Department
LJ. Institute of Eng. Technology.

Swarndeept Saket
Assistant Professor,
ME-CE Department
LJ. Institute of Eng. Technology.

Abstract - Nowadays everyone is investing in the stock market. The stock market is daily stock buying or selling stocks. Stock market prices fluctuating are difficult to forecast. Prices are unpredictable and there is a huge risk of the investor to get maximum profit or loss of investment in stock market. Fundamental analysis based on real-time news, companies market value, balance sheet or growth, stock market index and others facts. Based on historical data through technical analysis will predict stock market apply the various machine learning mechanism to forecast stock market. Various mechanisms are support vector machine, predictive analysis, supervised machine learning algorithm, regression algorithm and etc. The proposed system forecast the stock prices and evaluate result with best accurate predictable prices. The objective is to improve in predictive performance of the stock market with the technical indicators and verifies targeted value of forecast with the comparison of machine learning mechanism. One year historical data apply three machine learning mechanism. Compare the forecast set value and confidence- accuracy. Accuracy find good at level. Analysis trend based on the graphs and getting idea about current and future market situation.

Keywords - *Stock market forecasting; Machine learning mechanism; Predictive Analysis; Predictable Price; Support Vector Machine*

I. INTRODUCTION

Stock market is an electronic platform where investors come together to buy and sell their equity shares.[7]. Rapidly changes stock prices, it is a very challenging task to forecast. stock market is highly volatile [16]. Some real-life news impact on the stock market. Every country economic growth depends upon stock market performance. Stock market prediction using emerging machine learning mechanism to forecast targeted value. Predicting prices based on closing price of stocks and give accurate predictable price value.

Stock market using machine learning mechanism where support vector machine, supervised machine learning algorithms, k-nearest neighbor, random forest and many more mechanisms to improve stock market forecasting. Proposed flowchart shows the machine learning use for better prediction.

An investor gets maximum profit from the predictable price and detail analysis of the current scenario of the stock market. Stock market depends on the main three markets are currency market, commodity market and stock exchange. India's main stock market[7] are National Stock Exchange[NSE] and Bombay Stock Exchange[BSE].

I.1 MACHINE LEARNING MECHANISM

Machine learning is term as " field of study that gives computers the ability to learn without being explicitly programmed. Algorithm design with the maximum features. Features in terms of stock market are stock high price, close price, high price at ⁱth time on particular day/month/year.

Machine learning in finance[17]:

- Reduced operational costs thanks to process automation.
- Increased revenues thanks to better productivity and enhanced user experiences.
- Better compliance and reinforced security.

Applications of Machine Learning in Finance Services[17,18]:

- Process Automation
- Algorithmic Trading
- Fraud Detection
- Underwriting and credit scoring
- Customer services
- Security
- Sentiment/ News analysis

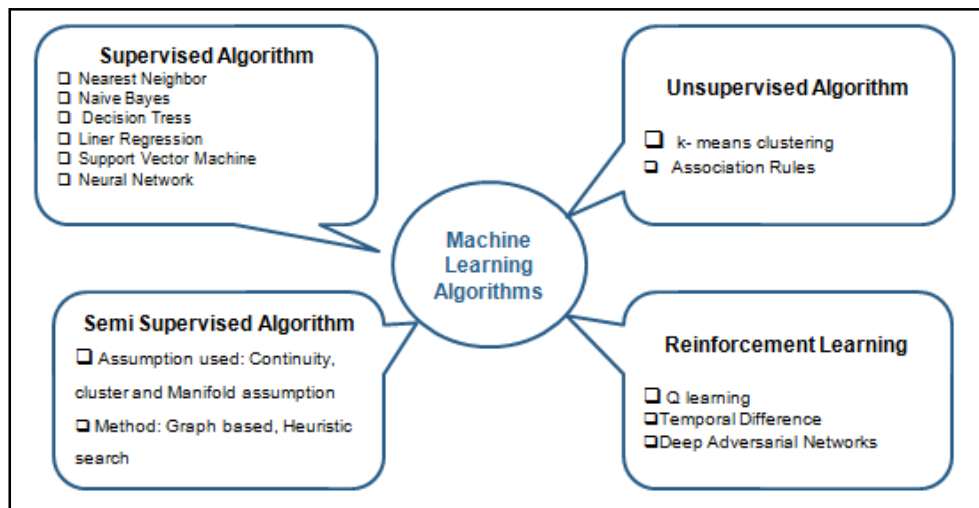


Figure1. Machine Learning Algorithms

I.2 PREDICTIVE ANALYSIS

It is the process where model created for better predictable outcome. Predictive analysis is based on machine learning and statistical analysis. Many other techniques are categories but improve forecast stock market and future of stock market activities.

