

STOCK PRICE ANALYSIS AND PREDICTION SYSTEM

Mini- Project

Submitted in partial fulfillment of the requirement of **Mumbai University**
For the Degree of **Computer Engineering**

by

Mr. Tanuj Avinash Palaspagar (C/25)

ID no. TUS3F181926

Mr. Bhushan Rajendra Patil (C/28)

ID no. TUS3F181930

Ms. Mitali Vyankat Mane (C/31)

ID no. TUS3F181934

Under the Guidance of

Professor Umesh Mantale

(Adjunct Professor , Department of Computer Engineering)



Department of Computer Engineering

TERNA ENGINEERING COLLEGE

Plot no.12, Sector-22, Opp. Nerul Railway station,

Phase-11, Nerul (w), Navi Mumbai 400706

UNIVERSITY OF MUMBAI

2021



Terna Engineering College, Nerul
Department of Computer Engineering
Academic Year 2020-21

CERTIFICATE

This is to certify that the mini project entitled “**STOCK PRICE ANALYSIS
AND PREDICTION SYSTEM**” is a bonafide work of

Mr. Tanuj Avinash Palaspagar

Mr. Bhushan Rajendra Patil

Ms. Mitali Vyankat Mane

submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the Bachelor of Engineering (Computer Engineering).

Guide

Head of Department

Principal

Approval Sheet
PROJECT REPORT APPROVAL

This Mini Project Report – entitled “**STOCK PRICE ANALYSIS AND PREDICTION SYSTEM**” by following students is submitted in partial fulfillment and is approved for the degree of B.E. in ‘Computer Engineering’.

Submitted by:

Mr. Tanuj Avinash Palaspagar

TUS3F181926

Mr. Bhushan Rajendra Patil

TUS3F181930

Ms. Mitali Vyankat Mane

TUS3F181934

Examiner:

Date: 27 May, 2021

Place: Navi Mumbai

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Mr. Tanuj Avinash Palaspagar	TUS3F181926
Mr. Bhushan Rajendra Patil	TUS3F181930
Ms. Mitali Vyankat Mane	TUS3F181934

Date: 27 May, 2021

Place: Navi Mumbai

ACKNOWLEDGEMENT

We would like to express our sincere gratitude towards our guide **Prof. Umesh Mantale** , Mini Project Coordinators **Prof. Ramesh Shahabade** and **Dr. Kiran Bhandari** for their help, guidance and encouragement, they provided during the project development. This work would have not been possible without their valuable time, patience and motivation. We thank them for making our stint thoroughly pleasant and enriching. It was great learning and an honor being their student.

We are deeply thankful to **Dr. Archana Mire** (H.O.D Computer Department) and entire team in the Computer Department. They supported us with scientific guidance, advice and encouragement, they were always helpful and enthusiastic and this inspired us in our work.

We take the privilege to express our sincere thanks to **Dr. L. K. Ragha** our Principal for providing the encouragement and much support throughout our work.

Mr. Tanuj Avinash Palaspagar	TUS3F181926
Mr. Bhushan Rajendra Patil	TUS3F181930
Ms. Mitali Vyankat Mane	TUS3F181934

Date: 27 May, 2021

Place: Navi Mumbai

INDEX

Chapter	Page Number
1. Abstract	6
2. Introduction 2.1) Stock 2.2) Stock Market Behaviour	7,8
3. Objective	9
4. Literature Survey	10
5. Software Analysis 5.1) Waterfall Model	11,12
6. System Design 6.1) Use-case Diagram 6.2) Component Design 6.3) Activity Diagram	13,14
7. Methodology 7.1) System Structure 7.2) System Variance and Prediction Error	15-18
8. Implementation 8.1) Libraries and Frameworks 8.2) Working System	19-24
9. Conclusions	25
10. Future Scope	26
11. References	27

Chapter 1

ABSTRACT

In the Stock Price Analysis and Prediction System, the aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic.

Prediction of the stock market is a long-time attractive topic to researchers from different fields. In this project, we propose a prediction algorithm that predicts the next-day stock trend of the financial company with the aid of **Time series regression**. Time series regression is a statistical method for predicting a future response based on the response history and the transfer of dynamics from relevant predictors. It is commonly used for modeling and forecasting of economic, financial, and biological systems.

