Stock Price Predictor

CSE 523 Machine Learning - Group-8

Krinal Boghani[1], Mansi Savaj[2],

Sanskruti Shingala[3], and Udit Kapadia[4]

Group-8, Ahmedabad University,
krinal.b[1], mansi.s4[2], sanskruti.r[3], udit.k[4]@ahduni.edu.in

Abstract - Stock Market Prices are highly volatile, affected by multiple factors stemming from internal company matters/decisions or external market forces and overall investors' sentiments. With the introduction of technological breakthroughs such as global digitalization, stock market forecasting has entered a technologically enhanced era, reviving the traditional trading methodology. The application of advanced Machine Learning techniques has improved the prediction accuracy of stock prices. However, due to non-linear, irregular, and chaotic data, stock market analysis and prediction remain some of the most challenging research areas. This project tries to achieve the task of stock price prediction on the basis of machine learning algorithms: Autoregressive Integrated Moving Average (ARIMA) and Moving Average, Linear Regression, Exponential Moving Average. The data of the companies 'Vodafone-Idea Limited', 'Axis Bank Limited' and 'Asian paints Limited' for the years 2015-2020 have been critically analyzed under the framework of these algorithms. Furthermore, the models were evaluated on the basis of Root-Mean Squared Error (RMSE) values.

Keywords - Machine Learning, Technical Analysis, Linear Regression, Moving Average, RMSE-Root mean square error, ARIMA, Stock Market Prediction, Optimal Prediction

INTRODUCTION

Stock prices are characterized as non-linear, dynamic, and irregular in nature. The prediction of stock prices is a challenging task as these do not depend on fixed variables. Multiple factors affect the ups-downs of stock prices such as risks, demand, and supply at a point of time, companies' earnings and profits, investor's sentiments, etc. [1] To maximize the profit and minimize the loss, stock market prediction examines previous data and estimates future data. In the financial sector, machine learning (ML) has been widely used to provide a new mechanism that can help investors to make better investments. Machine learning is commonly used in stock market prediction; while it does not provide an exact result, it can anticipate the movement of the stock price.[2] There are two types of analysis that are popularly used:

a. Fundamental analysis
b. Technical Analysis

While the fundamental analysis indicates prediction on the basis of the intrinsic value of a stock through assessment of assets of the company, financial statements, industry performance, management efficiency, promoter holdings etc. These values remain constant for a long period of time due to their very nature. Hence, fundamental analysis is used by investors for long term trading.

On the other hand technical analysis focuses on the trends and patterns followed by the stock for a given time frame. It is not concerned with the internal nature of the company or company matters. This is often utilized by the stock traders (intraday and short term traders) to gain profits off of these patterns.While fundamental analysis remains a tough problem for the Machine Learning to undertake, this project intends to conduct technical analysis of various stocks.

LITERATURE SURVEY

The set of articles includes studies that fundamentally focus on machine learning algorithms for stock market prediction. These machine learning techniques include linear regression, Auto-Regressive Integrated Moving Average (ARIMA), Prophet, etc. Moreover, some statistical formulas for evaluating the models have also been studied. Enke and Thawornwong (2005) investigated the predictive relationships for a variety of financial and economic factors using a machine learning data retrieval technique. By analyzing the gained information from each model, only significant variables should be selected to keep in the forecasting models.

I. Description of data:

A. Data Analysis

Firstly, we observed the trends of the adjusted closing prices of various stocks in the Indian stock market. Impact of external factors like component decisions, government announcements (Subsidies or newer projects, etc.), budget and earnings announcements by the board of directors, changes in central management etc. Other stimuli like International relations of nations and states of conflict, seasonal changes and changes in the availability of the resources in the nation.

In later stages, we did a technical analysis of several stocks observing the patterns that they have followed since the inception of particular stocks. We also Observed Chart Patterns of graphs in Linear and Candle Graph formats.

B. Data Collection:

The historical data of three companies- 'Vodafone-Idea Limited', 'Axis Bank' and Asian Paints', has been collected from the *Yahoo! finance* website[7]. We have used data recorded by the national stock exchange. The data set includes 5-year data from January 2015 to December 2019 of the above companies. This data is in tabular form and contains approximately 1200 rows and 7 columns.

C. Data Preprocessing

The Data contains information about the variables such as High, Low, Open, Close, Adjusted close price, and Volume. However, only the adjusted closing price of the stocks has been used for the prediction. From some research, we have found that an adjusted closing price is more effective for stock price prediction. It gives more accurate values of stock price.

	Data Set	Training Data Set	Test dataset
Time Interval	01/01/2015-30/12/2019	01/01/2015-01/01/2019	01/01/2019-03/12/2019

Table 1. Statistics of the dataset

II. Implementing Moving Average Model:

Simple moving average model is a smoothing method based on the concept of averaging random fluctuations in the time series to identify the underlying direction in which the time series is changing. Here, we have tried 2 , 3 and 4 period moving averages.

$$F_{t+1} = \frac{At + At-1 + \dots + At-k+1}{k}$$

Here A_t is the observed value in period t and F_t is predicted for period t.