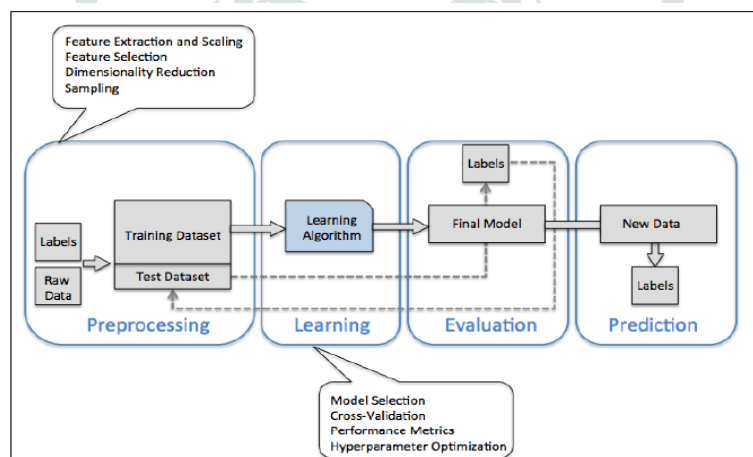


Figure2 : Machine learning in predictive analysis^[6]

- Preprocessing[12]- Normalization[10], Missing value, learn data
- Learning- Various machine learning techniques to train data, parameters
- Evaluation- Calculate performance like hit ratio, accuracy, moving average, percentage error, optimum time
- Prediction- Targeted value as an outcome.

I.3 STOCK MARKET ANALYSIS

Stock market analysis[2] for predication [18]based on technical analysis/ indicator[3] and fundamental analysis[5]. Investment of any stock investor analysis and making decision to buy or sell that stock. Investor invest in any sector but generally aim is to gain maximum profit at the end. Various stock market strategies[7] to analysis stock market situation.

A. Technical Analysis:

Analysis with historical stock data. Technical analysis consider short term approach for share price. Taking decision as short term movements in stock market price. Technical analysis is using for trade. Investor or market researcher get idea to trade in stock market for long term or short period of investment.

B. Fundamental Analysis:

Fundamental analysis revolves the company's balance sheet. It will take long term evaluation. Fundamental of company do not change on daily basis[7]. Fundamental analysis used to make an investment decision. Everyone hope to receive maximum outcome. Fundamental analysis through we can conclude investment decision.

Table 1: Stock Market Analysis

<i>Technical Analysis</i>	<i>Fundamental Analysis</i>
<ul style="list-style-type: none"> • Previous price to predict future price[2]. • Used for trading. 	<ul style="list-style-type: none"> • Study of economy and industry condition • Make an investment decision. • Evaluating future share price.

Some of the method[7] to find out fundamental analysis and evaluating future price bases on historical price chart, trends, pattern matching, statistical analysis etc..

- Moving Average Method
- Top down and Bottom up
- P/E ratio
- P/BV ratio
- Candlestick Techniques
- Trends: Triangle and flags
- Head and Shoulder Methods

II. RELATED WORK

Meryem Ouahilal et al.[1] proposed support vector machine for optimizing the prediction of stock prices. To demonstrate the improvement in predictive performance of stock market and verifies with the optimized model. Data collecting then after Hodrick- Prescott filter techniques used for noise filtering. Further support vector machine for predicting model and gives predict of best stock prices. Learning historical data to real-world data set for prediction. Stock prediction through very minimal execution time and improving accuracy. Bruno Miranda Henrique et al.[2] Efficient Market Hypothesis [EMH] through analyzing the market situation, financial news on daily bases predicting of stock market. Large and small dataset through Support Vector Regression[2] analysis and predict daily and up- to minute frequencies. Prediction error measures and compared with model by EMH. Up to minute data compared with the short period of time data will hit maximum data points. In support vector regression[2] selecting appropriate kernel to get the best outcome. Root Mean Squared Error(RMSE), Mean Absolute Percentage Error(MAPE) methods calculate the prediction error.

