Stock Prediction Using Technical and Sentimental Analysis

Prof. S. U. Mali[1], Ajinkya Admane[2], Pranav Dhanawade[3], Satyam Wagh[4], Harshad Gandhale[5]

Assistant Professor, Department of Computer Engineering, NBN Sinhgad School of Engineering, Pune[1]
Students, Department of Computer Engineering, NBN Sinhgad School of Engineering, Pune[2,3,4,5]

Abstract: *Stock market movement prediction is a well-known problem of interest. These days, social media accurately reflects public attitude and opinion on current events. Researchers have focused their attention on Twitter in particular to examine popular sentiments. Stock market forecasting using public emotion conveyed on Twitter has been a fascinating area of study. The approach employed in sentimental analysis is to look at how strongly stock price changes, such as gains and decreases, are associated to people's Twitter opinions. Sentimental analysis aids in the examination of public opinion on Twitter; we took this approach by employing sentimental analysis. Another approach in the same topic of our project is using technical analysis. We use a multiple kernel learning regression framework to model the stock price movement as a function of these input features and solve it as a regression problem. The machine learning coupled with fundamental and/ or technical analysis also yields satisfactory results for stock market prediction. We also evaluated the model for taking buy-sell decision at the end of day which is also known as intraday trading.*

Keywords: Sentimental Analysis, Regression, Machine Learning, Prediction, etc.

I. INTRODUCTION

Background

Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations. This concept came into existence since the innovation of social networking websites such as Twitter, Facebook, Tinder etc. and people started posting their honest online reviews about any particular brand or product on the Internet. Stock Prediction is one such field which is borne out of this. Sentimental Analysis encompasses several domains such as Natural Language Processing, text analysis, computational linguistics, and biometrics. Currently many traders and brokers use indicators based on mathematical formulae. Many of them even don't know how they work. Technical indicators will be in use if market or company stock repeats pattern; but many times, it doesn't. Through this project we have attempted to overcome most of the above flaws, make the system cheaper and easier to use and understand.

Relevance

Stock market prediction task is a fascinating topic and it divides researchers and academics into two groups, people who believe we'll devise mechanisms to predict the market and people who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, hence there's no space for prediction. Stock price prediction can be used to gain insight about market behavior over time, spotting trends that would otherwise not have been noticed.

Machine learning will be an efficient technique to handle this problem as computer computing power increases. However, the public stock data-sets is too limited for any machine learning algorithm alone to work with, while asking for more features may cost thousands of dollars every day.

In our project we will perform sentimental analysis of Twitter data (news or comments) to gain insight on customer behavior. It will also help us analyze the public sentiments. The researchers found a major correlation between stock returns and individual's reactions. In fact, valuable data in the domain of stock market should include several features like time, targeted audience, and brand but the most important feature for the decision makers who are looking to invest in the stock market are time and brand.

II. LITERATURE REVIEW

Several researchers have studied on various types of techniques for Concept of sentimental analysis is described and a model is proposed along with some ML methods. Sayavong Lounnapha[2], proposes a stock price prediction model based on convolution neural network, which has obvious self-adaptability and self-learning ability. Combining the characteristics of CNN (Convolution Neural Network) and Thai stock market, the data set is trained and tested after pretreatment. Three Thai Stock Exchange stocks (BBL, CAPLL&PTT) are examined and compared to the actual stock price on this basis. The results suggest that the CNN-based model can efficiently identify and predict changing stock price trends, providing a useful reference for stock price forecasting. In 2016, researchers used genetic algorithm and simulated annealing to find initial weights of an ANN, and then back-propagation algorithm is used to train the network [7]. In terms of Nikkei 225 index return prediction, this hybrid technique outperformed typical ANN-based algorithms.

The genetic algorithm was effectively used to obtain optimum weights of an ANN in which technical indicators were used to predict the direction of Nikkei 225 index movement with minor modifications [8]. Deep ANNs, that are basically neural networks with more than one hidden layer, among the first deep methods used in the domain. In 2016, some researchers [9] predicted NASDAQ prices based on the historical price of four and nine days ago. The number of hidden layers and neurons in ANNs with various topologies, including both deep and shallow ones, was investigated in order to determine the suitable number of hidden layers and neurons within them. Deep ANNs were found to be superior to shallow ANNs in the studies. Ze Zhang [3] proposes a method for optimising network weights and thresholds using a self-adapting version PSO algorithm.