Chapter 2

INTRODUCTION

In this mini project, application of **Time series regression** will be carried out using **Python**(computer language for coding), **FBProphet** and **Streamlit** that caters the requirement sufficiently. **Streamlit** is an open-source app framework for Machine Learning and Data Science teams. It has Python libraries that makes it easy to create and share custom web apps for machine learning and data science. In just a few minutes you can build and deploy powerful data apps. **FBProphet** uses time as a regressor and tries to fit several linear and nonlinear functions of time as components. At its core, the Prophet procedure is an additive regression model. A piecewise linear or logistic growth curve trend. Prophet automatically detects changes in trends by selecting changepoints from the data.

2.1 Stock:

A stock (also known as equity) is a security that represents the ownership of a fraction of a corporation. The stock market refers to the collection of markets and exchanges where regular activities of buying, selling, and issuance of shares of publicly-held companies take place. Now, the question arises for stock market investors is whether the price of stocks will rise or not. The fluctuation of the stock market is violent and there are many complicated financial indicators. Only people with extensive experience and knowledge can understand the meaning of the indicators and they use them to make good predictions to get fortune. Most other people can only rely on luck to earn money from stock trading.

2.2 Stock Market Behaviour:

The stock market is very unpredictable, any geopolitical change can impact the share trend of stocks in the share market, recently we have seen how Covid-19 has impacted the stock prices, which is why, on financial data, doing a reliable trend analysis is very difficult. Predicting the stock price trend by interpreting the seemingly chaotic market data has always been an attractive topic to both investors and researchers. Among those popular methods that have been employed, Machine Learning techniques are very popular due to the capacity of identifying stock trends from massive amounts of data that capture the underlying stock price dynamics. It is an efficient way to predict a market value close to the tangible value, thereby increasing the accuracy.

Chapter 3

OBJECTIVE

In this dynamic world of technologies everything has been digitized and people are overloaded with work. For such a situation there is a need to develop a system that is as accurate and trustworthy. We have developed an application that uses Machine Learning to help them navigate in the fast-changing stock market. The project aims to introduce and democratize the latest Machine Learning technologies for retail investors. No prediction is 100% accurate. Therefore, the upper bound and lower bound of the stock prices will be displayed to illustrate the trading range the investors should be looking at. This application serves as a supplementary quantitative tool for investors to see the market at a different perspective with the help of technology. A correct prediction of stocks can lead to huge profits for the seller and the broker.

Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of the respective stock market. In this project the primary objective of our work was to develop a robust system for analysing stock price movement based on the past stock price data. The system will be able to take input of parameters from history of the required stock value of a company. Based on the analysis of trends of the past data, it will be able to provide a graph of expected averaged prices over time, which will help people in investing at the right place.