Elham Ahmadi et al.[3] proposed hybrid model for timing of stock market based on technical analysis. Imperialist Competition Algorithm on SVM to optimize parameters. For feature selection Genetic algorithm is used. Based on stock data compared with model and as outcome Hit Rate measure for the correct prediction. Based on data selecting accurate features, further data checking and then calculating hit rate. Hit rate measure on the short period of time by both Raw and Signal approaches. GA used for selects optimum features for better prediction. Aparna Nayak et al. [4] Two models created. Historical prices of stock through sentiment combined for daily prediction models. Supervised algorithm are used for models. Sometimes prices are going same or trend based prediction to monthly prediction model. Analysis of historical prices where prices volatile changes up/down. Same day volume traded is compared with the trend on the same day to get volume pattern. where sentiment analysis used accuracy of support vector machine is less compared to other mechanisms. Impact of intraday price movement for prediction of stock market next day price for better accuracy.

Indu Kumar et al.[5] machine learning techniques used for stock price prediction. Models are support vector machine(SVM), Random Forest, K-Nearest Neighbor(KNN)[33], Naive Bayes and Softmax. Performance and comparing stocks with market trends. For the small term, dataset Naive Bayesian Classifier is used. For the large dataset, Random Forest algorithm is used. Technical indicator is analyzing the data with some techniques used to predicting stock market trend and reduced number of technical investors. Naive Bayes gives the best outcomes in term of accuracy.

Table 2: Comparative Analysis

Author and Year	Methodology	Parameters	Outcomes
Meryem Ouahilal et al. (2017)	SVR, Hodrick Prescott filter	Accuracy, Optimizing	Model is sufficient to predict stock prices. Best accuracy and optimize
Bruno Miranda	SVR	RAME,MAPE	Prediction Error

Henrique et al. (2018)			
Elham Ahmadi et al.(2018)	SVM,GA	Hit rate	SVM-ICA hit rate 76%
Aparna Nayak et al.(2016)	LR, SVM, Decison tree	Accuracy	Decision tree is better than others.
Indu Kumar et al(2018)	KNN, SVM, RF, Naive Bayes	Accuracy	Naive Bayes is better accuracy
Yang Jaio et al. (2017)	Elston's, LR, RF, NN, DT	Prediction Performance- Sequential split	Logistic Regression is better
Mahek Usmani et al. (2016)	NN, SVM	Accuracy	MLP- 77%
Li Tang (2018)	KNN,PCA	Hit rate	77.58% , PCA-KNN more effective for prediction.

III. METHODOLOGY

There are many machine learning mechanism but few of them compare with each other and gives better outcomes as stock market prediction in the end.

A. Support Vector Machine

Support vector machine[27,34]is very strong and useful in financial time series and predicting[1]. Data will be analysis for regression and classification analysis[13]. Selection of kernel[1]in SVM is beneficial for maximum margin[7,18]. SVM is separating two classes using linear boundary[3]. Closest to linear boundary are Support vectors. It's used in highly dimensional[8] data of previous stocks rapidly changes for time series volatile.

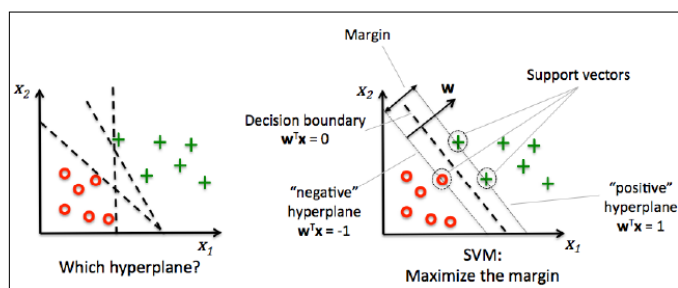


Figure 3. Support Vector Machine^[6]

B. K- nearest neighbor

K-nearest neighbor [33] is selecting number of k classifier among them nearest neighbor[7] classify. Assigning the labels for majority and get output value. A standardized range of all features guarantees equal contributions among predictors[9].

KNN for the prediction of the class it will analysis range 1 to n number of training dataset. Calculate the distance between the test data and train data based on different techniques to find the distance. Now it will divided distance value in the different class which is sort calculated distance. Now k value from the rows of the sorted array and most frequently class of these rows gives the range value. The forecast value set range based on the evaluation parameters gives the result.

