Stock Prediction Application using Machine Learning

Amir Gudekar[1], Salman Shaikh[2], Vikash Singh Bafila[3], Shital Agrawal[4]

[1,2,3]*Student, Department of Information Technology, ARMIET, Maharashtra, India*
[4]*Professor, Department of Computer Engineering, ARMIET, Maharashtra, India*

---***---

Abstract - *The stock market is an exceptionally powerful market where nothing is basically as steady as a stone however as the innovation is overhauling there are numerous ways and strategies one can attempt to realize this powerful change and be arranged appropriately. This paper centers around such various strategies for progressively learning the market and its patterns. We have involved two distinct models for this paper and have additionally performed sentiment analysis on the tweets in regards to the organization or the stock, the model with the least mistake is the ideal and the most favored technique for prediction. The aftereffects of this arrangement have given an unmistakable and shrewd thought regarding the random ups and downs of the market and furthermore another methodology for investors with the goal that they know where they can wager their cash. The ARIMA model is giving the best accuracy for each stock.*

Key Words: **ARIMA, Time Series, Forecasting, Sentiment Analysis, Stock Market Prediction, Tweets**

1.   INTRODUCTION

The historical backdrop of the world's stock markets returns to the seventeenth century when the Dutch East India Company was recorded on an authority stock trade in 1611. From that point onward, investors have been attempting to search for different procedures to accumulate information about the recorded organizations in the field of Stock Markets and track down ways of further developing their venture returns. At first, investors essentially depended upon their previous experience to distinguish designs and anticipate stock costs. Yet, with the field developing at such a high speed with north of 30 million exchanging accounts across India, the customary methods become irrelevant.

With the steadily developing technology, investors are moving towards utilizing Intelligent Trading Systems instead of fundamental analysis, which likewise assists them with pursuing better speculation choices. It appears to be an impossible assignment to coordinate the experience and impressive skill of an accomplished trader who has been in the field for a really long time however with how much data accessible and technological advancements, it's very conceivable to think of algorithms to foresee stock prices. The field of Machine Learning has seen critical improvement in monetary time series prediction. Eval Becker et al. in the collection of tweets regarding a particular company have shown positive results [1]. In this paper, we have gathered tweets utilizing the Twitter API called Tweepy. Following this

comes the test to group the dataset with the organization name and sentiments displayed in the tweets.

As a part of Qualitative analysis, news feed has a significant impact on how the stock behaves over a period of time. This proceeds to demonstrate that media and stock market trends are highly interconnected. [5]. It very well may be securely said that twitter is likely the most trustworthy and the quickest approach to consuming media. Discussing quantitative analysis, a great deal of information is accessible for preparing models. The model purposes the yahoo finance API to prepare the model and make predictions.

2.   LITERATURE SURVEY

Major stock price prediction methods have been built around the basic premise of fundamental and technical analysis, although recent studies have shown that stock prices have a strong correlation with the news articles of the company. [6].

Financial analysts utilized hourly stock prices of 30 unique stocks and their relating news stories and tweets with respect to the organization from the NASDAQ site. They likewise gathered the tweets connected with that multitude of 30 stocks for a range of a half year. Song et al. [2] collected six years of data from the Hong Kong stock market. [10]. They accumulated all the financials of those specific organizations and stocks and their comparing news stories and tweets for a similar time-frame to essentially draw a connection between news stories and stock patterns. Likewise for a specific trading day, they recorded the open, close, high, and low prices of a specific stock for each organization.

Dogra et al. [7] performed a detailed study on many classifiers such as KNN, Random Forest, SVM (Support Vector Machine), and Naive Bayes [4][8][11] on their efficacy in predicting stock trends. We can hence state that although the Naive Bayes model has substantially greater accuracy, the SVM [9] and f-measures of certain different algorithms such as Random Forest which states that they are overall better in performance. Deep learning models are extremely high-priced. Also, they do not provide a good perception on stock prices compared to other simplistic methods, so, they are not a suitable choice for an ensemble.

The latest things in the stock market show a connection with the past grouping of stock throughout a given time span. The Naïve Bayes approach just expects to anticipate the end price of a specific stock in view of the specific day's other market values like the opening price. In such a case working out the progressions in the price throughout a length of time can basically work on the exactness of the model. The stock patterns utilized de-pend on the components of the valuation

record like Working Capital, P/E Ratio, Price to Book Ratio, Price to Sales Ratio, Total Market Value and Cash Flow Ratio. This technique for trademark arranging with the assistance of the Genetic algorithm can be utilized as a pre-handling step which can then be considered to call attention to boss specialized measures which thusly is more useful to the merchants and can be just determined utilizing the closing price.

