

Stock market prediction and analysis

A project Report by Pyfindor

Team Members:

Harsh Gawai
Shree Rama Kamal Kumar Vegu
Gautam Swaminath Ganesh

hr912357@dal.ca
sh907407@dal.ca
gt728571@dal.ca

Abstract

Stock market is an integral part of our society and a valuable source of income for the general populous. However investing in stocks is difficult for a beginner, as one must be familiar with the trends of the stocks as well as how the company is performing in order to invest correctly to earn profits. In this project we aim to support the people who aim at buying stocks, by analysing and effectively predicting the stock value at a given point in time, we also aim to effectively predict trends and patterns in the rise and fall of stock prices. Our aim is to predict the price of the stocks at the closing period of each day. Currently as future employees to the wide world, new investors such as students who have just graduated would fall prey very easily to the trap of failed investment. Our model aims to reduce the amount of such predictions and investments so that each investor is successful and also provides visual aid to the user. This provides us with risk reduction and a proper visualization of the data obtained.

Introduction

Before we as a group started with stock market prediction, we initially went with the topic of effects of natural disasters in Canada, we aimed to make a model that would categorize the severity of the natural hazards that are occurring in Canada and rank them from the most hazardous to the least. This was discontinued due to the lack of real time or live data, thus we were not able to continue the model and hence we shifted to stock market prediction as we found an abundance of data and a clear idea on the model's perspectives.

Stock Market, an introduction

Stock market is an investment opportunity where people buy/sell shares of publicly listed companies. It offers a platform to facilitate seamless exchange of shares. It is important to note that a person can trade in the stock market only through a registered intermediary known as a stock broker. The buying and selling of shares take place through electronic medium.

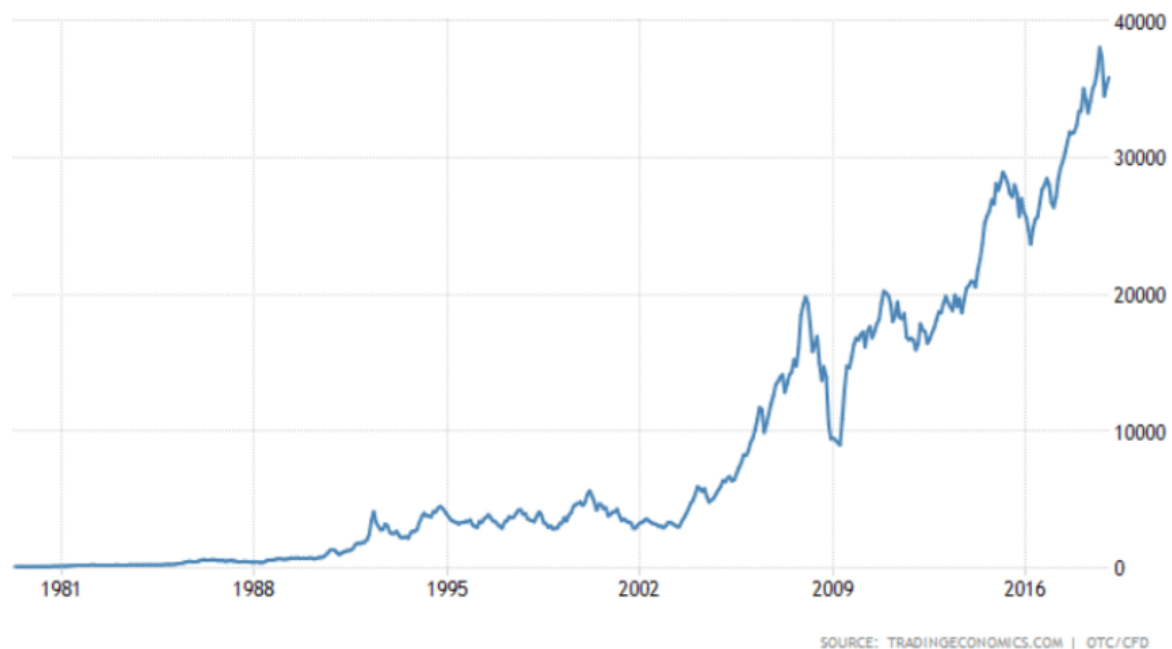
A main query that persists in the minds of many, is the doubt of why and what are stock markets and to whom do they serve as a beneficial investment.