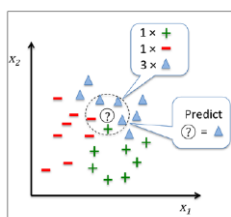


Figure 4. Example of k-nearest neighbor algorithm^[6]

C. Decision Tree

A decision tree is containing valid input and set of properties for an object we get output as Yes or No. Easy to interpret.

Two methods:

- Random forest – Independent classifier

Random forest is multiple decision trees and combine together for better accurate and prediction.[20]Random forest algorithm performs for large dataset[5]. It handles overfitting and reduce variance. Important Hyperparameters[20] are: Increasing predictive power and speed of model.

Features are selected randomly using bagging method. Based on dataset we will set featured, a number of training dataset are created by choosing random features with replacement. What this means is that one feature may be repeated in different training subsets at the same time. Once these features are selected, the trees are constructed based on the best split. Each tree gives an output which is considered as a 'vote' from that tree to the given output. The output which receives the maximum 'votes' is chosen by the random forest as the final output/result or in case of continuous variables, the average of all the outputs is considered as the final output.

- *Gradient boosting – Sequentially classifier*
It is sequential based classifier. Overfit can be possible and reduce bias and variance for dataset.

IV. PROPOSED MODEL

In the proposed model based on the historical data we can apply machine learning mechanism and compare with all mechanism. Targeted value for forecast stock market achieves accurately and minimal time. Selection of maximum features so investor gets an idea about the current trend, details analysis of company'. If a circuit breaker is happened then based on the closing price as well as maximum features to learn data and predict targeted value as an outcome. Investor will gain maximum profit with optimum time. More features are applied so easily predict with an accuracy of the stock market.

Four stage the proposed model is based on

1. Preprocessing
 - Collecting historical data and preprocessing data with methods are noisy data, missing value apply labels. Data will be fetched through datasets, libraries and real-time data. Further go to learning stage.
2. Learning
 - Apply machine learning mechanism for forecast whatever data we have to learn about data like open and close price of stock market at a particular time or day. eg are Open_high, Close_low
3. Evaluation
 - Selecting best features to forecast stock. Features are selected short term and long term period data or monthly or yearly data. According to data feature will be selected.
 - Different techniques through we can forecast learned data like filter methods, wrapper methods and embedded system.
4. Prediction:
 - Calculate performance like accuracy, hit rate, accurate time wise prediction , optimization and it will compare with all machine learning mechanism.
 - After comparison if actual value we get then further goes to different machine learning mechanism and targeted value get then evaluate result of prediction.

