STOCKS PREDICTION

A Project Work

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

CSE w/ IBM spec in
BIG DATA ANALYTICS

Submitted by:

Ishita Raghuwanshi
18BCS3753

Addhyan Pant
18BCS3780

Avinash Deshwal
18BCS3790

Under the Supervision of:

Ms. Monika

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
APEX INSTITUTE OF TECHNOLOGY
CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,
PUNJAB

February- May 2021

DECLARATION

We, **Ishita Raghuwanshi(18BCS3753)**, **Addhyan Pant(18BCS3780) and Avinash Deshwal(18BCS3790)** students of **'Bachelor of Engineering in Computer Science Engineering with Specialization Big Data Analytics'**, **session: 2018-2022**, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled '**Stock Trend Analysis and Prediction'** is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Ishita Raghuwanshi
18BCS3753
Addhyan Pant
18BCS3780
Avinash Deshwal
18BCS3790

Date: April 20, 2021
Place: Chandigarh University

ACKNOWLEDGMENT

Aside from our own efforts, the progress of every project is largely dependent on the support and guidance of many others. We'd like to take this opportunity to thank everyone who has contributed to the project's progress.

We would like to express our gratitude to Ms. Monika. We appreciate her great support and assistance. All of the members who contributed and continue to contribute to this project provided invaluable advice and encouragement, which was critical to the project's success.