Chapter 4

LITERATURE SURVEY

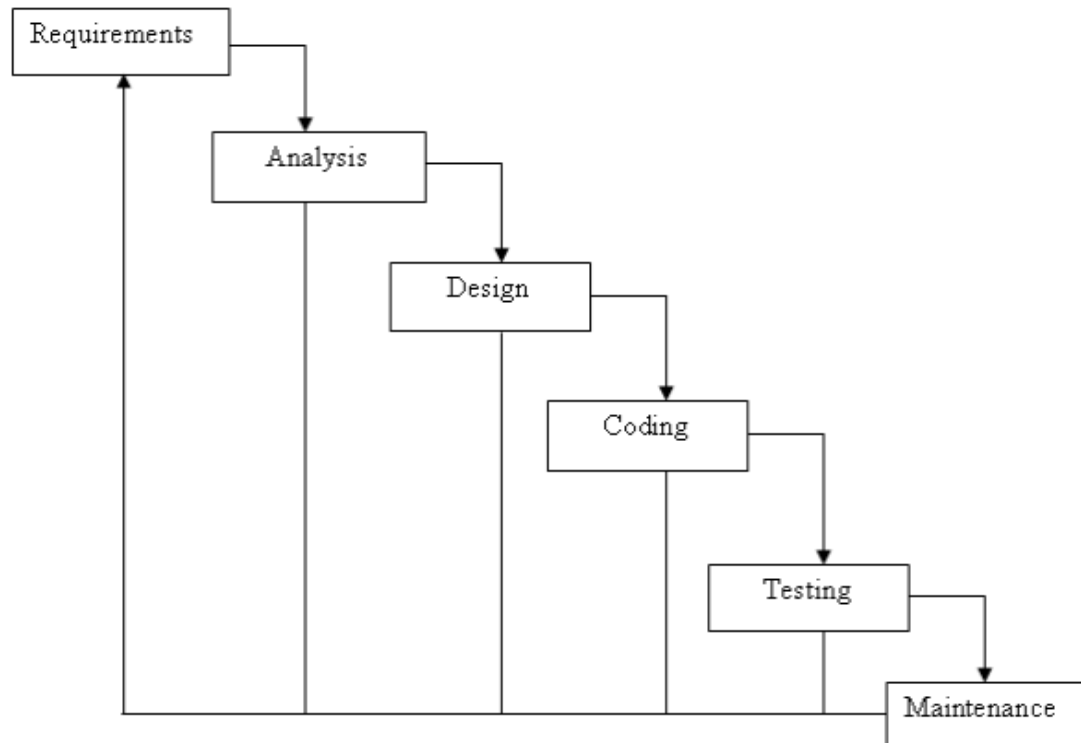
From the research paper “Machine Learning in Stock Price Trend Forecasting” written by Y. Dai And Y. Zhang in Stanford University, they used features like PE ratio, PX volume, PX EBITDA, 10-day volatility, 50-day moving average, etc. to predict the next-day stock price and a long-term stock price. The machine learning algorithms used in the research are Logistic Regression, Gaussian Discriminant Analysis, Quadratic Discriminant Analysis, and SVM. The accuracy ratio is defined as the number of days that the model correctly classified the testing data over the total number of testing days. With the short term model predicting the next day stock price, it has very low accuracy, the Quadratic Discriminant Analysis is the best among all models, it scored a 58.2% accuracy. With the long term model predicting the next ‘n’ day's stock prices, the longer the time frame, the better in the accuracy for SVM. With a time window of 44 days, the SVM model's accuracy reached 79.3%. Apart from that, it was found that by increasing the number of features, the accuracy increased. When all of the 16 features were used, the accuracy of the model reached 79%, while it fell to 64% when only 8 features were used, and 55% if only 1 feature was used.

The research paper “Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques” written by J. Patel, S. Shah, P. Thakkar, and K. Kotecha for the “Expert Systems with Applications” international journal demonstrated a way to use trend deterministic data to predict stock price movement. They conducted experiments using 10 technical indicators' signals as inputs, then they use prediction models to predict whether the stock will go up or down in the coming 10 days. Technical analysis indicators include SMA, EMA, Momentum, Stochastic SK, MACD, RSI, etc. The prediction models they have used include ANN, SVM, Random Forest, and Naive Bayesian models. The model outputs “up” or “down” movement signals. Experiments have shown random forest scored the highest performance with 83.56% accuracy with their inputs.

Chapter 5

SOFTWARE ANALYSIS

5.1 Waterfall Model



5.1.1 Requirements Analysis:

During this initial phase, the potential requirements of the application are methodically analyzed and written down in a specification document that serves as the basis for all future development. The result is typically a requirements document that defines what work the application should perform.

5.1.2 Design:

The requirement specifications from the first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture. Our system is designed in three phases: Firstly, data is collected and sorted for relevancy from various sources. Secondly, analysis is carried out on collected data by examining the market direction. And then lastly, training the model with the best possible technique.

5.1.3 Code/Implementation:

With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.

5.1.4 Implementation Testing:

All the units developed in the implementation phase are integrated into a system after testing of each unit. After integration the entire system is tested for any errors in the code