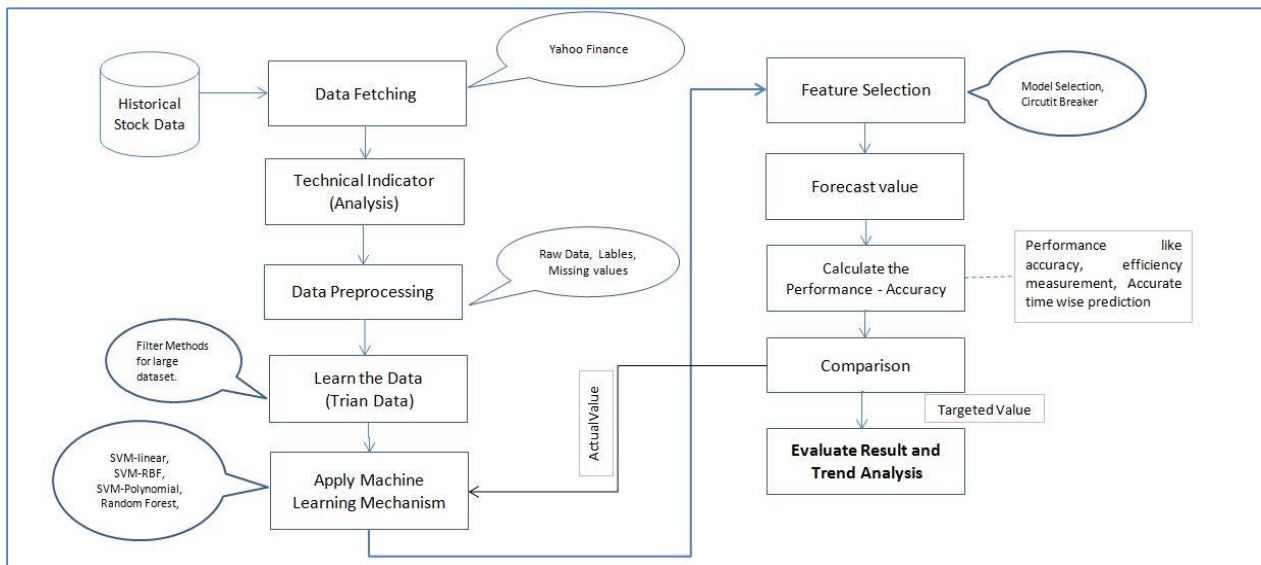


Figure 5 : Flowchart of Proposed Model

Steps:

1. Collecting historical data of stock from Yahoo finance.
2. Data fetching from libraries, dataset of the particular sector companies, short and long term data, different sector companies large dataset or required data will fetch.
3. Technical Indicator through analysis the historical data. In this phase we can easily compare the forecast value trend that is correct or not.
4. Data processing where missing value, normalization, raw data, set labels for the data. Then go to fetching phase.
5. Learn the data phase, data analysis based on feature selection.
6. Train the dataset ratio is 80-20%. Test data should be 20% of the entire dataset. Train data will be split for the forecast.
7. Applying machine learning mechanism for forecast targeted value in stock market. Values are not predictable. Based on historical data we can predict targeted value. Compare machine learning mechanism and give best result of forecast.

Machine Learning Mechanisms are:

- SVM- Linear
 - SVM-RBF
 - SVM-Polynomial
 - Random Forest
 - KNN
8. Feature selection is the best phase where hall proposed model is based on. Selecting maximum features gives best output.
 - Selecting features like Daily/Monthly/Yearly Data
 - Particular time wise selecting days, months, years
 - If Circuit Breaker is happens then close price of that day and applying features on historical data to forecast value.
 9. Forecast learned data with some methods to minimize data. Here filtering gives fastest time to forecast.
 10. Calculate performance like accuracy, optimum time to forecast, optimization, hit rate, time period wise prediction. It will compare with all mechanism.
 11. Comparison if targeted value gets it will go to step 10. If actual value gets then go to step 4 and hall steps will be follow.
 12. Targeted value get from step 9 then evaluate result in the end. Analysis future trend based on the forecast-set value.

V. SIMULATION AND EXPERIMENTAL RESULTS

Implemented proposed model in python language in Anaconda. Anaconda will perform on 4GB + memory, minimum 10 GB hard disk, Processor 1.8Ghz or higher, Intel 5th Generation processor in windows operating system. Python language for programming and libraries[32] used Pandas, Numpy, Scikit Learn, Scipy and Matplotlib.

V I Technical Indicators:

Technical Indicator is essentially a mathematical representation based on data sets such as price (high, low, open, close, etc.) or volume of a security to forecast price trends. There are several kinds of technical indicators that are used to analyze.

i. Simple Moving Average

A simple moving average (SMA) is an arithmetic moving average calculated by adding recent closing prices and then dividing that by the number of time periods in the calculation average.

ii. Relative Strength Index

It is a momentum oscillator that measures the speed and change of price movements. It oscillate between 0 and 100. Formula is given below and RS is relative strength.

$$RSI = 100 - \frac{100}{1 + RS}$$

iii. Bollinger Bands

These bands comprise of an upper Bollinger band and a lower Bollinger band and are placed two standard deviations above and below a moving average. Bollinger bands expand and contract based on the volatility.

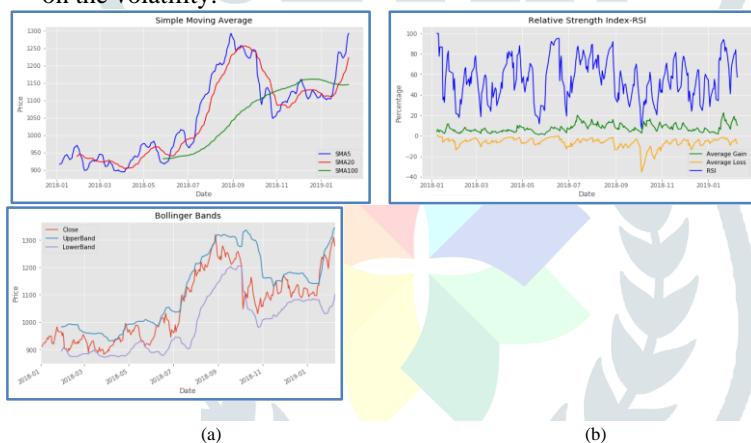


Figure 6: Technical Indicators: (a) Simple Moving Average (b) Relative Strength Index(RSI) and (c) Bollinger Band