III. Implementing Linear Regression Model:

Linear Regression, one of the most simple approaches, can be used to predict any continuous variable, including stock price predictions. As the name implies, linear regression is a linear technique that identifies the linear combination of X factors that are used to predict the Y variable (closing price).

$$Y = \Theta_1 * x_1 + \Theta_2 * x_2 + \Theta_3 * x_3 + \dots + \Theta_n * x_n$$

Here, days of the week are taken as independent variables i.e. $x_1, x_2, x_3 \dots, x_n$

IV. Implementing Exponential Moving Average

The exponential moving average is a type of moving average that places a greater weight and significance on recent data points. The newest price data has the most impact on the moving average while the older price has minimal impact.It applies more weight to data that is more current. So, EMA will predict trends better than a simple moving average. Here, we have taken 30 days as the number of observations.

$$EMA = (K \times (C - P)) + P$$

Here, K = number of observations
C = current price
P = previous period EMA

V. The ARIMA Method

Autoregressive moving average method (ARIMA), is a method of regression analysis that helps in representing the strength of a dependent variable relative to other changing variables. It represents a statistical analysis model applicable to the time series data for better understanding of the data and predicting forthcoming changes in the price of the stock.

The general forecasting equation for ARIMA is as follows:

$$\hat{y}_t = \mu + \phi_1\,y_{t-1} + \ldots + \phi_p\,y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$

Here, 'θ' represents moving average parameter, φ is slope coefficient. We have used the forecasting ARIMA(1,1,1) series.
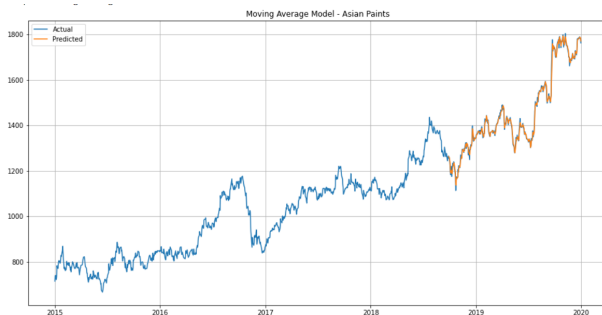
There are three important parameters in ARIMA:
p (past values used for forecasting the next value)
q (past forecast errors used to predict the future values)
d (order of differencing)

RESULTS

Here we have shown graphs of company Asian paints using our 4 models.
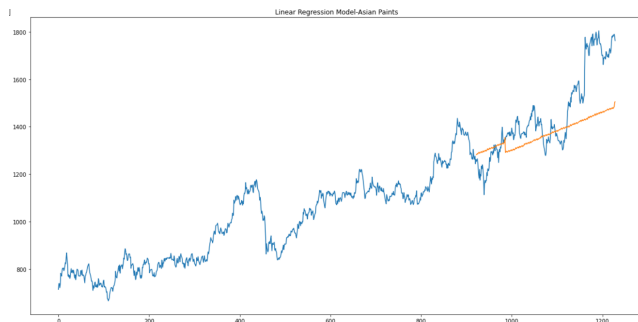
A. *Moving Average Models:*

RMSE Value -11.035834

RMSE values are quite low. The predicted values are of the same range as the observed values in the train set (it moves in trend and can predict the precise values).

B. *Linear Regression:*

RMSE Value - 151.609956

The RMSE values are very high. When applying a regression approach, the model overfits to the date and month columns. The model will examine the value from the same day a month ago, or the same date/month a year ago,

instead of the prior values from the point of prediction
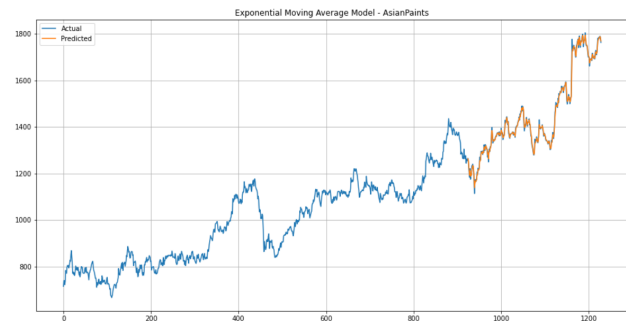
### C. *ARIMA Method*:

RMSE Value - 22.087966

The model was able to detect the trends in the series using these variables. Even while the predictions made with this technique are significantly better than those made with previously applied model, they are still far from accurate.

