Stock Market Prediction: Using Historical Data Analysis

Vivek Kanade	**Bhausaheb Devikar**	**Sayali Phadatare**	**Pranali Munde**	**Shubhangi Sonone**
I.T. DEPT	I.T. DEPT	I.T. DEPT	I.T. DEPT	HOD. I.T. DEPT,
India	India	India	India	India

Abstract— The Stock market process is full of uncertainty and it's affected by many factors. Hence the Stock market prediction is one of the important exertions in business and finance. There are two types of analysis possible for prediction, fundamental and technical. In this paper both fundamental and technical analysis are considered. fundamental analysis is done using social media data by applying sentiment analysis process. Social media data has high impact today than ever, it can helpful in predicting the trend of the stock market and Technical analysis is done using historical data of stock prices by applying machine learning algorithms. The method involves collecting news and also collect social media data and extracting sentiments expressed by individual. Then the correlation between the sentiments and stock values is analysed. The learned model can then be used to make future predictions about stock value. It's can be shown that this method is able to predict the stock performance and sentiment and social data are also closely correlated with recent news.

Keywords— prediction, Big data, social media analytics, machine learning.

I. INTRODUCTION

With growing popularity of Big Data Analysis, finding new patterns and meanings out of data which we were unable to do so far because of the available data silos, has now become a very powerful weapon for prediction in different sectors on the planet. Understanding the insights and the power to make quick decisions has geared us up from a level where we can't answer a question to a level where we literally can answer it that too within a fraction of a second. Stock Market continues to serves as a medium for companies to add up to their capital by introducing their company shares to the market and also turns out to be a handy platform for investors to earn beyond the threshold of the interest rates that the Banks offer. Considering the risk involved in stock market trading resulting out of the volatility which is influenced by a number of factor across the globe, predicting the behaviour and short term or long term potential of a company's stocks has been one area of interest of several Data Analyst from a long time now. Different algorithms have been designed by using different sciences of learning but have failed to provide an accurate prediction of any stock movement. In this paper we propose to build a system that will work on a similarity comparison model and will predict the stock prices by mapping it to its past behaviour under similar potential conditions.

II. OBJECTIVES

Problem Statement 1: Predictions are based on absolute theories and behaviours.

According to the Firm Foundation theory the market is defined from the reaction of the investors, which is triggered by information that is related to the "real value" of firms. The "real value" or else the intrinsic value is determined by careful analysis of present conditions and future prospects of a firm. On the other hand, according to the Castles in the Air theory the investors are triggered by information that is related to other investor's behaviour. So for this theory the only concern the investor should have is to buy today with the price of 20 and sell tomorrow with the price of 30, no matter what the intrinsic value of the firm he (or she) invests in is. Therefore, the Firm Foundation theory favour the view that the market is defined mostly by logic, while the Castles in the Air theory supports that the market is defined mostly by psychology.

Still there are some aspects which affect the accuracy of the predictions as they have some serious impact on the Stock Prices like

1. Global Indices
2. Indian Indices
3. Currency Prices
4. Sector Behaviour
5. Market Movers
6. News
   a. Economic Policies   b. Political News   c. Natural Disasters

Problem Statement 2: Unavailability of Prediction Tracking.

None of the applications provide regular predictions for the Stock Prices. It's if you are planning to invest on Stock "xyz" toady, then there is no application can predict its behaviour at that instance.

Problem Statement 3: No Guarantee on the accuracy of the Predictions

There is no application available today that can provide us a guarantee on their predictions or at least ensure the percent accuracy that their system can provide.

III. MOTIVATION

The Firm Foundation theory defines the Stock Exchange as the reaction to a Company's Real Value by the Investors. Analysis of Present and Future prospects of a Company helps in deriving its Real Value. Whereas, the Castles in the Air theory defines the Stock Exchange as the reaction of investors to actions of other investors. Hence this theory pitches the only concern for an investor as to purchase a Stock at let's say INR 60/- today and sell the same at INR 70/- tomorrow without being concerned about the Real Value of the Company in which he/she invests. Study leads us to conclude that either the Stock Market is influenced by pure logic (based on Firm Foundation Theory) or it is purely influenced by psychology.

