International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 01, 2020
ISSN: 1475-7192

Stock Market Analytics: Statistical and Machine Learning Techniques

J. Prathipa, S. Akshay and Rohan Shah

Abstract--- *Stock Prices tend to be erratic in behavior. They can be very volatile in nature, making it hard to predict. Thus, making an accurate analysis is beyond casual means. One method we use is to study historic data and learn patterns of uptrend and down- trend. Standard deviation is calculated on stock prices within a duration of quarter or a year under the close to close measure method. Many other statistical methods have been reviewed and analyzed. In this Project the efficiency of machine learning techniques including Random Forest (RF), Gradient Boosting (GB), Linear Regression (LR), and Decision Tree (DT) are proposed to be implemented and analyzed. This project aims to identify the most efficient Machine Learning Algorithm for consistent stock market analysis.*

Keywords--- *Learning Techniques, Statistical and Machine, Gradient Boosting (GB), Linear Regression (LR).*

I. INTRODUCTION

Stock market is the aggregation of buyers and sellers of stocks, which represent ownership links on businesses. The nature of the stock market is that of a dynamic sys- tem, it is influenced by a plethora of factors. Predicting stock market, considering factors such as market trend, social media analysis are the core challenging tasks to be considered.

Many business processes are automated to increase the efficiency and reduce risk involved. To achieve this, massive amounts of data is required to learn from case studies, this is where big data plays a vital role. The in- crease in complexity and data generation is transforming industrial operations along with the financial sector. Currently trillion bytes of data is produced in a spectrum of devices, which can be harnessed to be analyzed. Models can be created to provide solutions to our problems.

Many machine learning algorithms have been applied to the vast amounts of data to deduct patterns and establish models to help in predictive analysis, taking into consideration multiple variables.

Traditionally, every process of analysis was performed manually, including data collection, generation and the conclusions derived from them. This analysis was also done purely in a linear basis. But due to the non-linear nature of the stock market, the accuracy of the said pre- dictions was not efficient. The use of big data and algorithms can compute massive amounts of data and take in to factor many variables in a parallel manner. These models exhibit much better accuracy in their analysis.

II. LITERATURE SURVEY

Several methods have been implemented before to at- tempt at analysing stock values and its variations over a certain period.

J. Prathipa, Assistant Professor, CSE Dept., SRM IST, Kattankulathur. E-mail: prathipa.j@ktr.srmuniv.ac.in
S. Akshay, Student, CSE Dept., SRM IST, Kattankulathur. E-mail: sakshay_km@srmuniv.edu.in
Rohan Shah, Student, CSE Dept., SRM IST, Kattankulathur. E-mail: rohanshah_deepakshah@srmuniv.edu.in

DOI: 10.37200/IJPR/V24I1/PR200284
Received: 22 Nov 2019 | Revised: 26 Dec 2019 | Accepted: 10 Jan 2020

[4] A model for prediction was developed using the SAS suite of tools. The paper written by B.Siddhartha Reddy, the model emphasizes considerably on Stock Market (NSENifty, India) and its prediction along with the risk involved. There are many models such as Interrupted Time Series, Auto Regression (AR), Moving Average (MA), and Distributed Lags Analysis. Other types of statistical procedures namely, FORE CAST, ARIMA is also reviewed.