List Of Tables

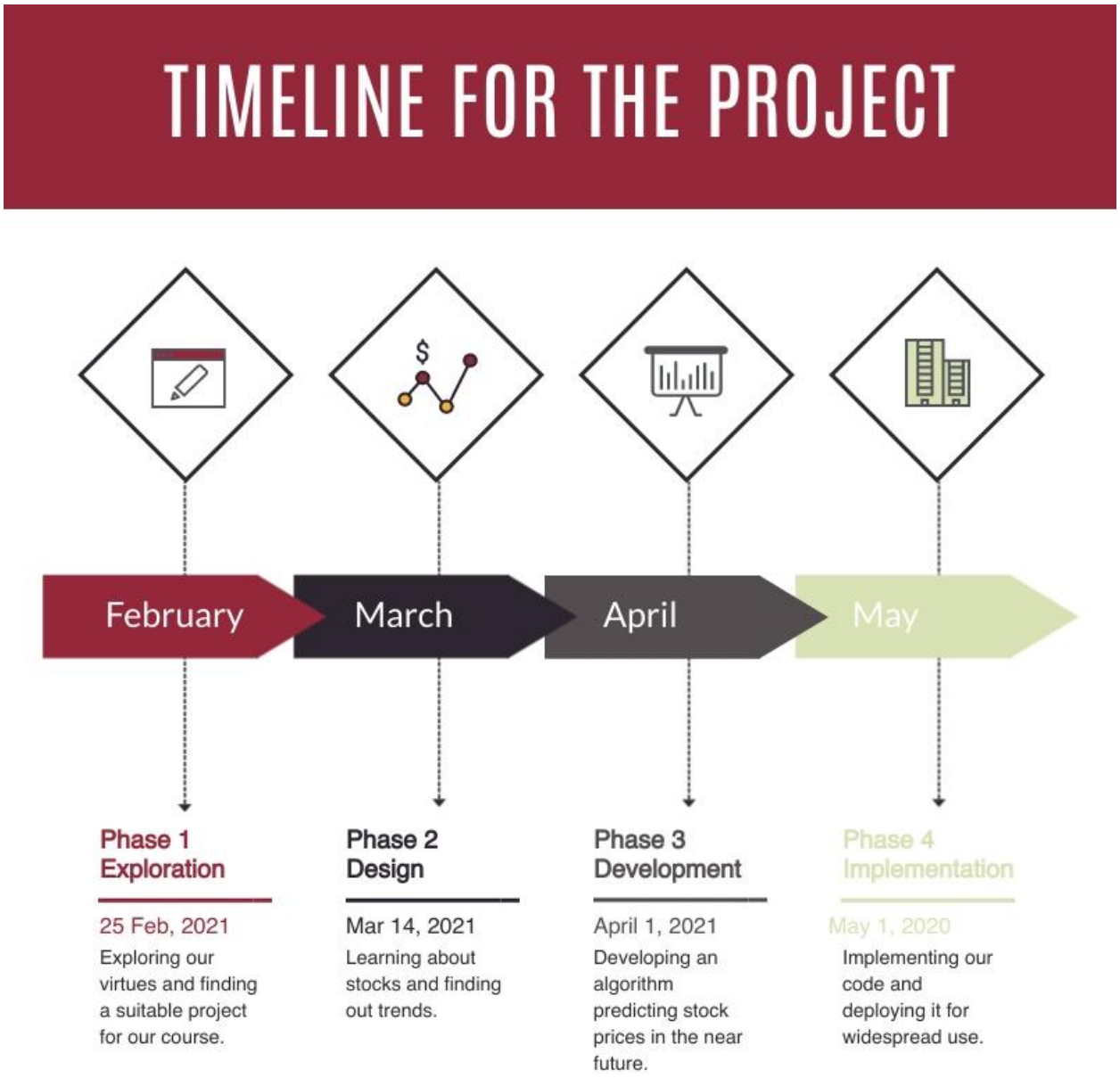
List of Figures

ABSTRACT

Forecasting stock prices is a common and significant subject in finance and academia. Since there are no substantial rules to estimate or forecast the price of shares in the capital market, it is an untidy place to predict. Many approaches are used to try to forecast the price in the stock market, such as technical analysis, fundamental analysis, time series analysis, and statistical analysis, but none of these methods has been proven to be a reliably appropriate prediction tool.

Our aim is to develop software that analyzes historical stock data for specific companies using various parameters that influence stock value. These values can be used in machine learning and data mining algorithms. This would also assist us in assessing the value of a given stock in the immediate future.

TIMELINE


TIMELINE FOR THE PROJECT
February
March
April
May
Phase 1
Exploration
25 Feb, 2021
Exploring our virtues and finding a suitable project for our course.
Phase 2
Design
Mar 14, 2021
Learning about stocks and finding out trends.
Phase 3
Development
April 1, 2021
Developing an algorithm predicting stock prices in the near future.
Phase 4
Implementation
May 1, 2020
Implementing our code and deploying it for widespread use.

INTRODUCTION

The act of attempting to forecast the future value of a company stock or other financial instrument traded on an exchange is known as stock market prediction and analysis. The stock market is an integral part of the country's economy and plays an important role in the development of the country's business and trade, which has an influence on the country's economy. Investors and companies alike are interested in the stock market and want to know if a particular stock can grow or fall over time. The stock market is the most common way for a business to raise capital for expansion. It is based on the demand and supply principle. While there is a higher demand for a company's stock, the company share price rises, and when there is a lower demand for a company's stock, the company share price falls.

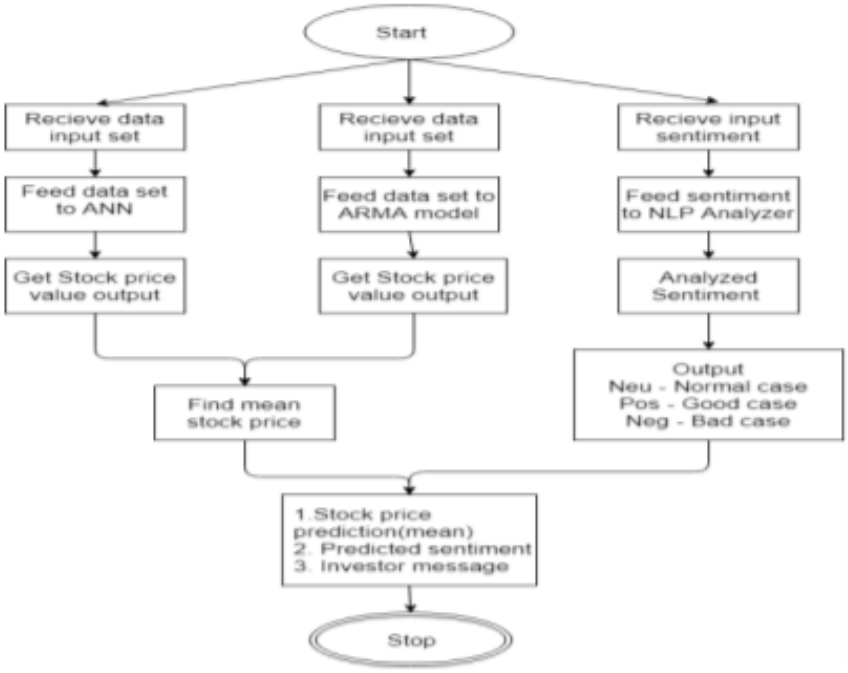
Another explanation for this field's study is the various theoretical and experimental obstacles it poses. The Efficient Market Hypothesis (EMH) notes that in an efficient market, stock market values completely represent available knowledge about the market and its constituents, and thus any potential to gain excess profit vanishes. The New York Stock Exchange is one example of a large exchange.

