PROJECT REPORT
ON
" STOCKS FORECASTING AND VISUALIZATION USING DASH "

SUBMITTED BY

TRIBE L MEMBERS

AMBRESH DESAI
HEENA NAWAZ
H ANJANA
GOURI POTDAR
VISHAKA RATHOD
SHWETA PAWAR

Under the guidance of

DINESH BEHERA

ABSTRACT

In the modern financial market, the most crucial problem is to find an essential approach to outline and visualize the predictions in stock-markets to be made by individuals in order to attain maximum profit by investments. The stock market is a transformative, non-straight dynamical and complex system. Long term investment is one of the major investment decisions. Though, evaluating shares and calculating elementary values for companies for long term investment is difficult. Stock price forecasting is a popular and important topic in financial and academic studies. Stock investments provide one of the highest returns in the market. Even though they are volatile in nature, one can visualize share prices and other statistical factors which help the keen investors to carefully decide on which company they want to spend their earnings on. In this project we have created a single page web application using the Dash library (of Python), we have made dynamic plots of the financial data of a specific company by using the tabular data provided by yfinance python library. On top of it, we have used machine learning algorithm to predict the upcoming stock prices.

CONTENTS

1. INTRODUCTION

Exchanging the stocks on money markets is one of the significant speculation exercises. Already, scientists developed different stock examination systems that could empower them to envision the bearings of stock esteem development. Predicting and foreseeing of significant future cost, in perspective of the present cash related information and news, is of colossal use to the financial pros. Financial masters need to know whether some stock will get higher or lower over a particular time-period. To obtain the accurate output, the approach used to implement is machine learning along with supervised learning algorithms. Results are tested using different types of supervised learning algorithms with a different set of features. Stock investment decisions require time, knowledge and awareness including historical data, the stock market contains a huge amount of data that varies over time. Stock prices are influenced by various factors ranging from the performance of the company itself to the conditions of the economy in general. Thus, to manage investment portfolios, stock market data has to be analyzed regularly to identify potential relationships between various stocks, hence to adjust investment based on related stocks trends. A picture is worth a thousand words presenting data in visual form can assist humans in exploring deep insight of vast amounts of complex raw data, especially when people have limited knowledge of the data. Visual representation is one of the most efficient ways to assist investors to have a clear overview of movements of the stock market, as well as providing a deeper understanding of each individual stock.

1.1. MOTIVATION

Predicting this stock value offers enormous profit opportunities which are a huge motivation for research in this area. Even a fraction of a second's knowledge of a stock's worth can result in large earnings. Similarly, in the repeated context, a probabilistically- correct prediction might be highly profitable. This attractiveness of finding a solution has prompted researchers, in both industry and academics to find a way past the problems like volatility, seasonality and dependence on time, economics and the rest of the market. However, the platform's prices and liquidity are highly unpredictable, which is where technology comes in to aid.

1.2. PROBLEM STATEMENT

The accuracy of the existing stock market prediction models is relatively low because only a small data-set is used for the training; the results will be less accurate. There is still a need to continually explore more new features that are more predictable. Even though multiple algorithms exist, there is no real-life implementation of these ideas for the benefit of people. Efficient algorithms should be made available with easy accessibility and interface.

1.3. OBJECTIVE

We will be creating a single-page web application using Dash (a python framework) and some machine learning models which will show company information (logo, registered name and description) and stock plots based on the stock code given by the user. Also the ML model will enable the user to get predicted stock prices for the dates inputted by the user.

2. DATA VISUALIZATION

- **What is Data visualization?**

  Data visualization refers to the techniques involved in graphically representing data, using visual elements like charts and graphs to spot trends, patterns, and outliers, for quick insights, and to help in real-time decision-making. It's increasingly important in today's world to understand the overwhelming volume of data being generated by businesses every single day.

- **Why is data visualization important?**

  Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

- Data visualization can also:

- Identify areas that need attention or improvement.
- Clarify which factors influence customer behavior.
- Help you understand which products to place where.
- Predict sales volumes.