V II Evaluation Parameters

Dataset is getting from Yahoo Finance Reliance Company 01/01/2018 to 08/02/2019. Dataset fetching from yahoo finance directly. Dataset train and test in 80-20% for the forecast value set.

Table 3 : Evaluation Parameters

No	Algorithm	Parameters	Forecast-Set (Values)	Accuracy (confidence)
1	SVM-Linear	kernel='linear', C=1e3	1208.3 to 1255.18	0.8544
2	SVM-RBF	kernel='rbf', C=1e3, gamma=0.2	1230.77 to 1273.23	0.8437
3	SVM-Polynomial	kernel='poly', C=1e3, degree=2	939.463 to 1170.61	0.0296
4	Random Forest	n- estimators=100	939.463 to 1170.61	0.2619
5	K-Nearest Neighbor	n- Neighbors =5	1218.02 to 1272.22	0.8779

As per above evolution parameter analyse that SVM-RBF and K-Nearest Neighbor gives highest accuracy in terms of accuracy. Comparing both the accuracy but range of the Forecast-set values will be different. The range of the forecast

set value based on the parameters. The graph trend of the KNN machine learning mechanism to analysis the future and current market.

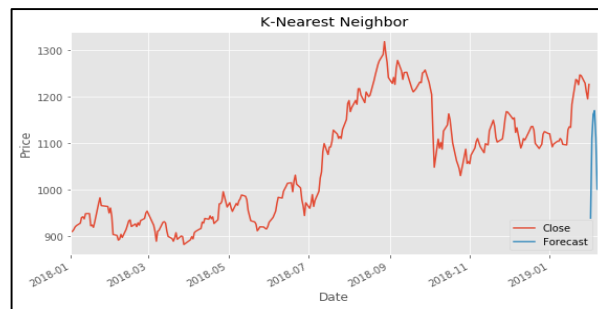


Figure 7. Graph trend K- Nearest Neighbor

Figure 10 Graph trend of K-Nearest Neighbor machine learning mechanism which shows the close price and forecast value set range prices of the RELIANCE company'.

VI. CONCLUSION AND FUTURE DIRECTION

Applying machine learning mechanism on the historical data with maximum analysis, comparison of various techniques and we will get targeted forecast set value of stocks market. Investor gets details analysis and accurate forecast price with optimum time. Try to apply different mechanism on different companies and compare with all of them to achieve the best forecast price in stock market. An investor makes the decision based on graph for analysis trend to gain profit. Further we can calculate performance for forecasting with maximum technical indicators. Investor make decision based on evaluating result to gain profit. Future direction also to apply machine learning mechanism on different sector company's to forecast stock market accurately with optimum time.