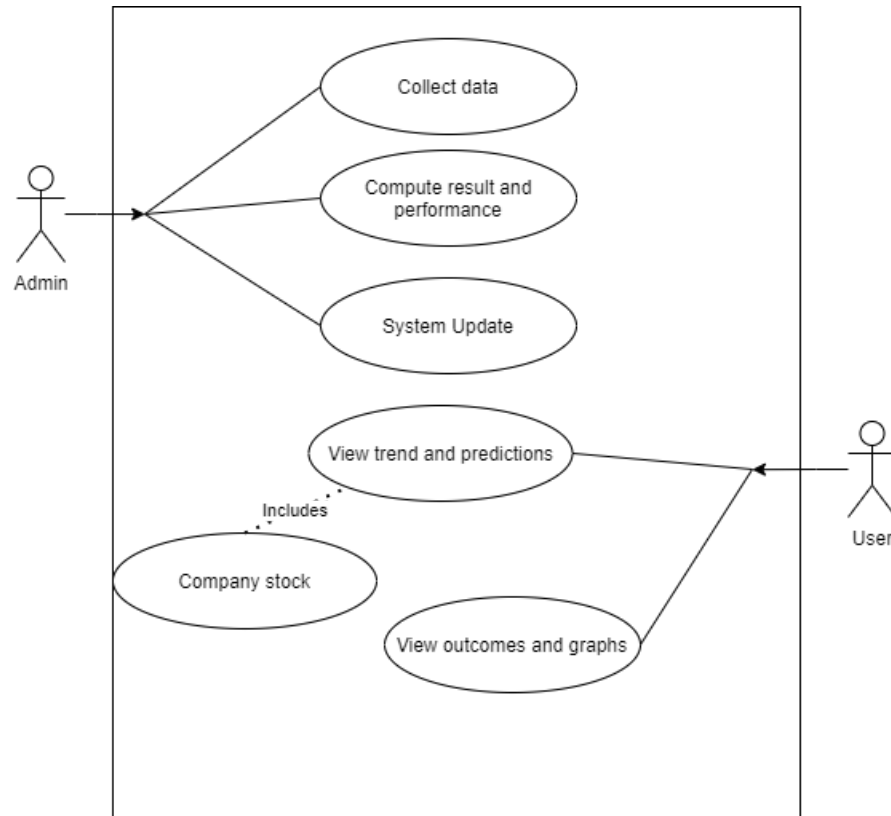
5.1.5 Deployment:

Once the functional testing is done, the product is deployed in the customer environment or released into the market. But for this project we have tested and deployed this project on the local server provided by streamlit.