2.1. Necessity of Data Visualization

According to the World Economic Forum, the world produces 2.6 quintillion bytes of data every day, and 90% of all data has been created in the last two years. With so much data, it's become increasingly difficult to manage and make sense of it all. It would be impossible for any single person to wade through data line-by-line and see distinct patterns and make observations. Data proliferation can be managed as part of the data science process, which includes data visualization. We need data visualization because a visual summary of information makes it easier to identify patterns and trends than looking through thousands of rows on a spreadsheet. It's the way the human brain works. Since the purpose of data analysis is to gain insights, data is much more valuable when it is visualized. Even if a data analyst can pull insights from data without visualization, it will be more difficult to communicate the meaning without visualization. The charts and graphs make communicating data findings easier even if we identify the patterns without them. The numerous types of data visualizations for example Line charts, Box plots, Area charts, Scatter plots, Bar charts. Population pyramids, Pie charts, Heat maps, Bar charts (actual vs. expected), Tree maps, Histograms, Bubble charts, Choloropleth, Network diagrams etc.

2.2. Sub-fields of Data Visualization

Data visualization is the presentation of quantitative information in a graphical form. In other words, data visualizations turn large and small datasets into visuals that are easier for the human brain to understand and process. Data visualizations are surprisingly common in our everyday life, but they often appear in the form of well-known charts and graphs.In terms of business intelligence (BI), these visualizations help users make better data-based decisions. Data visualization transforms raw data into information. The data visualization has three significant sub-fields.

2.2.1. Scientific Visualization

Scientific visualization is the representation of data graphically to gain understanding and insight into the data. This allows the researcher to gain insight into the system that studied information in ways previously impossible. It is also referred to as visual data analysis. The purpose is to convey the scientific data accurately, reveal underlying structures in data and encourage the exploration of the data. Scientific visualization is an interdisciplinary research and application field in science, focusing on the visualization of three-dimensional phenomena, such as architecture,meteorology, medicine or biological systems. Its purpose is to graphically illustrate scientific data, enabling scientists to understand, explain, and collect patterns from the data.

2.2.2. Visual Analytics

Visual Analytics can be perceived as an integrated approach that combines visualization, human factors, and data analysis. Visual analytics is a new field that has evolved with the development of scientific visualization and information visualization, with an emphasis on analytical reasoning through an interactive visual interface. Visual analytics methods allow decision makers to combine their human flexibility, creativity, and background knowledge with the enormous storage and processing capacities of today's computers to gain insight into complex problems. Visual analytics in the context of visualization relates to the areas of information visualization and computer graphics, and with respect to data analysis, it benefits largely from methodologies of information retrieval, data management & knowledge representation as well as data mining.

2.2.3. Information Visualization

Information visualization refers to the use of computer-supported, interactive visual representations of numerical and non-numerical abstract data sets in order to amplify human cognition. Information visualization, the art of representing data in a way that it is easy to understand and to manipulate, can help us make sense of information and thus make it useful in our lives. Information visualization is the communication of abstract data through interactive visual interfaces. Graphics such as histograms, trend graphs, flow charts, and tree diagrams all belong to information visualization and the design of these graphics transforms abstract concepts into visual information.

2.3. Goal of Data Visualization

The visual representation of data is more scientific than artistic in our modern world. The main goal of data visualization is effectively, efficiently, elegantly, accurately as well as meaningfully communicating information. It fulfills its objectives only if it encodes the given input in such a manner that our eyes can recognize and our brain can comprehend.

2.4. Data Visualization using Dash

Dash ships with a Graph component that renders charts with plotly.js. Plotly.js is a great fit for Dash: it's declarative, open source, fast, and supports a complete range of scientific, financial, and business charts. Plotly.js is built on top of D3.js (for publication-quality, vectorized image export) and WebGL (for high performance visualization).

Dash's Graph element shares the same syntax as the open-source plotly.py library, so you can easily switch between the two. Dash's Graph component hooks into the plotly.js event system, allowing Dash app authors to write applications that respond to hovering, clicking, or selecting points on a Plotly graph.

2.5. Data Visualization in the Stock Market.

Data visualization helps traders when making decisions quickly and enables them to easily synthesize large amounts of complex information.