PROBLEM STATEMENT

The stock market is highly complicated and difficult to grasp. Because of the market's extreme uncertainty, it is considered too volatile to be predictable. Stock market prediction tasks are fascinating because they split researchers and academics into two camps: those who believe we can formulate methods to forecast the market, and those who believe the market is efficient and absorbs new knowledge by correcting itself, leaving little space for prediction. Investing in a good stock at the wrong time can backfire, while investing in a good stock at the right time can pay off handsomely. Today's financial investors face a trading dilemma because they don't know which stocks to purchase or which stocks to sell in order to produce the best returns. As a result, the proposed project would solve the issue of acceptable accuracy in such a real-time scenario.

PROJECT OVERVIEW

The project would assist investors in making stock market investments based on a number of factors. The aim of the project is to create visualizations that analyze previous stock data from companies and use machine learning to predict the value of that stock in the near future with fair accuracy. Firstly, we have created an analysis of the stock on Microsoft Power BI as well as Tableau, which are both common BI tools used in a professional workspace. On Microsoft Power BI we have used a webhook of Yahoo Finance to get quotes of past and present stock price of the Company of your choosing. You can choose the company by typing its name over on the StockList excel sheet which is being used to gather the data. For tableau, we went a step further, we created a python script which gathers the lifetime stock quotes of certain companies from Yahoo finance, and stores them in a csv file. Those stock data are used for the Big Data Analytics being performed on Tableau. Over there we can see the market cap and volume of every major stock, classified on what type of company it is from. And finally we'll use Yahoo Finance to get stock data, visualize various aspects of it, and then look at a few different ways to assess a stock's risk based on its previous performance history on Jupyter Notebooks. We'll also use the neural network to forecast stock prices and create a dashboard for stock analysis using various Python libraries. We have worked on a lot of data sets. The output in the end is done in normalized and denormalized form.



SOFTWARE SPECIFICATION

Tools and Libraries:

Sequential	SKLearn	Matplotlib	ggplot	Dense, LSTM

Softwares Used:

Tableau	Jupyter Notebook	Microsoft Power BI

Hardware Specifications:

CPU	Dual core processor or above
RAM	4GB
Operating System	Windows 7 or above

EXISTING MODEL

Making money-related trade judgments is a complex and demanding errand of fiscal data guessing. For stock examiners, a figure regarding securities trading with high precision improves the return advantage. Expansion of efficient models for forecast conclusion is difficult, and it must be accurate, in light of the snare of budgetary trade financial data. Using data mining and AI techniques, this study attempted to create models for predicting securities trades and deciding whether to buy or keep a stock.

Stock price forecasting is a commonly discussed topic in a variety of fields, including trading, economics, statistics, and computer science. If stock market depositors can decide when to enter and exit a position, they can maximize their yield by exporting or selling their investment. Necessary and/or technical research are commonly used by specialized traders to inspect stocks before making investment decisions. The study of business fundamentals such as proceeds and earnings, market place, growth rates, and so on is part of vital research. Technical research, on the other hand, is focused on the analysis of price oscillations in the past. Because of the dynamics of market forces, economies tend to adopt a trend of long-term expansion and contraction. The stock market operates in a broad setting in which the economy shifts from one stage of the economic cycle to the next.

In comparison to previous research, this project examines stock market decisions using neural networks and LTSM in the context of a changing economic and business climate.

PROPOSED MODEL

As previously mentioned, stock market forecasting is a vast topic with many aspects to investigate, but one thing that all models have in common is their check on correctness of how well the models will adapt to a given dataset and whether the results and forecasting are similar or not.Still, each model has a few effects in common: they all need a list of companies from every stock exchange to forecast upon the three basic market buy, hold, and sell situations, and to do so, stock market data for each company against their tickers was stored in a machine (to reduce access time) and data manipulations were performed to prepare the dataset for further machine processing.