Stock markets exist to serve the wider economy. It helps individuals earn a profit on their income when they invest in the stocks and allows firms to spread their risks and receive large rewards. It also enables the government to increase spending through the tax revenue they earn from corporations that trade on the stock exchange. The government uses the revenue to increase reinvestment and employment capacity. If stock markets did not exist, companies would have to resort to borrowing from the bank to raise money for expansion. This would be a burden on the company as they would have to repay the loans with interest.

Fortunately, with stock markets, businesses have the ability to create an initial public offering and raise large amounts of cash without having to worry about repayment. Moreover, publicly traded companies have no obligation to pay dividends when they incur losses.

Capital raised this way can help companies expand operations inorganically and create jobs in the economy. From a greater economic perspective, consumer spending increases, governments can benefit from tax revenues and there will be lower levels of unemployment.

This action of investment in companies helps not only in the improvement of a company but also to raise the number of jobs offered. Thus it helps people who make judicious investments in growing their financial assets in an exponential way.



The performance of the stock market is a rough indicator of how well the economy is performing. This often depends on speculators and perceptions of investors in the market. A rise or fall in the price of shares represents what cycle the economy is in such as a recession or a boom.

Aim of the project

The model aims to provide a visual aid to the user by providing graphs and trends of the stocks. The data is live stock exchange data that is collected realtime by alpha vantage, a software that allows us to easily collect data on company stocks throughout the world. The data is filtered by a stock exchange broker. Close to 6500 company stocks can be seen collected and the data is real time, thus allowing for an accurate prediction of the trends. The stock trends are visualized by the use of various graphs such as candle light graph which is a box plot that gives us what are the indications if profit or loss was made over a course of time. The graph is differentiated by two colours red which indicates loss and green for profit. An individual has made a profit if he invested in stocks which are in a rise.



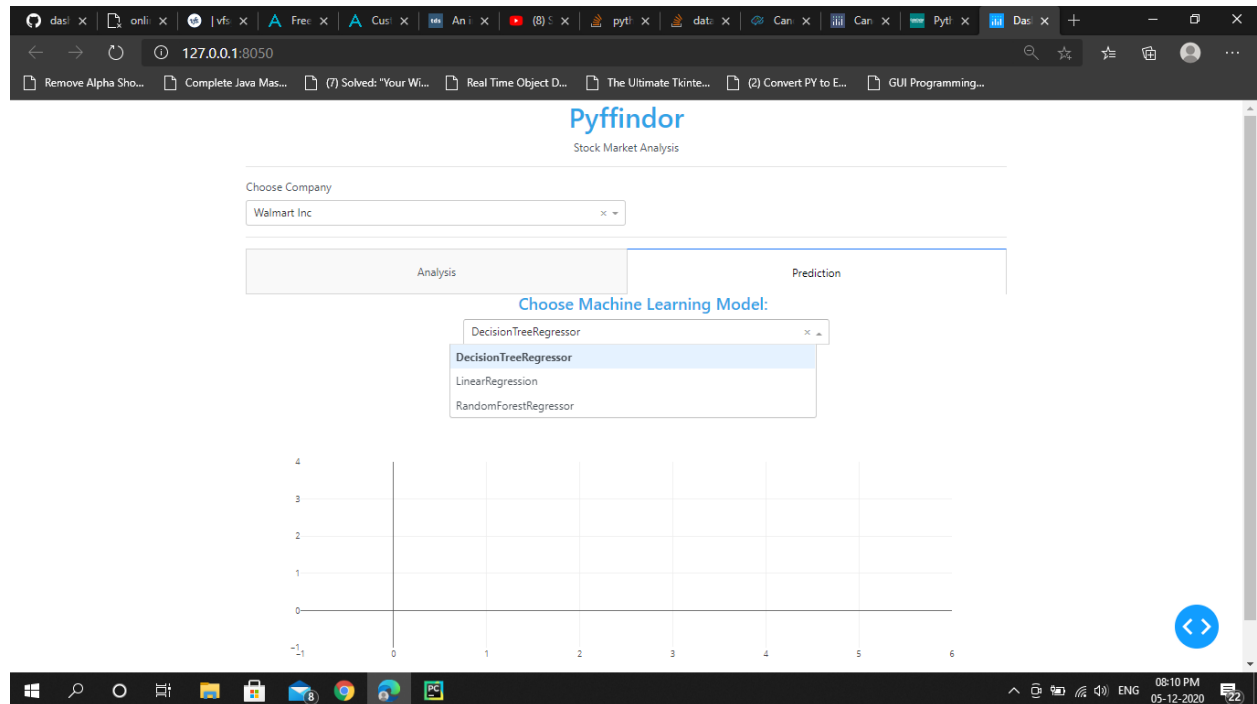
But the probability of the users to predict such trends are very low and are mostly incorrect which would lead to a failed investment. To prevent such failed investment the investor is to do substantial research in the chosen company. It is required for the user to check if the product made by each company is well received by the public and is sought for . One must also check public forums like twitter to see if the public image of said company is prospering before investing. This research is time consuming and is not convenient .