REFERENCES

- [1] Meryem Ouahilal et al. "A novel hybrid model based on Hodrick Prescott filter and support vector regression algorithm for optimizing stock market price prediction". Journal of Big Data(2017) DOI 10.1186/s40537-017-0092-5. Springer 2017
- [2] Bruno Miranda Henrique et al. "Stock price prediction using support vector regression on daily and up to the minute prices". The journal of Finance and Data Science4(2018).Pg no[183-201]. ScienceDirect 2018
- [3] Elham Ahmadi et al. "New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the Support Vector Machine and Heuristic Algorithms of Imperialist Competition and Genetic". Expert system with Application 94(2018) Pg. No[21-31]. ELSEVIER 2017.
- [4] Aparna Nayak et al. "Prediction Models for Indian Stock Market". Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016). Pg. No[441 – 449]. ScienceDirect 2016
- [5] Indu Kumar et al."A comparative study of supervised machine learning algorithms for stock market trend prediction". Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018)Pg no.[103-107]
- [6] Raschka, Sebastian. *Python Machine Learning*. [Packt] Publishing.
- [7] *Everything you wanted to know about STOCK MARKET INVESTING*. CNBC TV18, TV18 Mumbai.
- [8] Adam Atkins et al. "Financial news predicts stock market volatility better than close price". The Journal of Finance and Data Science 4 (2018) 120e137. Elsevier 2018
- [9] Tri Doan , Jugal Kalita "Predicting run time of classification algorithms using meta-learning". DOI 10.1007/s13042-016-0571-6 Springer 2016
- [10] Sameer Ranjan Jaiswal, Divyansh Sharma "Predicting Success of Bollywood Movies Using Machine Learning Techniques". <https://doi.org/10.1145/3140107.3140126>. ACM 2017.
- [11] Georgios N. Kouziokas "Machine Learning Technique in Time Series Prediction of Gross Domestic Product"
- [12] Man Li et al. "Stock Market Analysis Using Social Networks" <http://doi.org/10.1145/3167918.3167967>. ACM 2018
- [13] Osman Hegazy et al. " A Machine Learning Model for Stock Market Prediction" *International Journal of Computer Science and Telecommunications* [Volume 4, Issue 12, December 2013]
- [14] Alice Zheng, Jack Jin "Using AI to Make Predictions on Stock Market".
- [15] Vishal Parikh, Parth Shah "Stock Prediction and Automated Trading System". IJCSC Vol- 6 • Issue-1 Sep - Mar 2015 Pg no[104- 111]
- [16] Liheng Zhang et al."Stock Price Prediction via Discovering Multi-Frequency Trading Patterns". KDD 2017 Applied Data Science Paper. <https://doi.org/10.1145/3097983.3098117>. ACM 2017
- [17] <https://towardsdatascience.com/machine-learning-in-finance-why-what-how-d524a2357b56> [November 30th 2018]
- [18] https://en.wikipedia.org/wiki/Stock_market_prediction
- [19] Biao Huang et al. "Stock Prediction based on Bayesian-LSTM" . DOI: <https://doi.org/10.1145/3195106.3195170>.ACM 2018.
- [20] <https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd> [November 30th 2018]
- [21] <https://www.coursera.org/lecture/machine-learning/what-is-machine-learning-Ujm7v> Accessed on[November 30th 2018]

- [22] Mahek Usmani et al. " *Stock Market Prediction Using Machine Learning Techniques*". 2016 3rd International Conference on computer and Information Sciences(ICCOINS).
- [23] Yang Jiao et al. " *Predicting Stock Movement Direction with Machine Learning:an Extensive Study on S&P 500 Stocks*". 2017 IEEE International Conference on Big Data (BIGDATA) Pg no[4706-4713]
- [24] k- nearest neighbor <https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>
Access on[January 29th 2019]
- [25] Zsolt Bitavai, Trevor Cohn " *Day trading profit maximization with multi-task learning and technical analysis*".DOI 10.1007/s10994-014-5480-x Springer 2014
- [26] Shashank Tiwari et al. " *Stock Price Prediction Using Data Analytics*". 978-1-5386-1-17/\$31.00© IEEE2017
- [27] Abhinandan Gupta et al. " Stock Prediction using Functional Link Artificial Neural Network(FLANN)". DOI 10.1109/CINE.2017.25 IEEE 2017.
- [28] Li Tang et al." *K-Nearest Neighbor Regression with Principal Component Analysis for Financial Time Series Prediction*". <https://doi.org/10.1145/3194452.3194467> ACM 2018.
- [29] Tsai, C.-F. and Wang, S.-P "Stock Price Forecasting by Hybrid MachineLearning Techniques". Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I. ISBN: 978-988-17012-2-0
- [30] Li Tang et al. " *K-Nearest Neighbor Regression with Principal Component Analysis for Financial Time Series Prediction*" ACM 2018. <http://doi.org/10.1145/3194452.3194467>.
- [31] Mark Dunne. *Stock Market Prediction*. University College cork Computer Science Department
- [32] Capelo, Luis. *Beginning Application Development with Tensorflow and Keras*. [Packt] Publishing May 2018.
- [33] K- nearest neighbor <https://towardsdatascience.com/introduction-to-k-nearest-neighbors-3b534bb11d26>
Accessed on[3rd December 2018]
- [34] Kushal Patel at el. " Analysis of Stock Market Forecasting using Machine Learning Mechanism" JASC

Publication

December 2018.