PROBLEM FORMULATION

The popularity of stock market trading is growing rapidly, which is encouraging researchers to find out new methods for the prediction using new techniques. The forecasting technique is not only helping the researchers but it also helps investors and any person dealing with the stock market. In order to help predict the stock indices, a forecasting model with good accuracy is required. In this work, we have used one of the most precise forecasting technology using Recurrent Neural Network and Long Short-Term Memory unit which helps investors, analysts or any person interested in investing in the stock market by providing them a good knowledge of the future situation of the stock market.

RESEARCH OBJECTIVE

The following are the project's objectives:

1. To forecast the stock market by producing a trend from a wide collection of data.
2. Visualizations can be used to offer analysis to users.

The project would assist investors in making stock market investments based on a number of factors. The aim of the project is to create visualizations that analyze previous stock data from companies and use machine learning to predict the value of that stock in the near future with fair accuracy. Individuals will look at these predicted and evaluated data to learn about corporations' financial status and make comparisons. It can be used by companies and industries to solve their weaknesses and increase their stock value. Researchers, stock traders, market makers, the government, and the general public will all benefit from it. By generating a trend, the key feature of this project is to produce an estimated forecasting output and a general idea of future values based on previous results. This project's focus is restricted to a generalized suggestion generator.

METHODOLOGY

The proposed method for developing the system consists of mainly four main steps. Firstly, data is collected and sorted for relevancy from various sources. Secondly, the data cleaning process is done and analysis is carried out on the collected data. Thirdly, Prediction of the stock prices is done using a neural network. At last, a dashboard will be built using Plotly and different data visualization dash for stock analysis. This project aims to forecast the stock's value based on its previous performance and patterns. It necessitates stock market historical details, as the project also emphasizes data mining techniques. As a result, a reliable source with appropriate and sufficient data for the prediction is needed. We will be using two data sets one from Yahoo and the another by uploading a csv file. Both the data sets contain details such as Highest value, Opening value, closing value etc. The site is i.e Yahoo is updated on a regular basis and it has a repository for years of stock market. The stock market is a market with a lot of ups and downs. There are several companies in various industries, and the values and criteria will change over time.

Date	High	Low	Open	Close	Volume	Adj Close
2012-01-03	6.805000	6.662500	6.706250	6.681250	14123200.0	5.335590
2012-01-04	6.757500	6.638750	6.700000	6.650000	9913600.0	5.310636
2012-01-05	6.793750	6.640000	6.651250	6.751250	12347200.0	5.391493
2012-01-06	6.750000	6.640000	6.726250	6.740000	10824000.0	5.382509
2012-01-09	6.932500	6.752500	6.771250	6.917500	16880000.0	5.524260
...	...	...	...	...	...	...
2021-03-01	17.670000	17.379999	17.420000	17.660000	7592500.0	17.660000
2021-03-02	18.090000	17.790001	17.920000	18.020000	10305700.0	18.020000
2021-03-03	18.610001	18.110001	18.410000	18.459999	22907500.0	18.459999
2021-03-04	18.680000	18.010000	18.290001	18.170000	11839300.0	18.170000
2021-03-05	18.299999	17.840000	18.250000	18.280001	8945500.0	18.280001

2308 rows × 6 columns

The biggest difficulty in forecasting the stock market is that it is a chaotic system. There are many factors that may have a direct or indirect effect on the stock market. The variables have no significant relationships with the price. We are unable to establish a mathematical relationship between the variables. There are no rules that can be used to estimate the price of a stock using these variables. The neural network approach is appropriate for this type of chaotic system because we do not need to understand the solution. This is a significant benefit of neural network methods. Traditional methods, on the other hand, need a thorough understanding of the inputs, algorithms, and outputs. We just need to display the correct output for the given inputs with the neural network.