A combination of both theories may prove out to be one of the efficient ways to predict the behaviour of the market. This may focus on following factor that have a role in influencing the behaviour of the Stock Prices.

7. Global Indices
8. Indian Indices
9. Currency Prices
10. Sector Behaviour
11. Market Movers
12. General News
13. Company News
14. Company Financial Reports

In this paper we propose a Stock Market prediction algorithm which will be combining both the absolute behaviour as well as the relative behaviour of a company stock with analytical focus on the past behaviour. In this paper we will be implementing ways to identify the current situation of the Stock and map it up with a similar situation in the past.

IV.   LITERATURE SURVEY

The social media data, financial news and analysis of results are presented in a facilitated manner. By using the capabilities of Big Data a prediction model has been built, the trend of stock markets are predicted periodically by analytics of social media and machine learning.

Technical analysis methods such as regression analysis which proves social data complements of sentiment analysis of a model.

The volatility of the markets and the future performance of the system is changed by the political news, economical news and influence of the social media.

The numeric data adds Social media data which exploits the quality of the input and gives improved predictions.

The predictions at Real time with the aide of Big Data Technology.

The algorithm which is used for sentiment analysis that uses summative assessment of the sentiments in a particular news article or tweet, which can be improved for better calculation of sentiment, which would improve the accuracy of the prediction. [1]

The scaling law for price impact on the LSE market for a sample of six liquid equities.

For Smaller trades volume the price impact is decreasing, and for the larger trades there is a growing power-law dependence which is consistent with the findings for the US and the Australian stocks in previous studies.

For the stocks the price impact function is time-varying and the trading day advances the price impact is decreasing. [2]

The stock exchange data processing and presentation for Intelligent information system.

The results which are obtained that are expert system based on artificial neural networks combined with fractal and technical analysis is able to analyse and identify companies that will bring profits with 60% correctness (which means 60% of correct investment decisions) what in the case of tested period of time resulted in 11.1% of profits. The situation on the Warsaw Stock Exchange which was complicated because of political and economic situation in Europe in which the results which were obtained were better than before . [3]

The problem of analysing the mining text data to predict the stock market.

The authors proposed unified a model of latent factor to model the joint correlation between newspaper content and stock prices, which allowed them to make predictions on individual stocks, even which that did not appear in the news.

Authors solved the problem of factorization problem using AMM using sparse matrix from the formulated learning model.

Back testing which was performed extensively over these six years of Stock price data and WSJ showed the methods performed better than the market and a number of portfolio building strategies.

The methodology is applicable to text data of all sources and extended it to higher frequency data sources as Twitter. [4]

The series of stock market is generally non-parametric, noisy, dynamic and chaotic by nature.

Soft computing techniques are used widely, support Vector Machine(SVM) has gained lot of popularity and its has been outstanding against Artificial Neural Network(ANN) also.

SVM is unique and performs better in many applications, the suitable parameters of SVM (C, 0 and c) is the problem which is impacted by the practicality of SVM.

The Cuckoo Search (CS) is based on the Swarm Intelligence optimization technique and is very simple to tune the parameters of SVM. The accuracy of prediction ANN SVM and CS-SVM for Indian stock market was analysed.

The results which were obtained experimentally indicates the CS-SVM method can achieve higher accuracy rate than regular SVM method.

The accuracy of results improved of CS-SVM from 80.23% to 82.21% for BSE-Sensex and from 81.56% to 83.22% for CNX-Nifty.

The results of the proposed experiment demonstrates that the CS-SVM provides higher accuracy with lower MSE and MAPE among ANN and SVM.

The scope for improving the model by incorporating the social media sentiments and analysing the impact of various factors such as gold price, crude oil price and dollar price etc., towards the stock market index. [5]

The real task is to predict the financial market values and their impact.

The values are result of the attitude of investors and domain level experts it opens up an opportunity for researchers in this field by performing sentiment analysis over e-data depicting thoughts and views of these investors and experts.

The great opportunity is achieved by Electronic media for people to share thoughts and views, judgments and observations.