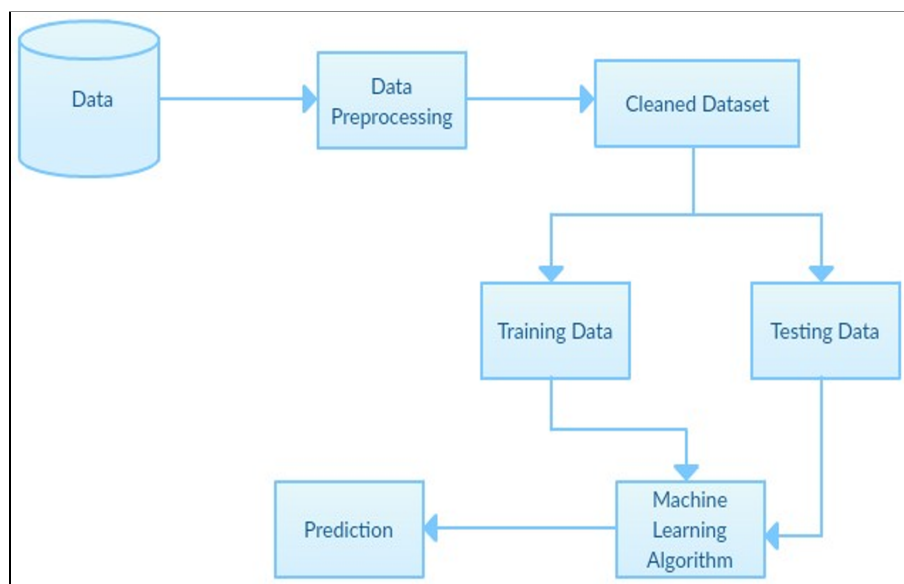
Chapter 6

SYSTEM DESIGN

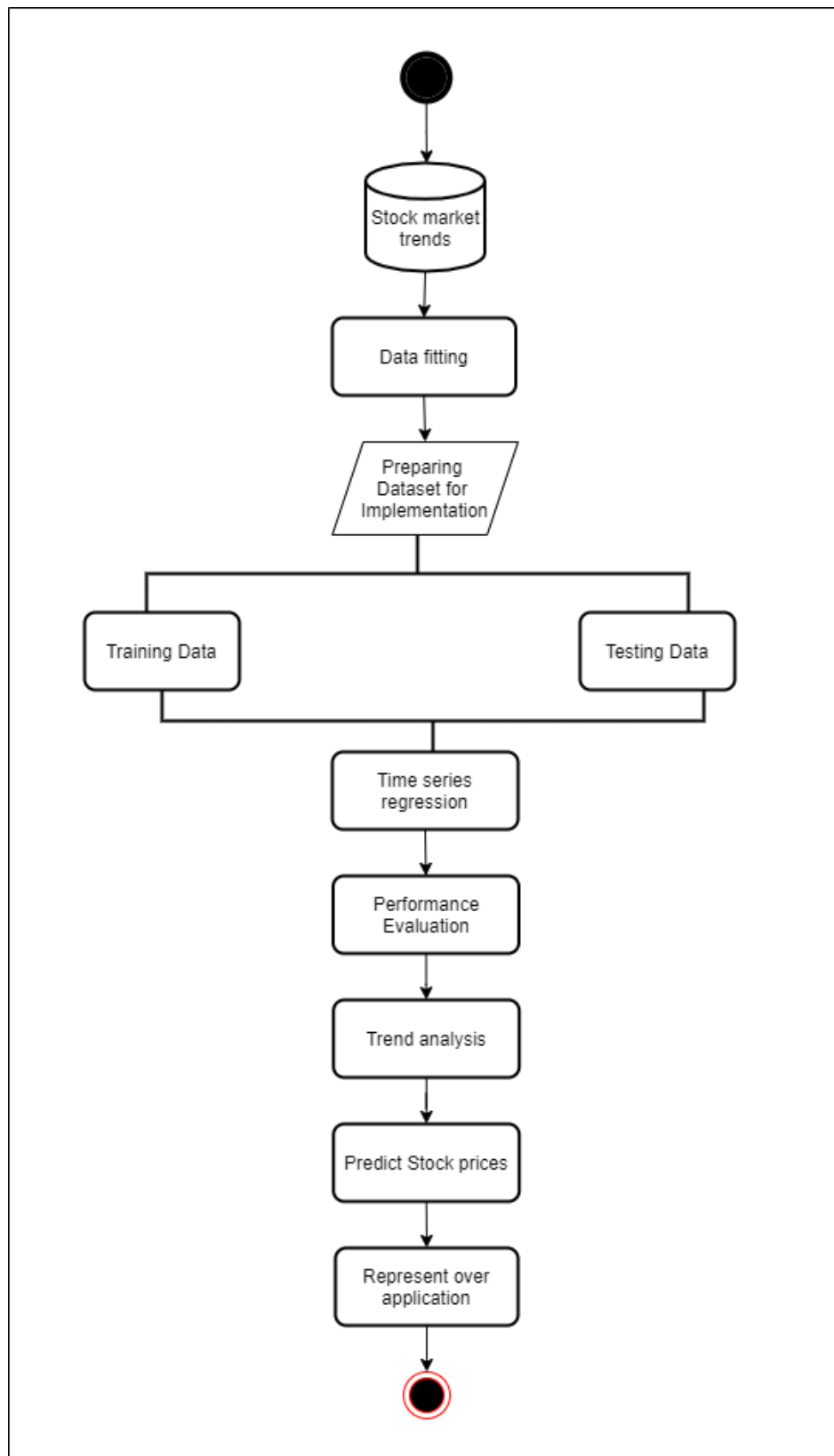
6.1 Use-Case Diagram:



6.2 Component Design



6.3 Activity Diagram



Chapter 7

METHODOLOGY

7.1 System Structure:

The proposed method for developing the system consists of mainly three parts. Firstly, data is collected and sorted for relevancy from various sources. Secondly, analysis is carried out on collected data by examining the market direction. Finally, training the model with the best possible technique.

7.1.1 Import Libraries

- a. We use `yfinance` for collecting and forming the datasets.
- b. The Prophet method from the `fbprophet` framework is imported.
- c. We use `plot_plotly` from the module `plot` of `fbprophet` as well as `graph_objs` from the `plotly` library for generating the different types of graphs.

7.1.2 Defining starting and ending date

- a. For a big enough dataset, we have found the optimum size of the dataset to be greater than 5 years.
- b. The system is set to today as the last date, with the first day of the year 2015 being the first day for the range of the dataset.
- c. Everytime this program is run, the latest stock values will be fetched.