Afterwards, the optimized data, regarded as initial weight and threshold value, is given to Elman network for training, accordingly the prediction model for opening price of stock market based on self-adapting variant PSO-Elman network is formed. Finally, they verify that model by some stock prices, and compares with BP network and Elman network, so as to draw the result that shows the precision and stability of this predication model both are superior to the traditional neural network. Heng Liu [4] introduce a deep residual network (ResNet) model for prediction, using the stock price graph as input.

The results show that the ResNet model has the average accuracy of 0.40, which is higher than the stochastic indicator of 0.33. Recently more and more researchers have applied deep learning models to stock prediction problems, and many studies have shown that deep learning models have strong learning ability and prediction accuracy [11-13]. In 2015, He et al. [6] proposed a special convolutional neural network called deep residual neural network, which can increase the number of network layers to hundreds or even thousands easily, while training time is within the acceptable range, greatly improving the accuracy of image recognition.
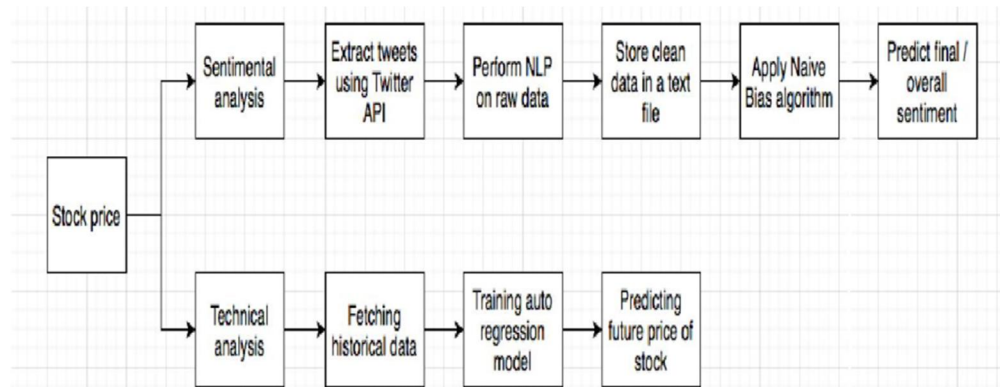
III. PROPOSED SYSTEM

Figure: Proposed System Architecture

IV. METHODOLOGY

Natural Language Processing

Natural language processing (NLP) is a subject of linguistics, computing, information engineering, and AI concerned with how computers interact with human (natural) languages, particularly how to design computers to handle and analyse huge volumes of lingual data. Challenges in tongue processing frequently involve speech recognition, tongue understanding, and tongue generation.

Naive Bayes Algorithm

Naive Bayes classifier uses applied mathematics to classify data. Naive Bayes classifier algorithms make use of Bayes' theorem. The basic insight of Bayes' theorem is that as new data is introduced, the probability of an event can be changed. What makes a naive Bayes classifier naive is its assumption that each one attributes of a knowledge point into account are independent of every other. Apples are red, spherical, and a specific size, but a classifier sorting fruits into apples and oranges would not assume all of these things at once. After all, oranges are round as well.

Time Series Prediction

Time series analysis refers to a set of techniques for evaluating statistical data in order to derive useful statistics and other information features. When a model is used to predict future values, it is said to be time series forecasting. While multivariate analysis is usually employed in such how on test theories that the present values of 1 or more independent statistic affect the present value of once more series, this sort of study of your time series isn't called "time series analysis", which focuses on comparing values of one statistic or multiple dependent time series at different points in time. Interrupted statistical analysis is that the analysis of interventions on one statistic.

Auto Regression

Auto regression is basically a time series model that uses observations from previous time steps or from the past data as input to a regression of y on x to predict the worth at the following time step. It is a simple concept that may end up in accurate forecasts on a spread of your time series problems which is a great way to predict the future price of a company's stock.

ARIMA from Stats Model

Auto regressive integrated moving average (ARIMA) is statistical model for time series prediction. The standard notation for the model is ARIMA (p, d, q) where p is total lag observations also called lag order, d is number of times that the raw observations are differenced also called degree of differencing, and q is size of moving average window. Model predicts future state by checking difference between values in time series instead of actual stock value.