Solution and Novelty

To enable a more convenient and less time consuming approach the model provides an accurate prediction on the stock trends for the day . It also gives a consolidated survey about the company and twitter feeds regarding the said company as well. This will allow for a more convenient way for new investors to put their money in the right place to earn profits. The model then predicts the closing price of the stocks at the end of each day to allow for an efficient estimation of how the trend would proceed . The model is then evaluated by a test-train accuracy by which can estimate how close or accurate our prediction is .

Machine learning model used

In our model we have used machine learning models to explain and predict the stock exchange value. To name a few, we have used linear regression and Decision tree regressor.



Linear Regression

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:

- (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable?
- (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they—indicated by the magnitude and sign of the beta estimates—impact the outcome variable?

These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.

Three major uses for regression analysis are (1) determining the strength of predictors, (2) forecasting an effect, and (3) trend forecasting.

First, the regression might be used to identify the strength of the effect that the independent variable(s) have on a dependent variable.

Second, it can be used to forecast effects or impact of changes. That is, the regression analysis helps us to understand how much the dependent variable changes with a change in one or more independent variables.

Third, regression analysis predicts trends and future values. The regression analysis can be used to get point estimates. A typical question is, "what will the price of stock be in 6 days?"

We aim to use such regression methodologies in order to predict the stocks at certain stages of time. The data is split between test and train datasets, these are sent to a linear regression model which enables the prediction of data at a future date and time.

Decision tree regressor

A decision tree is a decision support tool that uses a tree-like **model** of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

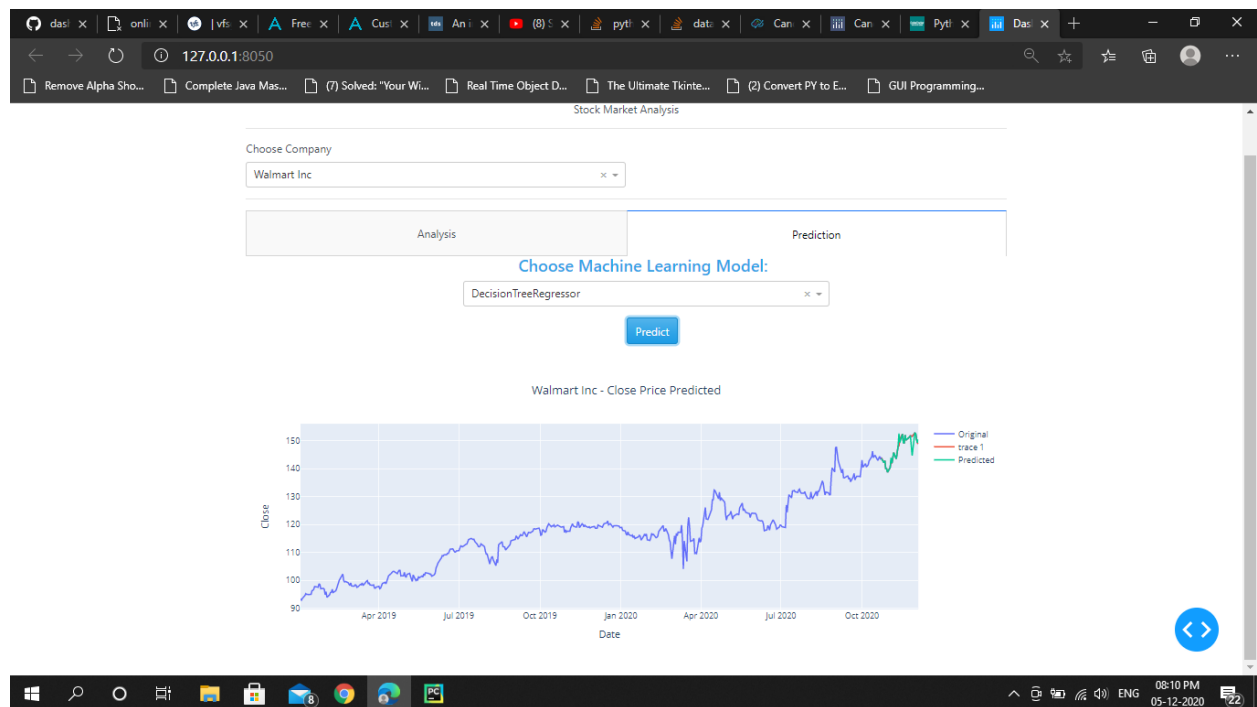
Decision trees build regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed.