- **How Data Visualization Can Improve Decision Making?**

- **Illustrating patterns:** Visuals are exceptional at presenting patterns. The human mind is wired to recognize patterns, but it's hard to make sense of those patterns when they appear as individual data points, such as in stock prices that fluctuate over a period of months. However, if those data points are plotted onto a linear graph, it's much easier to spot the highs and lows, and make a guess at how that pattern could unfold in the future.
- **Highlighting multiple variables:** Data visuals also allow you to highlight and control for multiple variables. Depending on your goals, this could help you factor multiple variables into your decision all at once, or help you drill down so you exclusively focus on one variable's effects on your investments. Either way, you can use visuals to make more informed decisions.
- **Reducing complex subjects:** One of the biggest advantages of data visualization is its ability to make complex subjects easier to understand, which is imperative in this age of big data. There are too many independent data points for any one person to track, but an automated platform can easily create a digestible version that appeals to your visual senses. Granted, not all financial topics are reducible to a simple conclusion, but visualization can make things more approachable, at the very least.
- **Adding customizing ability:** Most data visualization platforms have multiple controls that help a user customize their graphs and get exactly the visual they need to make a given decision. For example, you might be able to add or subtract specific variables, extend or contract the date range, or overlay competing subjects to get a clearer picture of what's going on.
- **Easing communication.** If you're investing with a partner, you should know data visuals also make it easier to communicate complex topics. Running a point-by-point analysis about each independent variable in your data set isn't going to go over well with a non-technical audience; but a graph can easily highlight your most important takeaways.

3. SOFTWARE REQUIREMENTS AND SPECIFICATION

3.1. Hardware Requirements:

- RAM: 4 GB
- Storage: 500 GB
- CPU: 2 GHz or faster
- Architecture: 32-bit or 64-bit

3.2. Software Requirements:

- Python 3.5 and above in Visual Studio code
- Operating System: windows 7 and above or Linux based OS or MAC OS

4. IMPLEMENTATION DETAILS

4.1. Python

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991. It is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc), it has a simple syntax similar to the English language. syntax allows developers to write programs with fewer lines than some other programming languages. It runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.Python can be treated in a procedural way, an object-oriented way or a functional way.

4.2. Details of Packages used

1. Dash: Dash is an open source framework which helps to create interactive, responsive and dynamic websites by using the Python or R programming languages. We used Dash core components to create tools to add inputs, sliders, drop down, graph, and other components to the web application to allow users to interact with the data. Dash's Hypertext Markup Language (HTML) components allows for composing the application layout using Python structures, rather than writing in HTML or using an HTML templating engine.Dash applications are made up of 2 building blocks : Layout and Callbacks.
    a. Layout describes the look and feel of the app, it defines the elements such as graphs, drop downs etc and the placement, size, color etc of these elements. Dash contains Dash HTML components using which we can create and style HTML content such as headings, paragraph, images etc using python. Elements such as graphs, drop downs, sliders are created using Dash Core components.
    b. Callbacks are used to bring interactivity to the dash applications. These are the functions using which, for example, we can define the activity that would happen on clicking a button or a drop down.

2. Plotly: Plotly is an open-source library that provides a list of chart types as well as tools with callbacks to make a dashboard.Plotly's Python graphing library allows for creating interactive and dynamic graphs. Dash is used to build the interface which is populated by the graphs that are created from and using Plotly.

3. Pandas: Pandas is a data analytics library in Python and commonly used for processing tabular data. Here we used Pandas to read the main data file, pre-process the unclean data, extract the necessary information and plot it using Plotly on Dash.

4. Flask is a small and lightweight Python web framework that provides useful tools and features that make creating web applications in Python easier.

5. numpy library is used for multi-dimensional array operations.

6. yfinance is a library that allows us to fetch financial data of a company (since its listing in the stock market) from its stock code directly.

7. gunicorn and lxml libraries will be used for the application's deployment i.e. to host the app on a target server.

8. sklearn and scikit-learn are tools used in the development of Machine Learning (ML) models.

9. dash_core_components: It is used for making beautiful charts, other visual components such as selectors, drop-downs, dates, and other visualizations.

10. dash_html_components: It includes various HTML tags for designing the dashboards. Dash HTML Components are similar to HTML syntax-es, so try relating the two syntax together.

11. dash-bootstrap-components is a library of Bootstrap components for Plotly Dash, that makes it easier to build consistently styled apps with complex, responsive layouts.