7.1.3 Choosing the stocks to show in dropdown menu

- a. Here, the programmer has the control of which stocks the prediction system will provide for the end-user.
- b. This does not have any restrictions, and the appropriate symbols should be included that match the `yfinance` framework
- c. For the demo, we have chosen to add some top stocks from the NSE Stock Exchange, but as long as the stocks are registered with a nationally/internationally recognized Stock Exchange, the symbols and data will be available with `finance` for our purpose

7.1.4 Front-end components

- a. The front-end of the program is initialized in the `streamlit` framework
- b. Components like select box, slider, comments are created

7.1.5 Fetching and loading dataset

- a. From the previous set parameters, we try to fetch the stock history from yahoo
- b. The selected stock is sent as a ticker to the server with the start and today parameters
- c. This data is cached for future references to preserve resources
- d. Since this fetches data real-time, there is a requirement for internet connectivity to fetch and update the data set

7.1.6 Plotting the data in graphical format

- a. The raw data is plotted in scatter plot by the plotly module with Date on x-axis and Price value on the y-axis
- b. Two graphs are created: one for closing price and one for opening price and they are superimposed onto same plane
- c. This figure is then passed to the streamlit framework and the figure is connected with a rangeslider that allows for expanding and collapsing the density of graph on the x-axis (time) and panning through the graph

7.1.7 Training the model

- a. Appropriate train data frames are created with the date and close parameters
- b. Using the Prophet module, we fit the data into the data frame
- c. The future predictions are generated using 'Time Series Regression' and the prediction is stored in appropriate variables
- d. The forecast data is displayed as graph and tables, similar to how the raw data was displayed via streamlit

7.1.8 Showing the data variance trend

- a. The forecasted data from the previous method is passed through Prophet for determining the trend
- b. Trend over time in different intervals is analyzed and plotted through tools available from Prophet itself
- c. The trend over Yearly, Monthly and Weekly is plotted and displayed through Streamlit to the front-end
- d. This trend provides great insights about the nature of the stock market and the behaviour of its investors