The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy), each representing values for the attribute tested.

Leaf node (e.g., Hours Played) represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

The core algorithm for building decision trees called ID3 by J. R. Quinlan which employs a top-down, greedy search through the space of possible branches with no backtracking. The ID3 algorithm can be used to construct a decision tree for regression by replacing Information Gain with Standard Deviation Reduction.

A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). We use standard deviation to calculate the homogeneity of a numerical sample. If the numerical sample is completely homogeneous its standard deviation is zero.

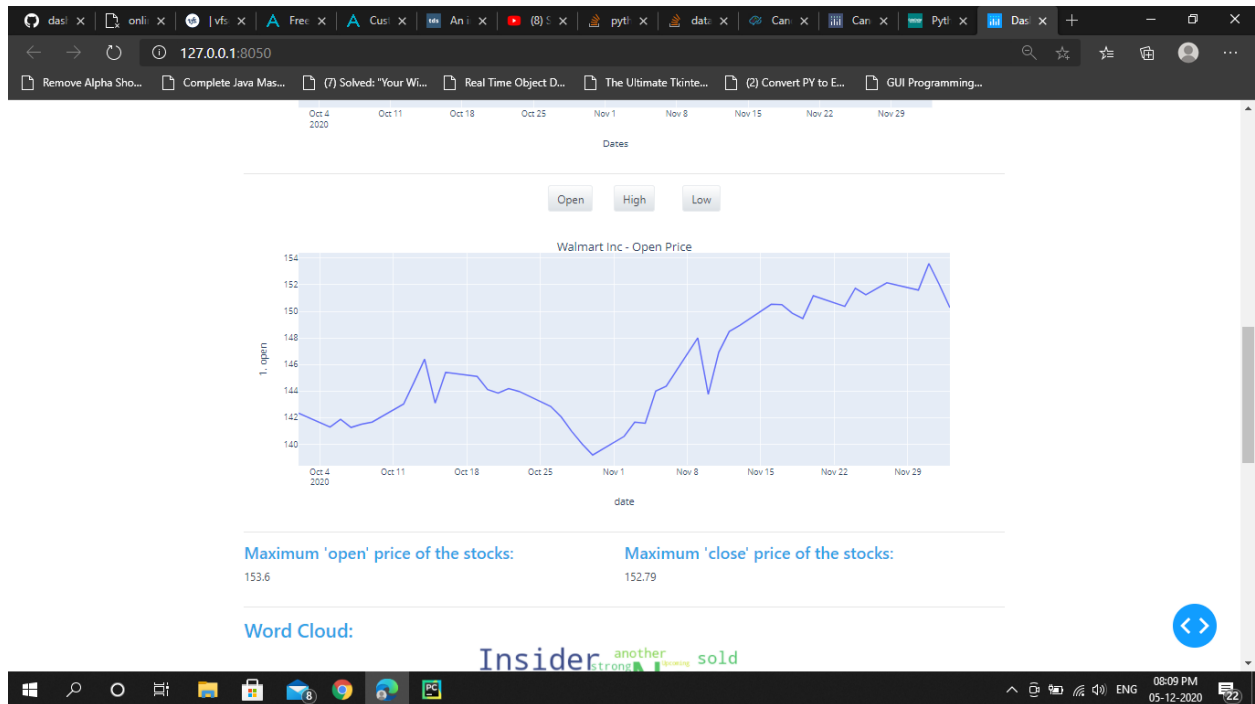


In our model again similar to the linear regression implementation, we apply a test-train accuracy model to predict the stock value at a future date or time. We then compare the accuracy that is derived from the linear regression and the decision tree in order to evaluate and compare both models.

These predictions are done completely independent of other company influences and thus are not susceptible to external influences. The predictions made are only an estimate on the rise and fall of stock prices in accordance with previous data.

visualization module

The main goal of this project was to make it easier for the user to purchase stocks at a very efficient and less time consuming manner. A simple solution to explain a concept is through images, this is where the visualisation techniques come to play.