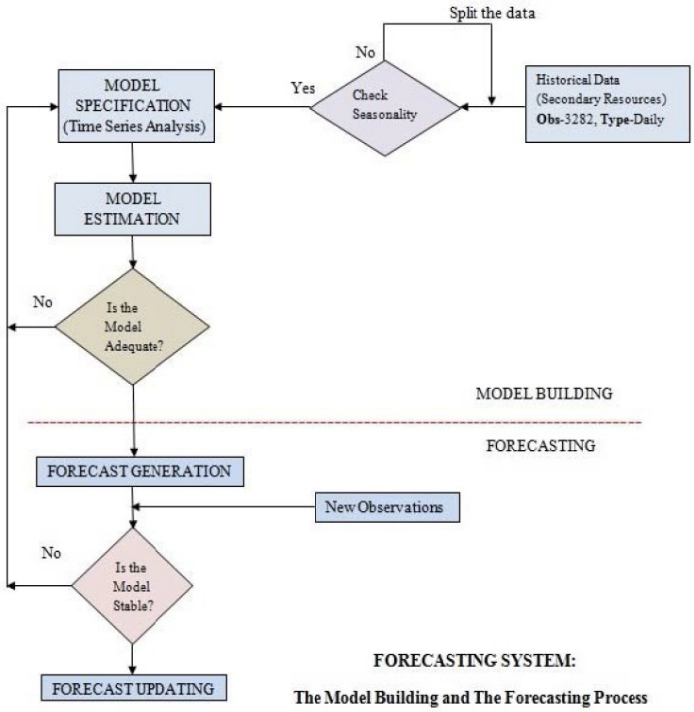


Figure 1: The Model Building Prediction of Stock Mar- ket Indices – Using SAS, figure 3.[4]

[5] A similar study was documented in which the risk factor in the stock market values was analyzed. Data from BSE and NSE was used for the purpose of analysis. The study revealed that research on values using time series data using algorithms ARIMA and AR have proven to be useful and efficient.

Beta calculation of risk analysis.

Closing price (bank) = rp

Current closing price = ra

β = Covar (ra, rp) /Var (rp)

If β < 1 implies that the stock has low risk, Otherwise

Stock has high risk

[3]Ryota Kato and Tomoharu Nagao of the Yokohama

National University, Japan, proposed in their paper in

2012 a stock market prediction model, which involved the use of interrelated time series data.

DOI: 10.37200/IJPR/V24I1/PR200284
Received: 22 Nov 2019 | Revised: 26 Dec 2019 | Accepted: 10 Jan 2020

They focus on extracting interrelationships between the predicted stock and various time series data, for example different stocks, world stock market indices, foreign exchanges and oil prices. And test the method for predicting the daily up and down changes in the closing value by using discovered interrelationships.

While there are some causes of changes in stock prices, information about other than the predicted stock should have effects on the predicted stock. For example, a stock related to exports is affected by foreign exchanges or foreign stocks. Therefore they extract interrelations of changes in stock prices between the predicted stock and various time series data from real data, then predict the stock by using extracted interrelationships. This method is composed of two phases, interrelation discovery phase and prediction phase.