D. *Exponential Moving Average:*

RMSE Value -7.8764

RMSE values are very low as compared to other models. Exponential MA gives more weight to most recent price data, which is considered to be more relevant than older data.
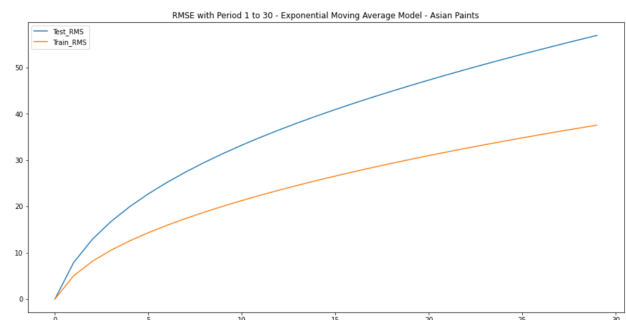
RMSE Value for weight=2 is minimum for test and train dataset.

Date	Open	Adj_Close	Pred_Close
17-12-2019	1735.2	1723.65	1720.436946
18-12-2019	1749.25	1756.29	1744.339252
19-12-2019	1778.8	1780.63	1768.535285
20-12-2019	1804.95	1776.83	1774.06588
23-12-2019	1795.8	1785.03	1781.373848
24-12-2019	1812	1780.98	1781.110537
26-12-2019	1799.4	1787.05	1785.071758
27-12-2019	1811	1789.67	1788.136794
30-12-2019	1810	1779.84	1782.607854
31-12-2019	1805	1762.71	1769.342185

Table 1. Predicted Closing Price Values (for Asian Paints Ltd) for the Test Set and Adjusted Close Value (EMA)

CONCLUSION

For this project, 4 different approaches were used to perform technical analysis on the stock prices of Indian Share Market. These are as follows: ARIMA, Simple and Exponential Moving Average and Linear Regression. After experimentation it was observed: while models like Linear Regression and simple moving average seem to predict the trend these couldn't forecast the values effectively rendering them useless for price prediction. On the other hand it is observed that Exponential Moving Average turns out to be the best model for predicting the stock prices due to its lowest RMSE values compared to other methods. This is due to greater emphasis (weight) being given to the recent (present-day) data that helps in better prediction and, hence, better fitting. More recent the data is, the more effective it would be to analyze the trend as well. Thus, we conclude that Exponential Moving Average serves as the best model amongst the aforementioned models for stock price prediction.

REFERENCES

[1] BUSINESS, F. (N.D.). RETRIEVED APRIL 24, 2022, FROM HTTPS://PEOPLE.DUKE.EDU/~RNAU/411ARIM.HTM

[2] CHEN, J. (2022, MARCH 18). EXPONENTIAL MOVING AVERAGE (EMA) DEFINITION. RETRIEVED APRIL 24, 2022, FROM HTTPS://WWW.INVESTOPEDIA.COM/TERMS/E/EMA.ASP#:~:TEXT=AN %20EXPONENTIAL%20MOVING%20AVERAGE%20(EMA)%20IS %20A%20TYPE%20OF%20MOVING,THE%20EXPONENTIALLY%20 WEIGHTED%20MOVING%20AVERAGE.

[3] HARPER, D. (2022, FEBRUARY 03). WHAT DRIVES THE STOCK MARKET? RETRIEVED APRIL 24, 2022, FROM HTTPS://WWW.INVESTOPEDIA.COM/ARTICLES/BASICS/04/100804.AS P

[4] HAYES, A. (2022, FEBRUARY 08). AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA). RETRIEVED APRIL 24, 2022, FROM HTTPS://WWW.INVESTOPEDIA.COM/TERMS/A/AUTOREGRESSIVE-INTEG RATED-MOVING-AVERAGE-ARIMA.ASP#:~:TEXT=ARIMA%20IS%2 0A%20METHOD%20FOR,POINTS%20INFLUENCE%20FUTURE%20D ATA%20POINTS.\

[5] OBTHONG, TANTISANTIWONG, JEAMWATTHANACHAI, & WILLS. (2020, MAY 06). A SURVEY ON MACHINE LEARNING FOR STOCK PRICE PREDICTION: ALGORITHMS AND TECHNIQUES. RETRIEVED APRIL 24, 2022, FROM HTTPS://EPRINTS.SOTON.AC.UK/437785/

[6] WEST, Z. (2021, JULY 31). CALCULATING MOVING AVERAGES IN PYTHON. RETRIEVED APRIL 24, 2022, FROM HTTPS://WWW.ALPHARITHMS.COM/CALCULATING-MOVING-AVERAGES -IN-PYTHON-585117/

[7] Y. (N.D.). YAHOO FINANCE - STOCK MARKET LIVE, QUOTES, BUSINESS & FINANCE NEWS. RETRIEVED APRIL 24, 2022, FROM HTTPS://FINANCE.YAHOO.COM/