4.3. Visual Studio Code

Visual Studio Code is a streamlined code editor with support for development operations like debugging, task running, and version control. It aims to provide just the tools a developer needs for a quick code-build-debug cycle and leaves more complex workflows to fuller featured IDEs, such as Visual Studio IDE. It's available on platforms like Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node. VS Code uses experiments to try out new features or progressively roll them out. Our experimentation framework calls out to a Microsoft-owned service and is therefore disabled when telemetry is disabled. Code editing in Visual Studio Code Visual Studio Code is a lightweight but powerful source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages (such as C++, C#, Java, Python, PHP, Go) and runtimes (such as .NET and Unity).

4.3. Time-series graph

A time series chart, also called a times series graph or time series plot, is a data visualization tool that illustrates data points at successive intervals of time. Each point on the chart corresponds to both a time and a quantity that is being measured.

Generally, the horizontal axis of the chart or graph is used to plot increments of time and the vertical axis pinpoints values of the variable that is being measured. When the values are connected in chronological order by a straight line that creates a series of peaks and valleys, a time series chart may also be referred to as a fever chart.

A time series chart can be thought of as a series of snapshots that have been taken at regular intervals. The apples-to-apples comparison provided by this type of chart makes it an ideal tool for executive dashboards that help end users quickly identify a trend, spot an outlier in a cyclical pattern or analyze how a key metric is changing over time.

5. PROPOSED SYSTEM

A single-page web application using Dash (a python framework) and SVR(Support Vector Regression) machine learning model which will show company information (logo, registered name and description) and stock plots based on the stock code given by the user. Also the ML model will enable the user to get predicted stock prices for the date inputted by the user.

5.1. PROJECT STAGES

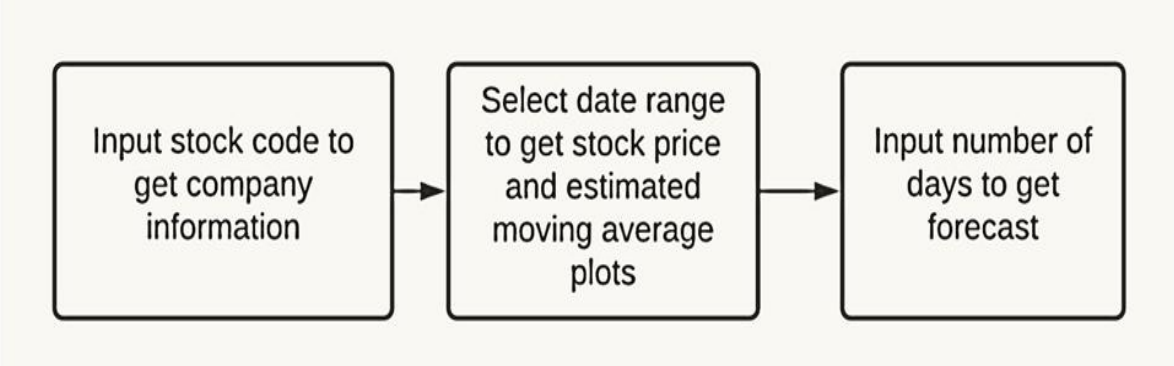

Fig: 5.1. Project Stages

5.2. APPROACH USED

- Install all the required python libraries and framework
- Create a basic website layout.The basic layout of the application will be built using Dash. Create the main website's structure using mainly Dash HTML Components and Dash Core Components.
- Enhance the site's UI by styling using CSS. Using CSS we will style our web page to make it look more neat and user friendly.
- Generate plots of data(company's information) using the plotly library of Python. By fetching the data using yfinance python library.We are going to use the yfinance python library to get company information (name, logo and description) and stock price history. Dash's callback functions will be used to trigger updates based on change in inputs.
- Implement a machine learning model(Support Vector Regression) to predict the stock price for the dates requested by the user. Using the support vector regression (SVR) module from the sklearn library. Fetch the stock prices for the last 60 days. Split the data-set into a 9:1 ratio for training and testing respectively. Use the rbf kernel in GridSearchCV for tuning your hyper parameters. Then, train the SVR model with the training dataset .
- Test your model's performance by using metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) on the testing dataset.
- Use the following command in the terminal of the working directory to run your Dash app's server locally - $python filename.py