7.2 System Variance and Prediction Error

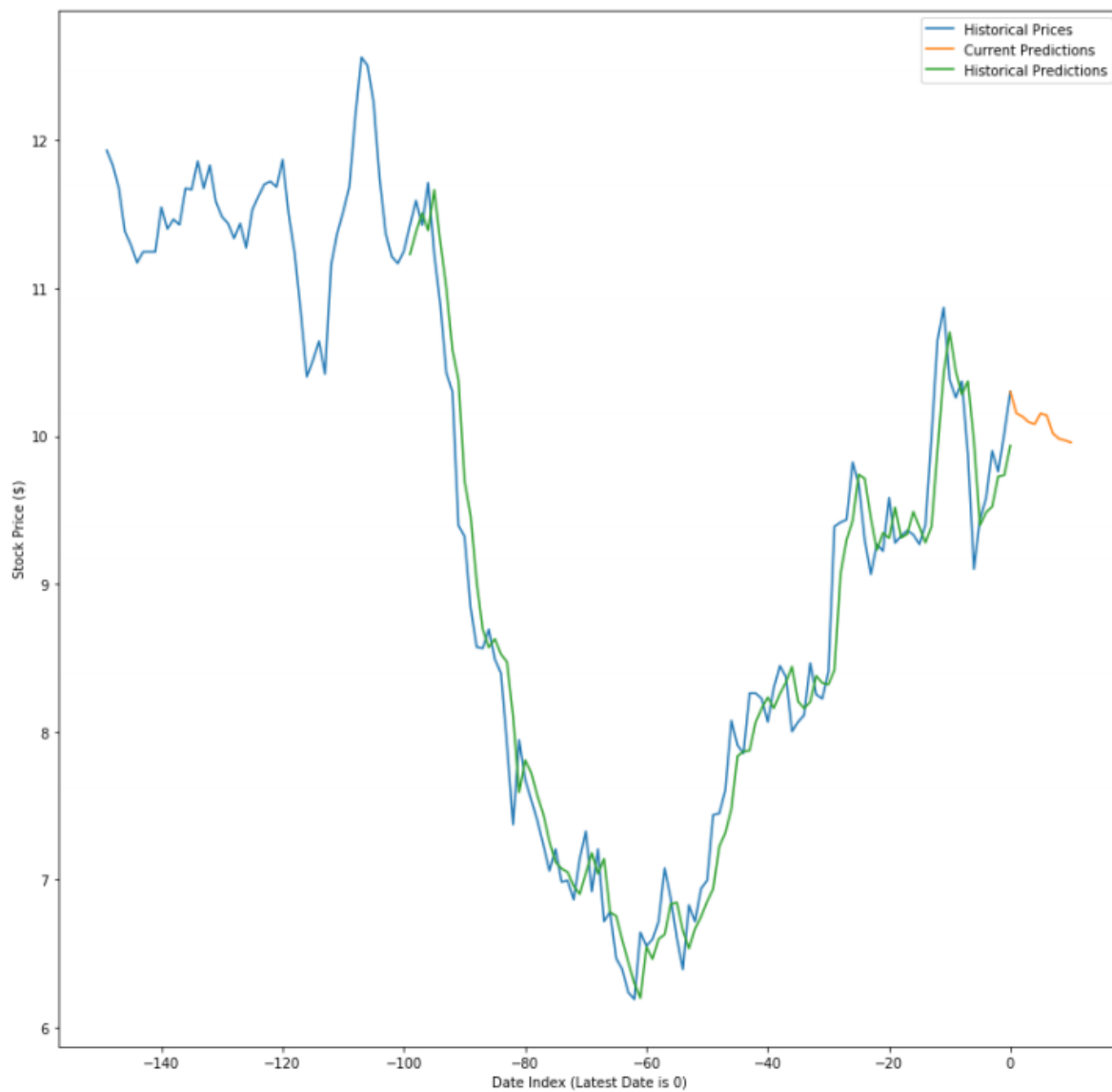
Accuracy of Stock Price Predictions, while the 1-day stock price prediction follows closely with actual stock prices, the predictions for stock prices after 10 days deviate considerably from the actual stock prices. This shows that the machine learning model fails to provide accurate stock price predictions to financial stocks of a company. The model has been shown to outperform predictions based on random walks and therefore might still serve as a reference for more savvy investors, who might be able to compare the results with their own analysis findings to discover meaningful trends.

According to the results for most stocks, most models that predict 10-day stock prices directly have a higher error than predicting individual stock prices. However, the errors in predicting different days in the future are relatively constant for models that predict 10-day stock prices directly, while the error increases with the time to predict stock prices one day at a time.

One possible explanation for such observation is that since the learned abstractions need to be applicable throughout 10 days, the high error from further prediction because of low correlation is propagated to other closer predictions. It results in a constantly increasing error for predicting all days.

On the other hand, predicting stock prices one at a time allows the model to learn the relationships between more correlated data points. It can be observed from the results that the first-day prediction is more accurate than predicting 10 days directly. However, the error for further predictions is higher.

Figure: 1-day interval historical predictions (GE, USD)



Chapter 8

IMPLEMENTATION

8.1 Libraries and frameworks:

8.1.1 Streamlit:

Streamlit is an open-source python library that makes it easy to create custom web apps for machine learning and data science. We have used this library to create the application user interface and to run the web application on the server.

8.1.2 Fbprophet:

Prophet is a forecasting procedure implemented in Python. Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality. This library helped us in training the historical datasets for predicting the future values.

8.1.3 Yahoo Finance (yfinance):

It is a python library which offers a temporary fix to the problem by fetching the data from Yahoo Finance and returning the data in the DataFrame format. So you can still use Yahoo Finance to get free stock market data. Yahoo Finance provides access to more than five years of daily OHLC price data. The real time stock data of different stocks in the web application is availed by yfinance library

8.1.4 Plotly:

The plotly python library is an interactive, open-source plotting library that supports many unique chart types covering a wide range of statistical, financial charts etc. The data is plotted in graphs dynamically with the help of this library.

8.2 Working System

8.2.1 From the predefined list of Stock symbols in the system, the user must choose the stock for analysis and prediction. The predefined symbols are upto the developer to include, which are updated in real-time through yfinance. In this demo, we have included a few stocks from the NSE stock exchange, but it is not limited to a particular Stock Exchange.