3. PROPOSED SYSTEM
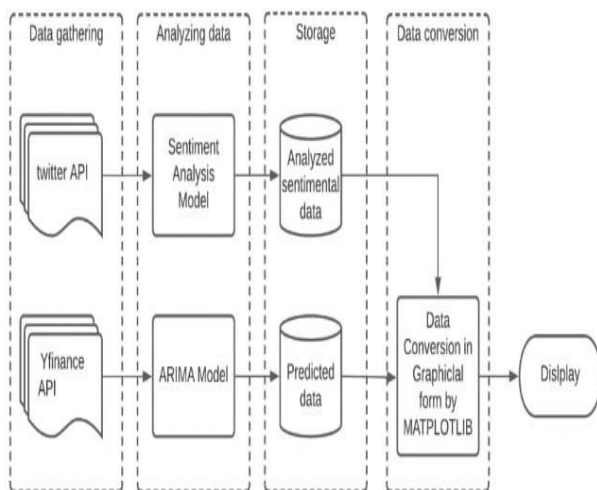
3.1. System Overview

Fig -1: System Architecture

The scope of this paper is to arrange the stocks as per their areas and analyze the sentiment from the tweets with respect to a specific stock or an organization as entirety. The flowchart in Fig. 1 exhibits the overall system architecture. The system can be isolated into these primary parts: a) The assortment of the stock information like the opening, closing, high, low prices and pre-processing of tweets for sentiment analysis, b) The classification of stock ticker utilizing model and c) Performing analysis on the labels acquired after characterization and d) Converting the Results in Graphical form.

3.2. The Proposed Method

i. Collection of Stock Ticker

The dataset is made utilizing tickers of various stocks from Yahoo Finance API and Twitter API called Tweepy is used to extract tweets from Twitter for sentiment analysis. The benefit of this technique is the capacity to determine watchwords which the model believes the tweets should have by characterizing a series of rules. For instance, we can characterize the string with explicit words either with organization names or stock names, also certain words can be featured, e.g.: open, close and so forth. In the Indian stock market, the Nifty file is the best 50 performing stocks so to take every one of the patterns for each of the stocks in that rundown like the Volume, Opening, Closing prices and burden in the model. For the tweets the retweets have not considered as there the given information or data probably won't be precise and subsequently during the forecast that information could modify the accuracy of the framework.

ii. Data preprocessing

Following the removal of such tweets, text pre-processing was finished by utilizing Regular Expressions. The presence of hashtags and mentions in the text expands the length of the sentence in this manner diminishing the capacity of the model to order the sentiments. [3]. Henceforth, they have been taken out. Furthermore, the tweets have been switched over completely to lowercase and different whitespace characters have been supplanted with only a solitary event. Subsequent to pre-processing the dataset additional tree elements will be utilized to choose the highlights which are the best foreseeing the end price of the stock. Scikit-learn executes an assessor that utilizations randomized choice tree to fit different sub tests of the dataset while utilizing the mean method. Henceforth the expectation accuracy of the model is expanded and the overfitting of the dataset is managed.

iii. Testing the accuracy of different models

Stock market prices contain sequential data i.e., everything could be planned to a specific event at some given time and can conjointly attempt to foresee the data for the future. The current brain network has numerous self-circles in the secret layers of the calculation to empower RNN to utilize the past arrangement of stowed away neurons so it can become familiar with the present status with the given input. Since RNN has the issue of disappearing inclination it's anything but an appropriate fit for the model. As the stock prices of different stocks vary by a huge sum the information needs to initially standardize and must be kept in a scope of 0 to 1 this information is then given to the model for preparing. The model which we have utilized has been prepared for 25 ages by changing the size of the layers for additional superior tuning. In the wake of preparing this large number of models were tried and the model with the least RMSE score is viewed as the best model to go on for the forecast.