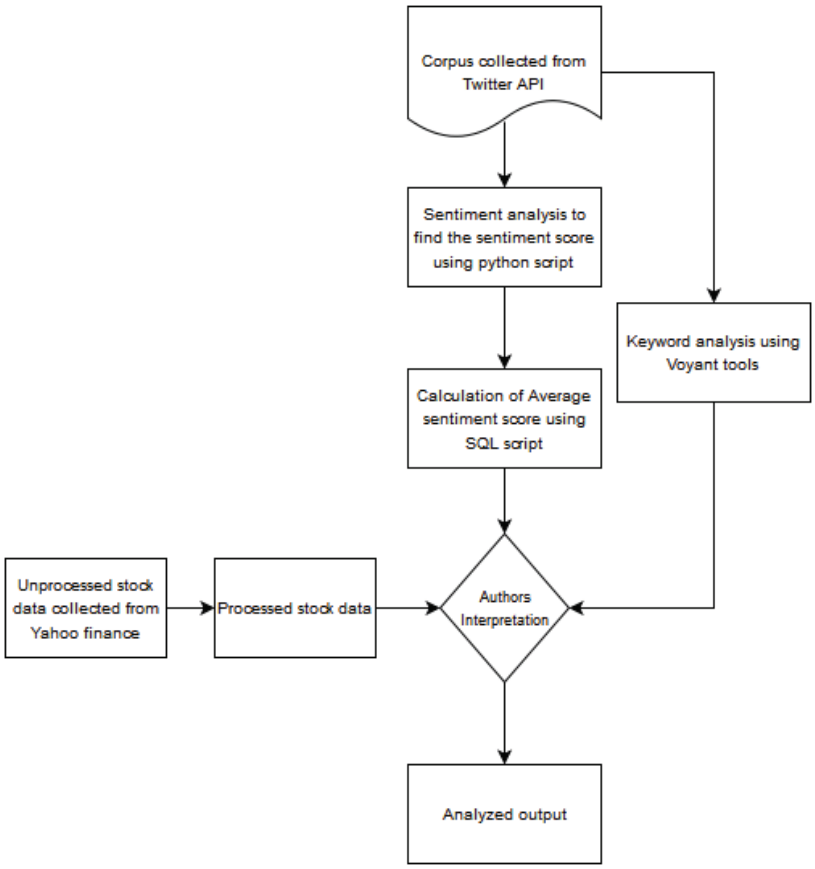


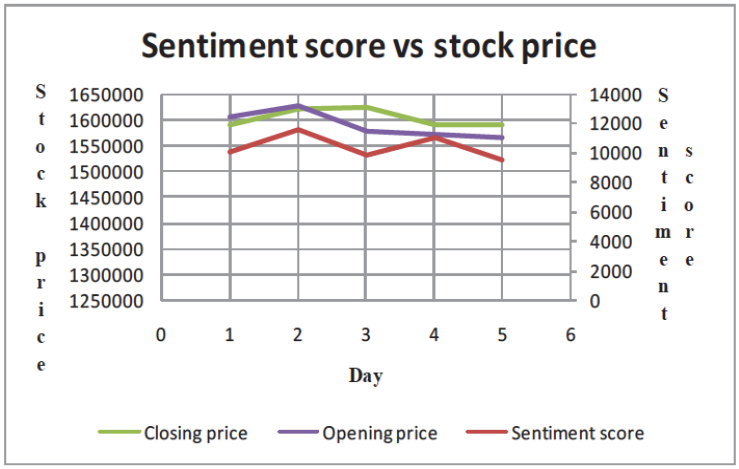
Figure 2: Analyzing stock price changes using event related Twitter Feeds, Figure 1.[7]

Later, [7] machine learning algorithms were starting to be implemented in the analysis of stock values. A paper published in Feb'17 explains the analysis of stock price changes using event related Twitter Feeds. This paper describes about the stock prices changes over the event related day wise tweet sentiment score. The authors have analyzed the stock market values collected from yahoo finance. They have done the keyword analysis of tweets to find the frequent keywords used during the period about the event and also seen the trend of keyword used over the period. After all the analysis and observation some positive correlation was found between the stock prices changes and the average sentiment score.

The sentiment analysis involves assigning scores to each related tweet in the range of -1 to +1.

DOI: 10.37200/IJPR/V24I1/PR200284
Received: 22 Nov 2019 | Revised: 26 Dec 2019 | Accepted: 10 Jan 2020

Table 1: Analyzing stock price changes using event related Twitter Feeds, Figure 5.[7]



[1]A paper written by R. Yamini Nivetha and Dr. C. Dhaya illustrates a comparative study of three algorithms used to analyse stock market values namely, Multiple Linear Regression, Support Vector Machine and Artificial Neural Network.

The prediction model predicts daily prediction and monthly prediction. Sentiment Analysis is combined with a best predicting algorithm to refine the outcomes in the stock market. Hence the prediction model acts as a stock market broker in finance and business streams.

Stock markets generate a lot of data provided by various sources including the stock exchange, economic reports, and company financial reports, region-specific financial information.

[9] In this paper, The authors , Poonam Somani, Shreyas Talele, Suraj Sawant, aim at surveying recent literature in the area of Neural Network, Hidden Markov Model and Support Vector Machine used to predict the stock market fluctuation. Neural networks and SVM are identified to be the leading machine learning techniques in stock market prediction area. Also, a model for predicting stock market using HMM is presented. Traditional techniques lack in covering stock price fluctuations and so new approaches have been developed for analysis of stock price variations. Markov Model is one such recent approach promising better results. A predicting method using Hid- den Markov Model is proposed to provide better accuracy and a comparison of the existing techniques is also done.

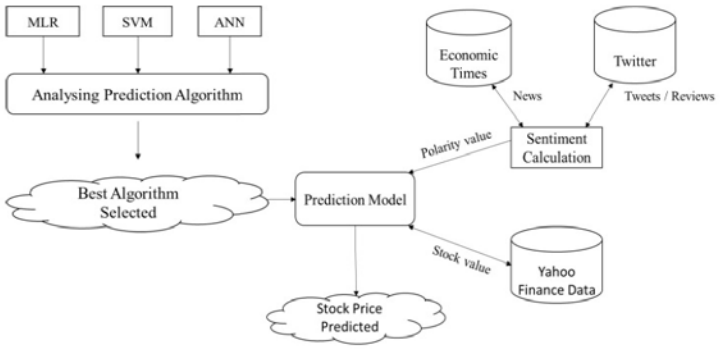


Figure 3: Stock Prediction Model, Developing a Prediction Model for Stock Analysis, Figure 1. [1]

DOI: 10.37200/IJPR/V24I1/PR200284
Received: 22 Nov 2019 | Revised: 26 Dec 2019 | Accepted: 10 Jan 2020

A model of HMM is developed by training data accord- ing to the Baum Welch algorithm. The obtained model is then tested using a Maximum a posteriori(MAP) ap- proach. A best probability value is chosen from the mod- el using the Viterbi algorithm. This output represents the closing stock value of the following day.