The trading dealt by online means is encouraged by blogs, forums and website.

This e-data if analysed in an efficient manner which helps the financial experts to predict the index values more accurately.

The authors tired to collect the e-data form various sites and after classification tried to analyse it with ANN.

The results has opened many opportunities for research and exploiting media for trade at the end.

The authors have tried to reduce the error in its prediction up to few – possible value whereas e-data can prove to be much more efficient and utilized in an optimal manner.

The authors have tried to set the target on index value of BSE which can be extended to other indices. [6]

V. PROPOSED SOLUTION

❖ *Design*
  o A Web Application will be designed to give a friendly interface to the users for the following
❖ *Market Watch*
  o The customers can have a glimpse of what's happening on the Stock Exchange on a Real Time basis as also during the non-working hours. This section of the application will provide the users with a stock ticker displaying the current price and change in the price of few stocks on a random basis. This ticker can be configured to display stock movements of companies the user would be interested in tracking. A detailed view of the movements in the Indian and Global Indices along with Commodity Watch would be another feature of this section. This data will be fetched from the Yahoo Finance API using Web Services and API calls.
❖ *News*
  o The live news feeds will provide news in different categories like financials, Government Policies, Natural Calamities, Disasters, FII, Foreign Policy, Global Indices and Sector News. A near real time view of headlines from Yahoo News and RSS Feeds will be constantly supplied in this area. Every news can be read in detail if it is of interest to the user. The news will also have mentions about the sentiments of Social Media on topics relevant to the news and possible impact on the Stock Exchange. If the user is registered, he will also be able to see classification of every news as a positive or negative sentiment.
❖ *Stock Search*
  o This section will enable the user to find the Stock Code of any company and get the detailed Quote for any Stock code. The quote will consists of details including but not limited to Day Open, Day High, Day Low, Previous Day Close, Top 5 Bids and Top Five Offers. If the user is registered, he will also be able to see the predictions for the stock (details in the next segment).
❖ *Portfolio Management*
  o This section will be available only for registered users. Here a user can add, modify and delete the details of his transactions. He will be able to see the detailed prices, new and latest predictions on the stocks that he has in the portfolio. Moreover based on the purchase price, current price or selling price the user will also be updated by the profit/loss or each equity he has transacted and also the total profit/loss that he has incurred on his portfolio. The user will also be advised with some helpful tips to improve his portfolio. The user can set alerts on the stock prices for him to get intimated for further actions.
❖ *Stock Predictions*
  o This section will display the detailed report of the performance analysis of any stock. Based on certain assumptions, the future stock value for a period of over a week will be portrayed in here with interactive visual graphs that will make understanding the assumptions much easier. Moreover, this section will also have the accuracy meter which will reflect the margin of error that we hold in the prediction of the stock behaviour. The margin for error will vary from stocks depending on the past performance of our respective predictions. The predictions made would be depended on certain prevailing conditions like the Sector Behaviour, Indian Markets Behaviour, Global Market Behaviours and New. A detailed report of the expected stock behaviour in different

conditions can be found in this section. Moreover this section will be on a continuous update and will change depending on the volatility in the pertaining conditions.