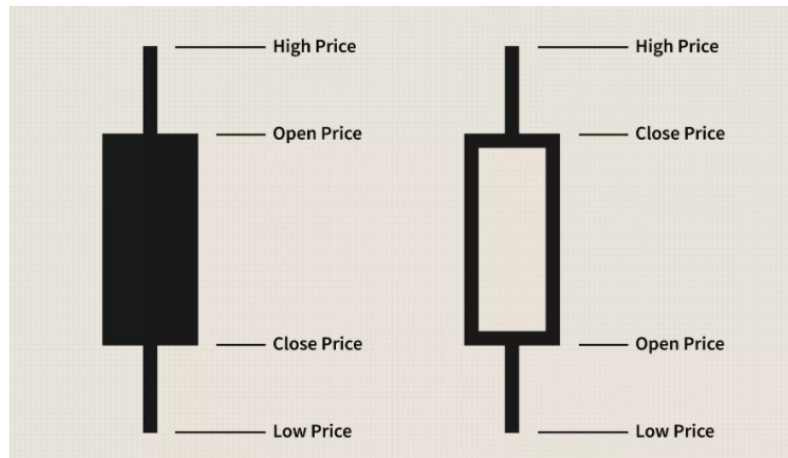


With the help of such visualisation techniques we aim to allow the user to experience a much easier time in reading the data and the trends among other companies. We aim to do this by providing a couple of graphs such as candle light graphs, a smoothing curve and a comparison graph to compare graphs of trends between companies.

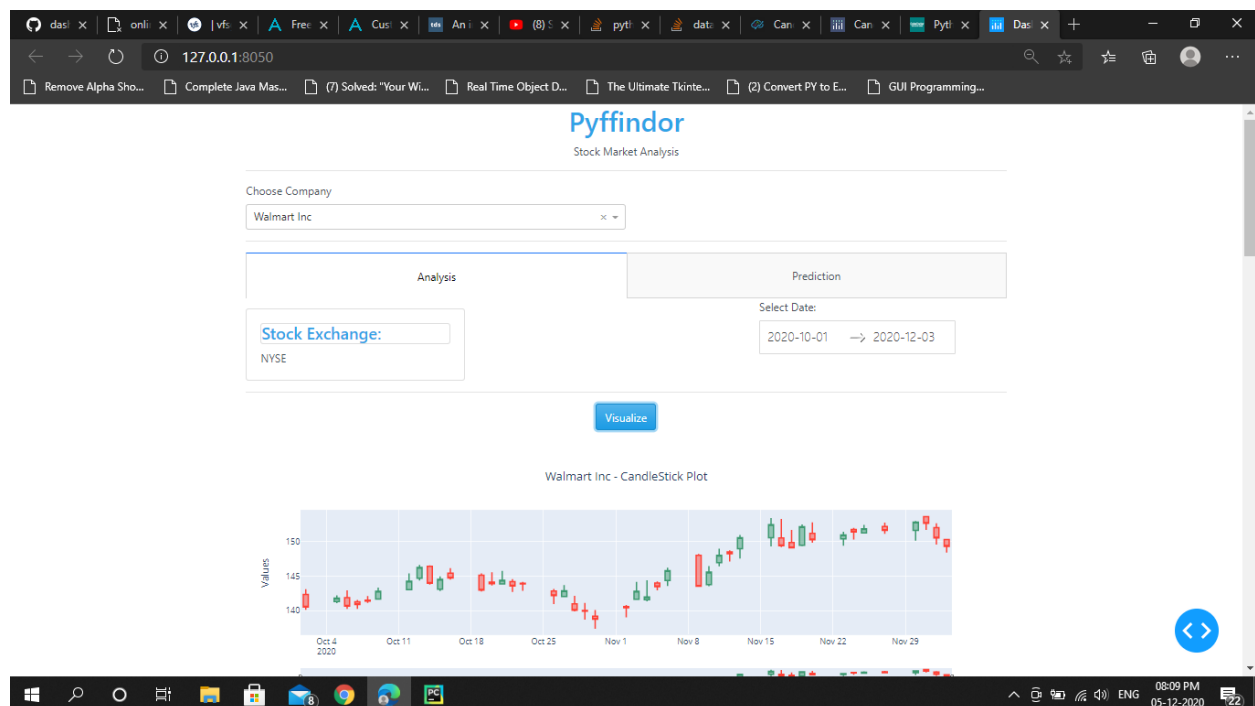
Candle Light Graph

In the candle light graph we can see the trends of stock prices through a period of time. It shows the profits or loss made on a daily basis. Just like a bar chart, a daily candlestick shows the market's open, high, low, and close price for the day. The candlestick has a wide part, which is called the "real body."