Stock Forecast App

Select dataset for prediction

TCS.NS

TCS.NS

HDFCBANK.NS

SBIN.NS

INFY.NS

WIPRO.NS

HINDUNILVR.NS

ITC.NS

1573	2021-05-25T00:00:00+05...	3092	3,128.2500	3,082.1001	3114
1574	2021-05-26T00:00:00+05...	3120	3165	3,103.8000	3,158.5000

8.2.2 After selecting the stock all the historical data is loaded dynamically which includes daywise OHLC(Open High Low Close) values. We can choose the period(in years) for which we want to predict the data.

Stock Forecast App

Select dataset for prediction

TCS.NS

Years of prediction:

1

1

4

Loading data... done!

Raw data

	Date	Open	High	Low	Close
1570	2021-05-20T00:00:00+05...	3,067.1001	3,088.8000	3,052.1001	3060
1571	2021-05-21T00:00:00+05...	3061	3,088.2000	3,055.1001	3,080.5000
1572	2021-05-24T00:00:00+05...	3,081.5000	3105	3072	3,081.5000
1573	2021-05-25T00:00:00+05...	3092	3,128.2500	3,082.1001	3114
1574	2021-05-26T00:00:00+05...	3120	3165	3,103.8000	3,158.5000

8.2.3 This historical data is automatically plotted in graphical format(using plotly library). We can zoom in and zoom out the graphical data by changing the position of the slider below. The darker slider allows panning of the curve over sparse time periods.

Time Series data with Rangeslider



8.2.4 The predicted data will be shown in the next frame in both numerical and graphical form. The predicted graph shows the possible market price range for the stock.

Forecast data

	ds	trend	yhat_lower	yhat_upper	trend_lower
1935	2022-05-22T00:00:00+05...	4,395.1714	3,760.0926	4,994.2847	3,781.1487
1936	2022-05-23T00:00:00+05...	4,398.2429	3,757.0518	4,966.9122	3,784.8257
1937	2022-05-24T00:00:00+05...	4,401.3144	3,731.3158	4,978.8734	3,786.9228
1938	2022-05-25T00:00:00+05...	4,404.3860	3,738.7455	4,993.5586	3,787.0200
1939	2022-05-26T00:00:00+05...	4,407.4575	3,746.9517	4,989.4436	3,786.5376

Forecast plot for 1 years



Predicted Data with Graphs

8.2.5 It is also possible to analyze the trend, weekly data and yearly data by using the following graphs

Forecast components



Chapter 9

CONCLUSION

The project lays the foundation for democratizing machine learning technologies for the future value of the financial stocks of a company. It helps investors navigate through the stock markets with additional analysis and helps them make more informed decisions.

It has led to the conclusion that it is possible to predict stock markets with more accuracy and efficiency using machine learning techniques.

In this project, we have applied the Time Series Regression technique in predicting the stock price trend of a financial company. Time Series forecasting is really useful when we have to make future decisions or we have to do analysis. The findings concluded the usefulness of the evolution algorithm in lowering the mean squared error when predicting stock prices, which is helpful for improving the trend prediction for retail investors.

Chapter 10

FUTURE SCOPE

This system can further be modified or developed to feature the following:

- News Gathering and Rectification
- Real time re-analysis using dynamic fetching of LTP(Last trade Price) and CMP(Current Market Price)
- Meter reading to suggest user for Buying/Selling/Holding
- Testing our predictor on different stocks to see its robustness. Try to develop a “more general” predictor for the stock market.
- For more accurate prediction, more features are needed to provide more useful information

Chapter 11

REFERENCES

- [1] “Survey Finds Hong Kong Securities Market Attracts Wide Range of Investors,” HKEX, 13 Jul 2017;
http://www.hkex.com.hk/news/news-release/2017/170713news?sc_lang=en
- [2] Y. Dai and Y. Zhang, “Machine Learning in Stock Price Trend Forecasting,” Stanford University;
<http://cs229.stanford.edu/proj2013/DaiZhang-MachineLearningInStockPriceTrendForecasting.pdf>
- [3] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques,” Expert Systems with Applications: An International Journal, Vol. 42, Jan. 2015, pp. 259-268
- [4] Streamlit, version 0.81.0 Available: <https://docs.streamlit.io/en/stable/>
- [5] FbProphet version 0.7.1 Available:
https://facebook.github.io/prophet/docs/quick_start.html