VI. PROPOSED ARCHITECTURE
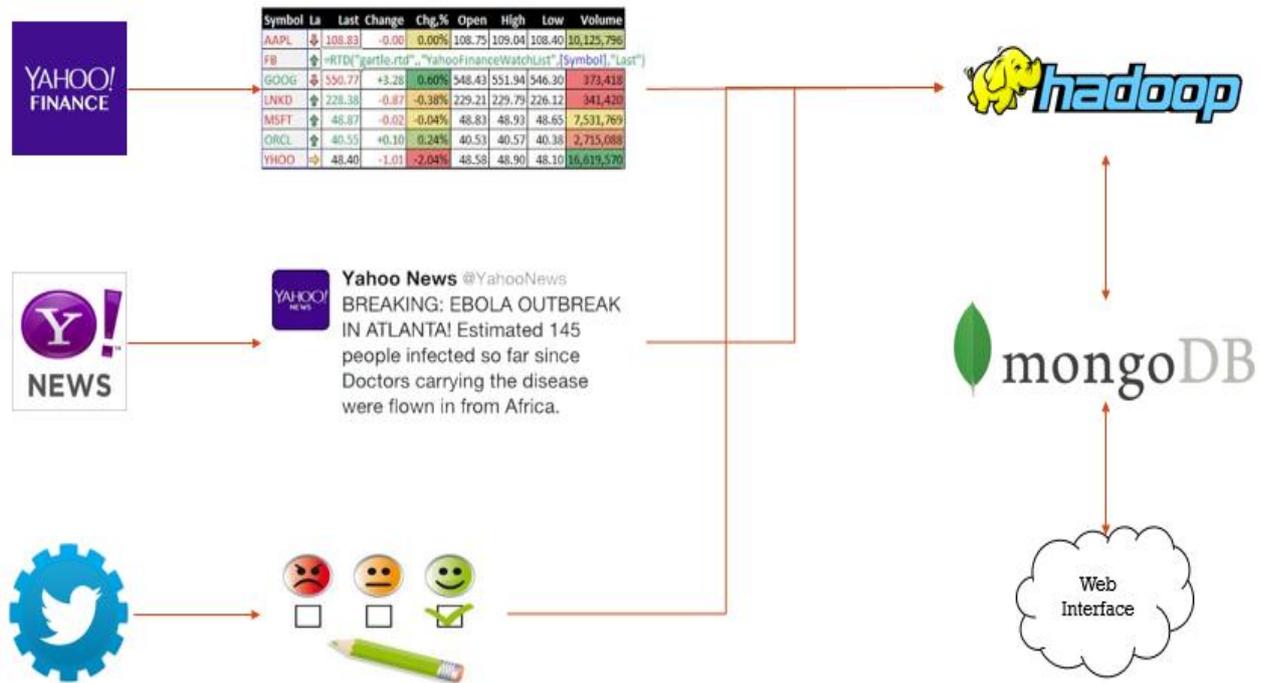
VII. CONCLUSION

In this paper developed facilities techniques and for exploiting especially social media data, financial news and analysis results are presented. A prediction model has been built that uses social media, big data analytical capabilities analytics and machine learning to periodically predict the trend about stock markets. Model shows that sentiment analysis of the social data complements proven to the technical analysis methods such as regression analysis. It's shows that the future performance and volatility of the markets of the system is affected by the political and economic news and influence of the social media. Exploiting social media data in addition to numeric data increases the gives improved predictions and quality of the input . The aide of big data technology allows predictions for real-time. However the algorithm used for sentiment analysis uses summative assessment of the sentiments in a particular tweet or news article, this could be improved for better sentiment calculations, which would improves for the accuracy of the prediction.

REFERENCES
[1]   Girija V Attigeri , Manohara Pai M M, Radhika M Pai, Aparna Nayak "Stock Market Prediction: A Big Data Approach", 978-1-4799-8641-5/15/$31.00 © 2015 IEEE.
[2]   Mateusz Wilinski, Wei Cui, Anthony Brabazon, "An Analysis of Price Impact Functions of Individual Trades on the London Stock Exchange".
[3]   Micheal Paluch Lidia Jackowska-Strumillo "Intelligent information system for stock exchange data processing and presentation", 978-1-4673-6936-7/15/$31.00 ©2015 IEEE.
[4]   Felix Ming Fai, Zhenming Liu, Mung Chiang "Stock Market Prediction from WSJ: Text Mining via Sparse Matrix Factorization", 2014 IEEE International Conference on Data Mining.
[5]   Ms.K.Nirmala Devi, Dr.V.Murali Bhaskaran, G.Prem Kumar, "Cuckoo Optimized SVM for Stock Marl(et Prediction", IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded and Communication systems (ICJJECS)2015.
[6]   Sunil Kumar Khatri, Himanshu Singhal, Prashant Johri "Sentiment Analysis to Predict Bombay Stock Exchange Using Artificial Neural Network", 978-1-4799-6896-1/14/$31.00 ©2014 IEEE .
[7]   Yahoo live news API.