Forward Moving Average

It is the most used technical indicator used to find momentum of trading stock. It is calculated for fixed time frame by adding next successive items to previously calculated average of group of items. The most common time frame for predicting stock momentum is 14 days, 22 days and 50 days. It is used to understand short term fluctuations. Simple moving average (SMA) and exponential moving average (EMA) are two basic types of moving forward average.

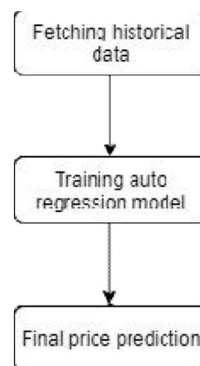## V. STAGES OF IMPLEMENTATION

Technical Analysis

Figure: Steps for technical analysis
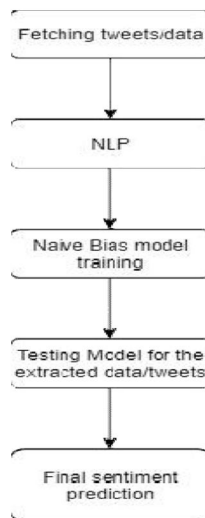
Sentimental Analysis

Figure: Steps for sentimental analysis

VI. IMPLEMENTATION TECHNIQUES

6.1 Raw Data Preprocessing Component
It takes content of news and comments as input, and also takes historic price and volume data as raw data for further process the content of tweets require cleaning which helps us in creating the sentiment analysis.

6.2. Feature Extraction Component
It breaks down each and every word in the fetched tweets and compares it with bag of words, if that word exists in bag of words, then it assigns it labels from that respective bag of words.

6.3 Sentiment Analysis Component
It analyzes the overall sentiments (objective, subjective, positive, negative), a threshold value for overall sentiments is set to classify each sentiment. The output of process will be the actual percentage of positive and negative sentiment from the fetched data.

6.4 Technical Analysis Component
It formulates different indicators based on price and volume. These indicators generate signals movement of market. Then we combine outputs of each signal using simple combination or using Neural Network combination. The trend of market is predicted as output of technical analysis.

VII. PROBLEM STATEMENT
We aim to create a system for stock prediction based on a set of technical trading rules and sentimental analysis of date present on social media. The aim of the research is to check if it is possible to obtain a set of trading pattern, which could be used to take decisions while trading such as a Buy, Sell, Exit, Stop Loss, etc.

VIII. CONCLUSION
The proposed system will help beginner traders as a decision support tool and help them take decisions accordingly. Collective analysis of news regarding market have been made easy using machine learning algorithms

REFERENCES
[1] Pahwa, K., & Agarwal, N. (2019). Stock Market Analysis using Supervised Machine Learning. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing.

[2] Akba, F., Medeni, I. T., Guzel, M. S., & Askerzade, I. (2020). Assessment of Iterative Semi-Supervised Feature Selection Learning for Sentiment Analyses: Digital Currency Markets. 2020 IEEE 14th International Conference on Semantic Computing.

[3] Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., & Anastasiu, D. C. (2019). Stock Price Prediction Using News Sentiment Analysis. 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (Big Data Service).

[4] Sarode, S., Tolani, H. G., Kak, P., & Lifna, C. S. (2019). Stock Price Prediction Using Machine Learning Techniques. 2019 International Conference on Intelligent Sustainable Systems.

[5] Ravikumar, S., & Saraf, P. (2020). Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms. 2020 International Conference for Emerging Technology.

[6] Mankar, T., Hotchandani, T., Madhwani, M., Chidrawar, A., & Lifna, C. (2019). Stock Market Prediction based on Social Sentiments using Machine Learning. 2018 International Conference on Smart City and Emerging Technology.

[7] Kim, J., Seo, J., Lee, M., & Seok, J. (2019). Stock Price Prediction Through the Sentimental Analysis of News Articles. 2019 Eleventh International Conference on Ubiquitous and Future Network.

[8] Gupta, R., & Chen, M. (2020). Sentiment Analysis for Stock Price Prediction. 2020 IEEE Conference on Multimedia Information Processing and Retrieval.

[9] Moghaddam A H, Moghaddam M H, Esfandyari M. Stock market index prediction using artificial neural network: [J]. Journal of Economics Finance & Administrative Science, 2016.