This real body represents the price range between the open and close of that day's trading. When the real body is filled in or black, it means the close was lower than the open. If the real body is empty, it means the close was higher than the open.



Traders can alter these colors in their trading platform. For example, a down candle is often shaded red instead of black, and up candles are often shaded green instead of white.



So why do we use such a visual representation over a simple bar chart?

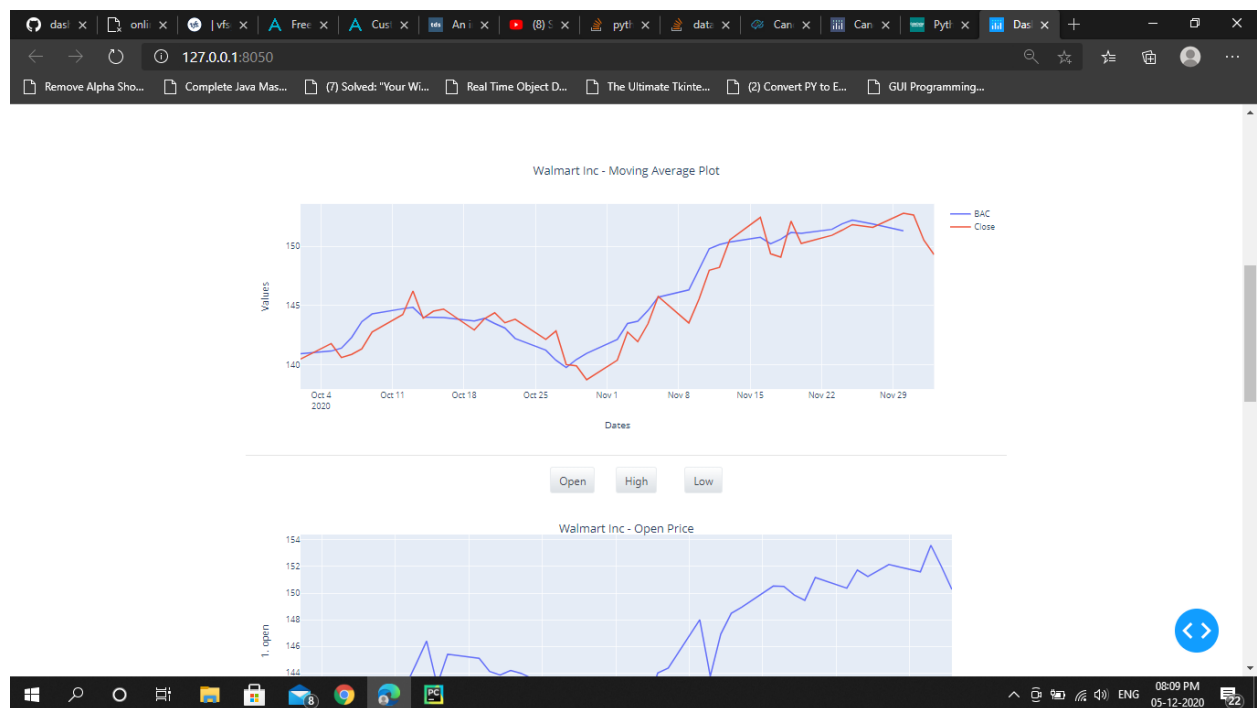
Just above and below the real body are the "shadows" or "wicks." The shadows show the high and low prices of that day's trading. If the upper shadow on a down candle is short, it indicates that the open that day was near the high of the day.

A short upper shadow on an up day dictates that the close was near the high. The relationship between the days open, high, low, and close determines the look of the daily candlestick. Real bodies can be long or short and black or white. Shadows can be long or short.

Bar charts and candlestick charts show the same information, just in a different way. Candlestick charts are more visual, due to the color coding of the price bars and thicker real bodies, which are better at highlighting the difference between the open and the close.