5. RESULT ANALYSIS

1. Our Web App looks as shown below in Fig 5.1:

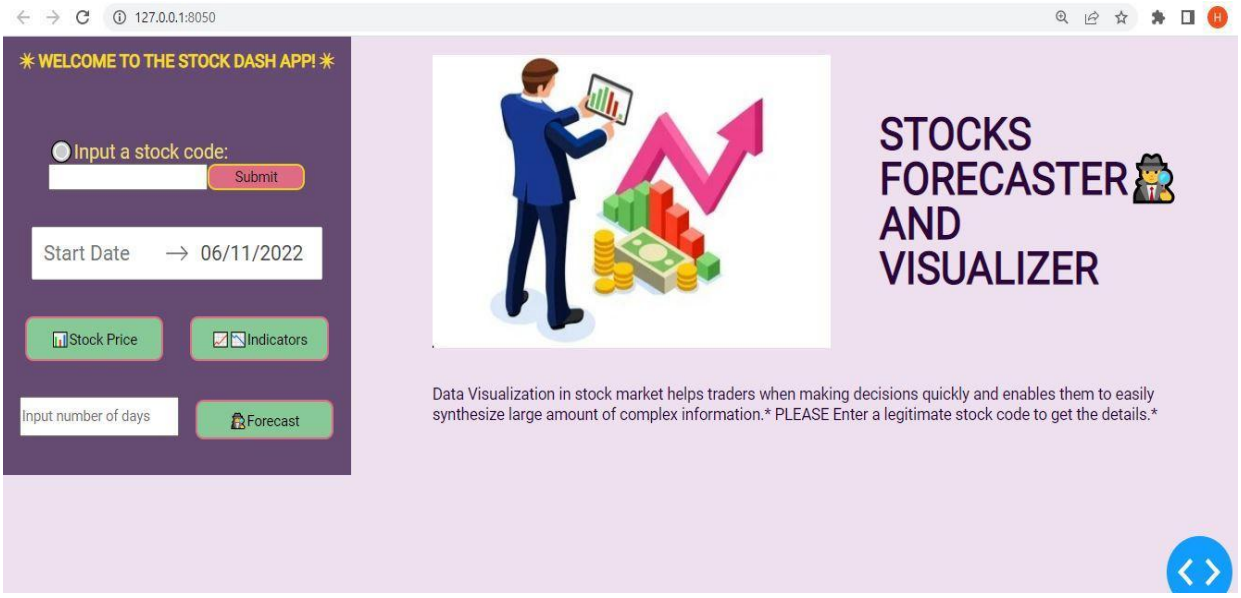

Fig 5.1

2. Input a stock code to get company's information as shown in fig 5.2.

Fig 5.2

3. Using the date range selector provide the start and end date to get the stock prices(closing and opening prices) of the company whose stock code you have provided earlier as shown in fig 5.3.

Fig 5.3

4. You can choose the Indicator button to get the indicators by providing the company stock code and date range as shown in fig 5.4.

Fig 5.4

5. You can predict the company stock prices(closed price) by providing the number of days and choosing the forecast button as shown in fig 5.4 you will get a time series plot between Date and Closed Price .

Fig 5.5

7. CONCLUSION

- In recent years, it has been noted that most people are investing in the stock market in order to make fast money. At the same time, an investor stands a good risk of losing all his or her money. Data visualization helps traders when making decisions quickly and enables them to easily synthesize large amounts of complex information

- Stock forecasting and visualization using dash is a machine learning project. We have built a single page web application using the plotly dash python framework and using some machine models to get company's information and plot graphs for visualization based on it. Stock market forecasting is done using a machine learning model(SVR).

- Dash provides a fast and quick way of creating beautiful and interactive dashboards without extensive knowledge of web development. It combines web development and analytics and provides a useful mechanism for serving dashboards for web response.

REFERENCES

- Dash HTML Components | Dash for Python Documentation | Plotly
- Dash Core Components | Dash for Python Documentation | Plotly
- CSS Tutorial
- A Complete Guide to Flexbox | CSS-Tricks
- Part 3. Basic Callbacks | Dash for Python Documentation | Plotly
- Build A Stock Prediction Program
- Python LSTM (Long Short-Term Memory Network) for Stock Predictions | DataCamp
- An introduction to Grid Search. This article will explain in simple… | by Krishni | DataDrivenInvestor
- Plotly Express
- Hands-On Guide To Using YFinance API In Python
- https://www.crio.do/projects/category/python-projects/
- Part 2. Layout | Dash for Python Documentation | Plotly