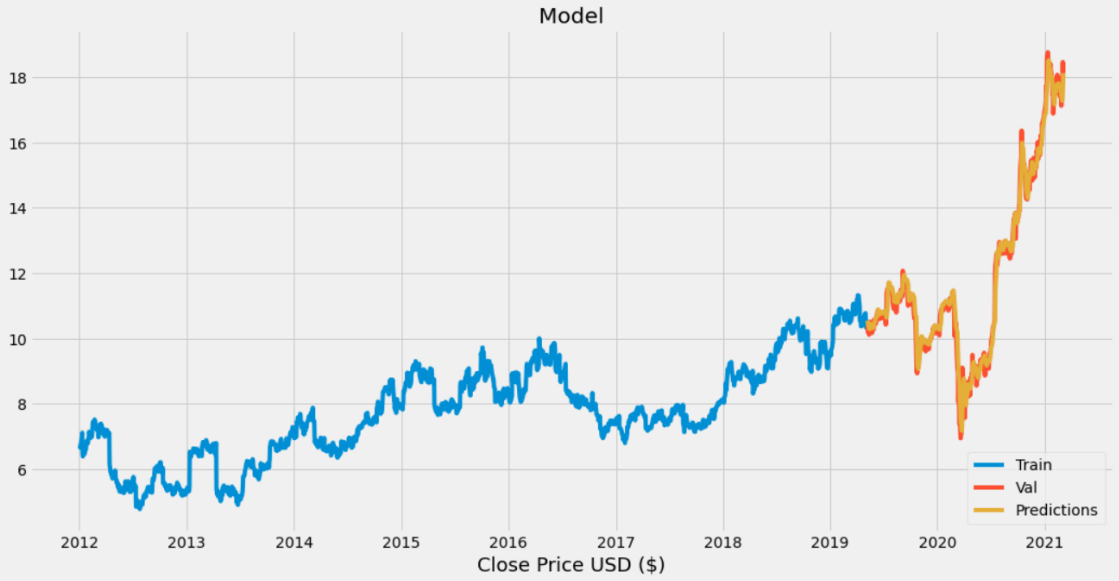
Before being fed into the neural network, the data is normalized. The training data input vectors are normalized such that all of the features are zero-mean and unit variance. The target values are normalized using the min-max equation, resulting in all values falling within the range of 0 to 1. The minimum and maximum values are expressed by 0 and 1, respectively. z = x min(x) max(x) min(x).

Testing modules built from the device design is done using unit testing. The inputs for each module are used to compile each element. During the unit testing process, all modules are combined into a larger unit. Each stage of project design and coding has been thoroughly tested. Module interface testing is done to ensure that information flows properly into and out of the program unit when it is being tested. By inspecting the local data structure, the integrity of the temporarily generated output data is guaranteed during the algorithm's execution. Last but not least, all error-handling paths are put through their paces.

The testing procedure is a subset of the larger topic of verification and validation. We must accept the device specifications and make every effort to satisfy the customer's requirements, and we must check and test the product for this sole purpose. Validation and verification are not the same thing. One is done to ensure that the program correctly incorporates a particular functionality, and the other is done to ensure that the final product meets the customer requirements.The project was checked to ensure that it complied with all of our project's requirements and specifications. We ensured that our project met the standards we set out for it at the outset of its production.

RESULT AND ANALYSIS

The following factors are considered in this project for changes in a company's closing price: General Index, Price Gap, Highest Value, Lowest Value, Share Amount, and Closing Price. We analyzed the data we collected to see if there was a connection between our performance parameters and the factors we chose.



Following the collection of data, a neural network is used to forecast the future share price.

Date	Close	Predictions
2019-05-08	10.390000	10.541470
2019-05-09	10.200000	10.484051
2019-05-10	10.260000	10.406861
2019-05-13	10.110000	10.360822
2019-05-14	10.210000	10.299530
...	...	...
2021-03-01	17.660000	17.274855
2021-03-02	18.020000	17.323788
2021-03-03	18.459999	17.546034
2021-03-04	18.170000	17.911610
2021-03-05	18.280001	18.120380

461 rows × 2 columns

CONCLUSION

As a result, it can be concluded that no trading algorithm can be 100 percent effective; not only that, but it will almost never be close to 70 percent, but even a 40 percent or 35 percent accuracy is sufficient to obtain a good forecast spread. To make our expectation more efficient, we can provide large data sets with millions of entries, which will help the machine learn more effectively. Various stock activities may result in a variety of price increases or decreases in the forecast price; use these fluctuations to judge if a business should be traded in or not.There has been no teaching. Data is never completely stable, so there will always be some unevenness, as seen in the above data spread, but forecasting close to a result is still a successful approach if the precision is greater than 33%. When creating a plan, traders should always consider having nominal imbalance while remaining above 33% accurate.

It can also be calculated that in a stock market, certain firms will not be associated at all, but most can be associated with each other, which can aid in the fairness of stock movements. We can scale affairs to see how much they are linked in percentages. Including massive data sets to maximize effectiveness, and in data sets with nan values in tables, replace the null values with 0 in both cases, which is something that a trader may want to adjust when creating a trading strategy.

REFERENCES

https://bit.ly/3ejlW5A
https://bit.ly/32t54DE
https://bit.ly/3wZW3QJ