Comparative Graph

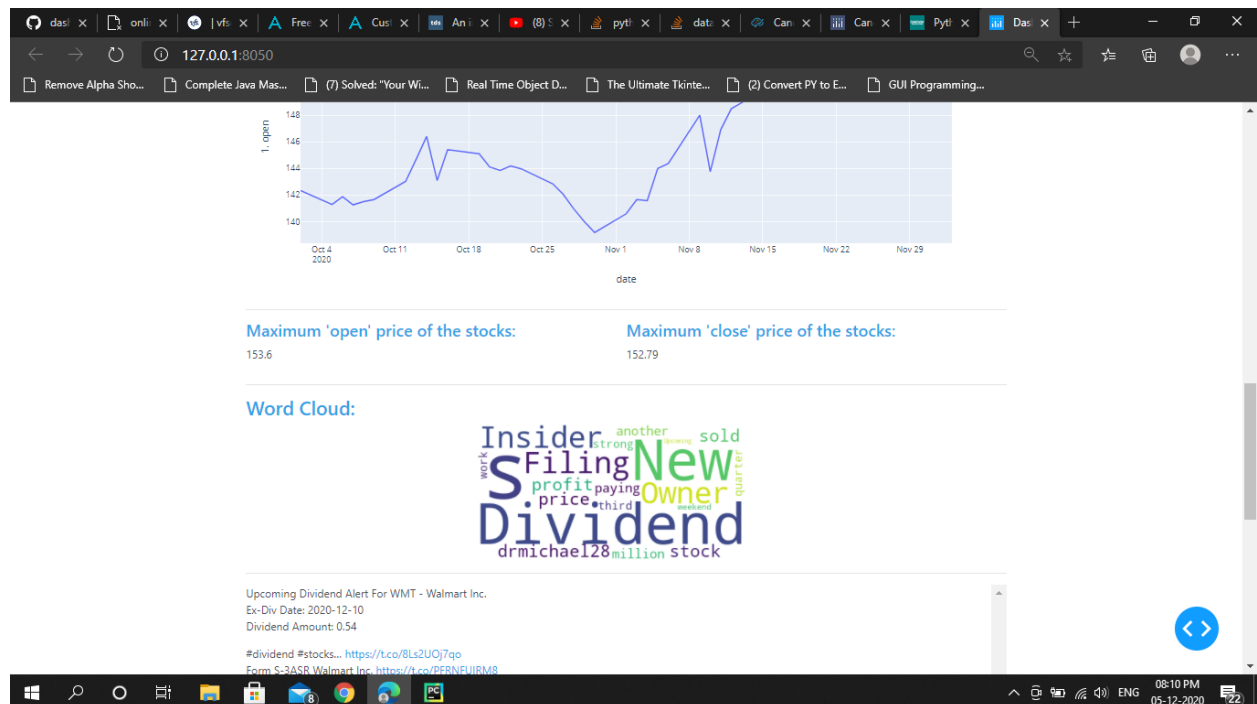
For any purchase one does not simply purchase the first article that he/she sees, It is common for comparison between objects to take place. Thus to offer a visual aid in the comparison of stocks of different companies, the model provides a comparative graph. The data is taken from the real time dataset and is printed. The user is given an option to choose between a number of companies and compare them against each other. This is presented in a graphical manner to provide a visual aid to the user.



This sort of representation gives the user some idea regarding a specific company's progress in their field against their competitors.

Company presence in social media

To know how good a company is performing, it could also be done if one could see twitter or any social media feed regarding the said company. The model provides a web mining of twitter with the company tags and averages them based on the number of views. This would allow the user to see the influence of the company on social media.