[2] In this paper, Rui Li, DianZheng Fu, Zeyu Zheng , study the impact of social media opinions on the stock market. They gathered data from the forums of the Shanghai composite index and composed structured emotion time series data using sentiment analysis and feature extraction. A study between the emotional index and the change in stock prices is also illustrated.

It was evident after the analysis that stock market values were impacted considerably based on the stock emotional index. Therefore, they concluded that emotional index is a viable factor in predicting the nature of the stock market values.

[6] This paper written by Kamran Raza,Mehak Usmani, Syed Hasan Adil, Syed Saad Azhar Ali, illustrates the aim to predict the market performance of Karachi Stock Exchange (KSE) on closing values by the methods of machine learning algorithms such as Multi-layer Percep- tron and Support vector machine. They have used various variables such as foreign exchange rate, the news, social media feeds in addition to the stock market values.

In conclusion, they inferred that MLP algorithm was proven to be most efficient and that Foreign exchange rate wasn't of considerable factor in the analysis of the change in stock values.

III. CONCLUSION

In this paper we have proposed to apply statistical meth- ods and machine learning algorithms to analyse stock market values.

It is observed that machine learning algorithms generally produce more accurate results in the analysis of stock market values. The machine learning algorithms proposed to be implemented: Random forest, gradient boost- ing, Linear Support Vector Machine. A visual representation of the implementation of above stated algorithms will conclude a higher efficiency of stock values analysis compared to statistical methods.

REFERENCES

[1] Dr. C. Dhaya, R. Yamini Nivetha, "Developing a Pre- diction Model for Stock Analysis", 2017 *International Conference on Technical Advancements in Computers and Communications.*

[2] DianZheng Fu, Rui Li, Zeyu Zheng ," An analysis of the Correlation between Internet Public Opinion and Stock Market",

[3] Ryota Kato, Tomoharu Nagao , "Stock Market Predic- tion Based on Interrelated Time Series Data",

[4] B.Siddhartha Reddy," Prediction of Stock Market Indices – Using SAS", *IEEE.*

[5] B. Uma Devi, D. Sundar, Dr. P. Alli ," A Study on Stock market Analysis for Stock Selection - Naive In- vestor's Perspective using Data Mining Technique", *International Journal of Computer Applications*, 2011.

[6] Kamran Raza,Mehak Usmani, Syed Hasan Adil, Syed Saad Azhar Ali, " Stock Market Prediction Using Ma- chine Learning Techniques", 2016 *3rd International Conference On Computer And Information Sciences (ICCOINS)*

[7] Bijay Bhaskar De, Hee-Cheol Kim, Mangal Sain, Satyabrata Aich, "Analyzing stock price changes using event related Twitter feeds", February 19 ~ 22, 2017

DOI: 10.37200/IJPR/V24I1/PR200284
Received: 22 Nov 2019 | Revised: 26 Dec 2019 | Accepted: 10 Jan 2020

[8] Dai Feng, Ma Ruobing, Wu Songtao," Market De- mand Risk Analysis on the Decision of Enterprises' Op- timum Stock" 2008 IEEE DOI 10.1109/IITA.Workshops.2008.127

[9] Poonam Somani, Shreyas Talele, Suraj Sawant, " Stock Market Prediction using Hidden Markov Model", 978-1-4799-4419-4 /14/ ©2014 IEEE

[10] 10.Yaojun Wang, Yaoqing Wang," Using Social Media Technology to Assist in Price Prediction of Stock Mar- ket", 10.1109/ICBDA. 2016.7509794

DOI: 10.37200/IJPR/V24I1/PR200284
Received: 22 Nov 2019 | Revised: 26 Dec 2019 | Accepted: 10 Jan 2020