4. RESULTS

The ARIMA model gives best outcomes while anticipating short term outcomes though the time series model is better for long term predicating. For the stock market expectation, we require transient forecast to be more exact as the client needs to run the model every day for each stock, so for this specific situation we don't need long term forecast to be more precise and consequently the ARIMA model is giving the best RMSE score.
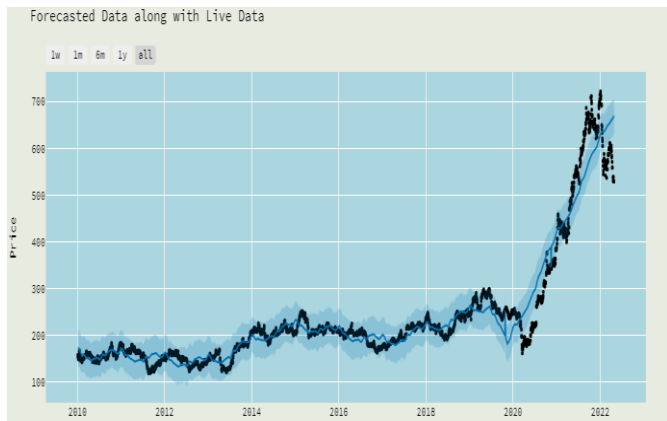
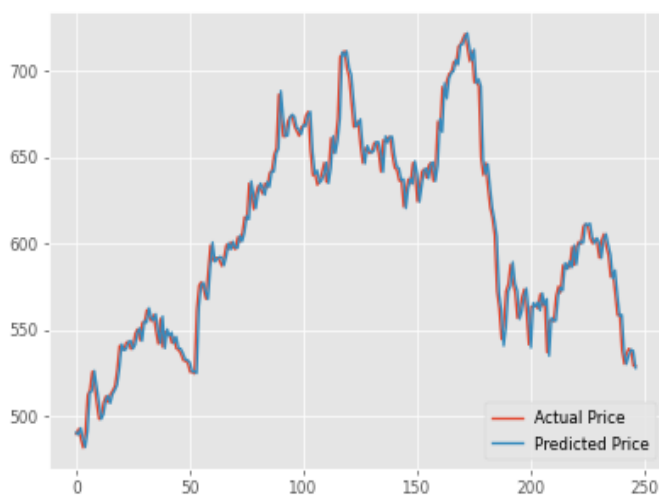**Chart -1**: Actual vs Predicted graph of Wipro NSE using Time Series (FbProphet)

Chart -2: Actual vs Predicted graph of Wipro NSE using ARIMA.

5. CONCLUSIONS

In this paper, we have compared and analyzed the ac-curacy of two algorithms specifically ARIMA and Time Series to foresee stock prices. We have utilized Tweepy, a python library to access. Twitter API to perform sentiment analysis of tweets. The App estimates stock prices of the following days for any stock recorded under NSE. The sentiment analysis of tweets joined with the anticipated prices prescribes the client whether to trade a specific stock or not.

REFERENCES

[1] Y. Wang and Y. Wang, "Using social media mining technology to assist in price prediction of stock market," 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, China, 2016, pp. 1-4, doi: 10.1109/ICBDA.2016.7509794.

[2] X. Li, H. Xie, Y. Song, S. Zhu, Q. Li and F. L.Wang, "Does Summarizing Help Stock Prediction? A News Impact Analysis," in IEEE Intelligent Systems, vol. 30, no. 3, pp. 26-34, May-June 2015, doi: 10.1109/MIS.2015.1.

[3] Yadav, A., Vishwakarma, D.K. Sentiment analysis using deep learning architectures: a review. Artif Intell Rev 53, 4335–4385 (2020). https://doi.org/10.1007/s10462-019-09794-5

[4] Amit Kumar Sirohi, Pradeep Kumar Mahato and V. Attar, "Multiple Kernel Learning for stock price direction prediction," 2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014), Unnao, India, 2014, pp. 1-4, doi: 10.1109/ICAETR.2014.7012901.

[5] W. Zhao et al., "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 1, pp. 185-197, 1 Jan. 2018 doi:10.1109/TKDE.2017.2756658.

[6] 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 (BigDataService).

[7] I. Kumar, K. Dogra, C. Utreja and P. Yadav, "A Comparative Study of Supervised Machine Learning Algorithms for Stock Market Trend Prediction," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp.1003-1007, doi:10.1109/ICICCT.2018.8473214.

[8] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017,pp.1643-1647,doi:10.1109/ICACCI.2017.8126078.

[9] Zainuddin, Nurulhuda & Selamat, Ali. (2014). Sentiment analysis using Support Vector Machine. I4CT 2014 - 1st International Conference on Computer, Communications, and Control Technology, Proceedings.333-337.10.1109/I4CT.2014 .6914200.

[10] Xiaodong Li, Pangjing Wu, Wenpeng Wang, Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong, Information Processing & Management, Volume 57, Issue 5, 2020, 102212, ISSN 0306-4573

[11] Sarkar, A. K. Sahoo, S. Sah and C. Pradhan, "LSTMSA: A Novel Approach for Stock Market Prediction Using LSTM and Sentiment Analysis," 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2020, pp. 1-6, doi: 10.1109/ICCSEA49143.2020.9132928.