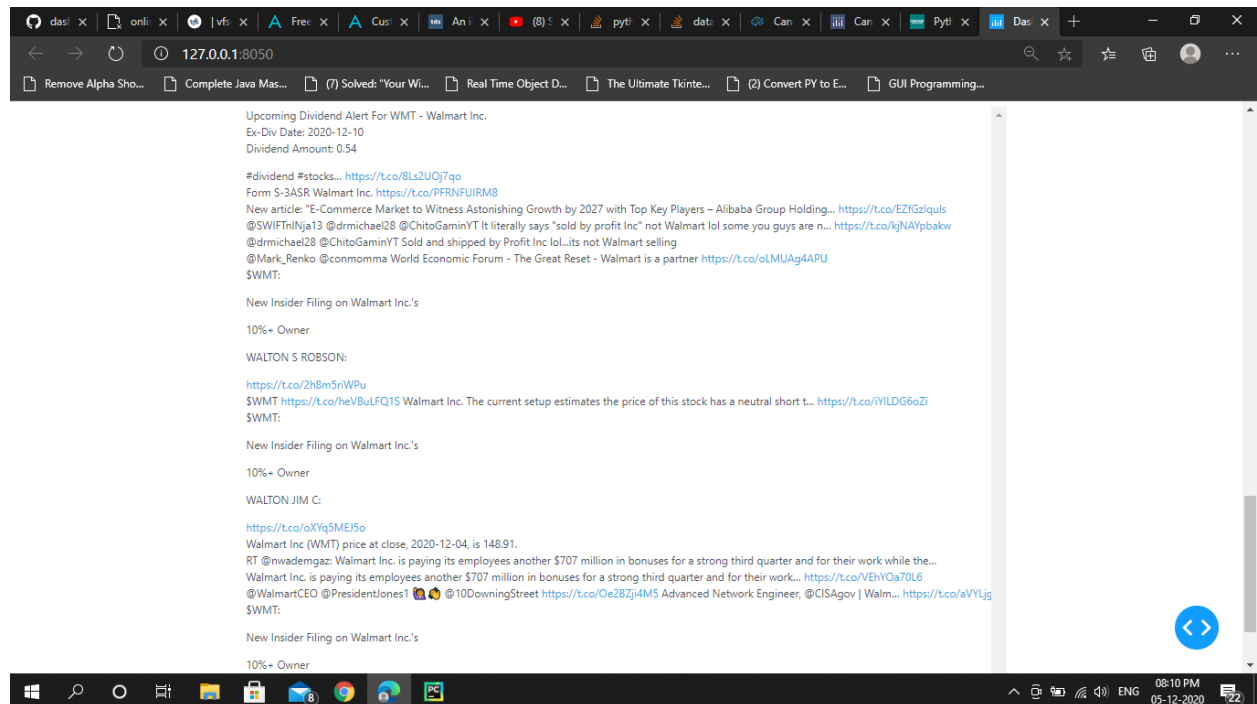


Sentiment analysis

Sentiment analysis is the process of using natural language processing, text analysis, and statistics to analyze customer sentiment. The best businesses understand the sentiment of their customers—what people are saying, how they're saying it, and what they mean. Customer sentiment can be found in tweets, comments, reviews, or other places where people mention your brand. Sentiment Analysis is the domain of understanding these emotions with software

One of the most well documented uses of sentiment analysis is to get a full 360 view of how your brand, product, or company is viewed by your customers and stakeholders. Widely available media, like product reviews and social, can reveal key insights about what your business is doing right or wrong.

Sentiment analysis is done using algorithms that use text analysis and natural language processing to classify words as either positive, negative, or neutral. The input to the algorithm is given in the form of the twitter feeds that is obtained in the previous step. From this we obtain the analysis on if the product is happily received and the company is doing well or the vice versa.



The user thus can obtain quality information on the company and can then further decide on his choices on investment based on the abundant results that the model provides. With the aid of the visual graphs representing the stock predictions and the sentimental analysis we can see if in the future the stock will be going down or up thus assuring the investor in making the right decision.

Evaluation metrics

Evaluating your machine learning algorithm is an essential part of any project. Your model may give you satisfying results when evaluated using a metric say accuracy_score but may give poor results when evaluated against other metrics such as logarithmic_loss or any other such metric.

Most of the time we use classification accuracy to measure the performance of our model, however it is not enough to truly judge our model.

We have also implemented a confusion matrix in order to verify and see the effectiveness of the product as a whole.

However it is important to note that stock prices are very volatile and are easily affected by external influences. A small change or a new product may offset the current trend and thus completely throw off the predictions.

These predictions that have been made are done on the premise, there are no external influences.

Tools, libraries

complete this

Conclusion

This project helped us to understand the fundamental value of the stock market and its inner workings. We were able to establish a safe, working and less time consuming model that allowed for the prediction of close period stock prices by the use of machine learning models. We were also able to compare the machine learning models and check for better results. Visualisation was provided in the form of several graphs that also had interactivity with the user to provide specific and detailed information on certain crucial data points. The candle light graph is used extensively to provide the user the overall data that he/she requires. It also answers the big question of whether the investment resulted in a profit or a loss. Text mining on relevant data corresponding to the companies is also given to the user to provide clear information on whether the company's product is leading to success or failure. This is also supported by sentimental analysis of the data mined. This would provide information if the general public is in good terms and the company product is well liked or not. The model is designed in a fashion to make the investors' job less taxing and less time consuming. All these factors help the user to make the right investment option on the company he/she desires.

Future Work

If further time and resources were available the project could be extended and a prediction for opening price at the stock market would be a good addition to the existing model. Comparison among a larger group of companies and better GUI for the users comfort would be a welcoming addition to the existing model.

