Predicting the Price of Amazon Stock

Since the start of the stock market, people have tried to predict the direction of stocks. We were able to create a model that can predict the price of Amazon stock.

Data

The data gathered for this project comes from several sources. All stock data comes from yahoo financial. Data concerning economic indicators were gathered from the United States Department of Treasury.

Method

The process for the project is to collect data, prepare for analysis, create different models, compare model results to determine the best model.

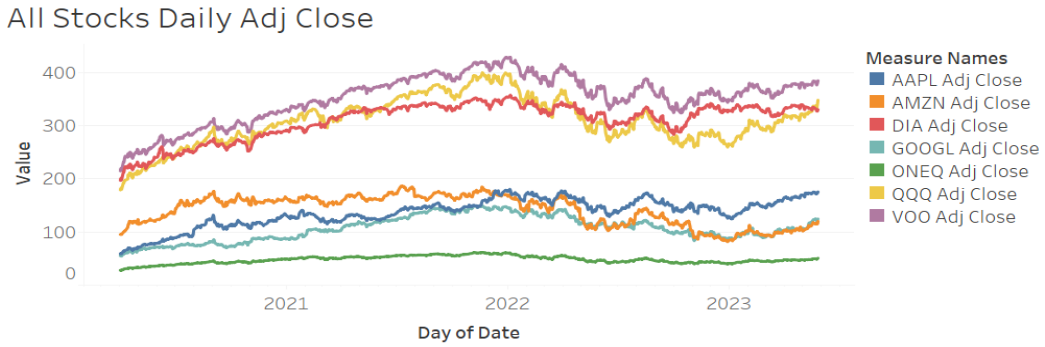
Data Cleaning

Overall, the data was clean. Any null values came from statistics that are not updated daily buy monthly quarterly etc… These stats were filled with the previous released value.

Also, some technical data require twenty values before it can compute. For this reason the first 20 days were deleted.

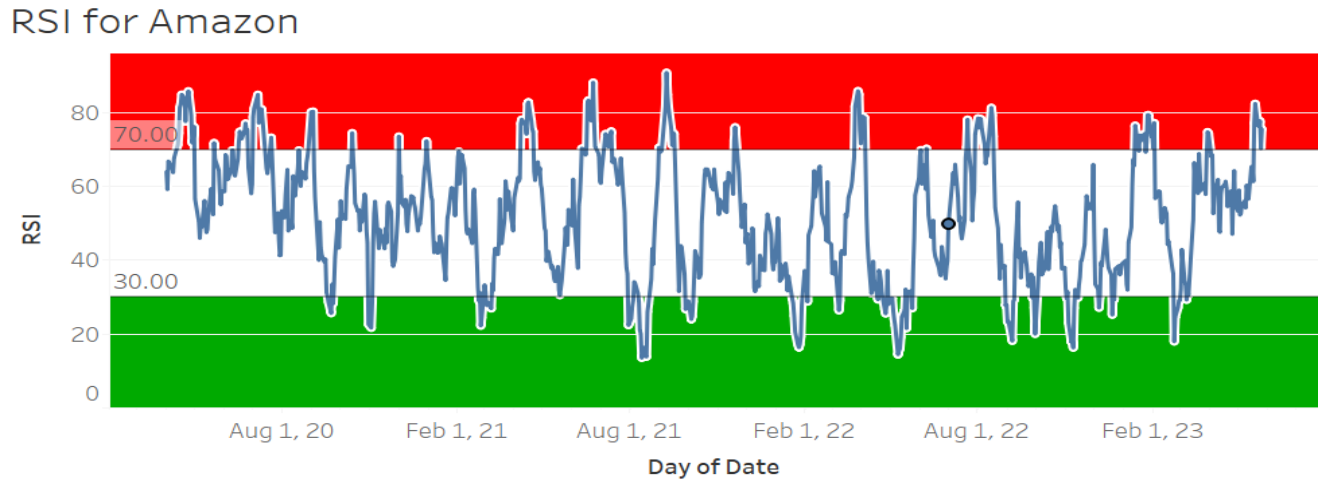
Exploratory Data Analysis

This is a graph showing the stock prices of different stocks used for data.

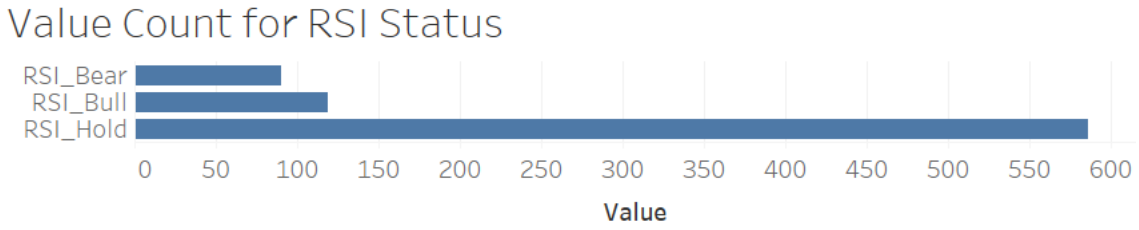


The trends of AAPL Adj Close, AMZN Adj Close, DIA Adj Close, GOOGL Adj Close, ONEQ Adj Close, QQQ Adj Close and VOO Adj Close for Day of Date. Color shows details about AAPL Adj Close, AMZN Adj Close, DIA Adj Close, GOOGL Adj Close, ONEQ Adj Close, QQQ Adj Close and VOO Adj Close.

RSI is a technical tool used to determine buy/sell based on momentum trading.



The trend of RSI for Day of Date.



RSI_Bear, RSI_Bull and RSI_Hold.

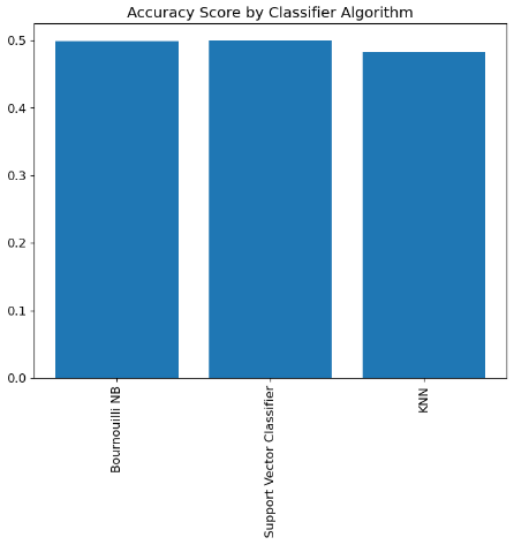
Algorithm and Machine Learning

I wanted to analyze the stock using both regressor and classifier algorithms. For regression I chose logistic regression and Random Forest Regression. For classification algorithms, the algorithms used were Bournoulli Naives Bayes, Support Vector Classifier and KNN. Different training sets were created to cater to the needs of the different types of models.

Initial Results

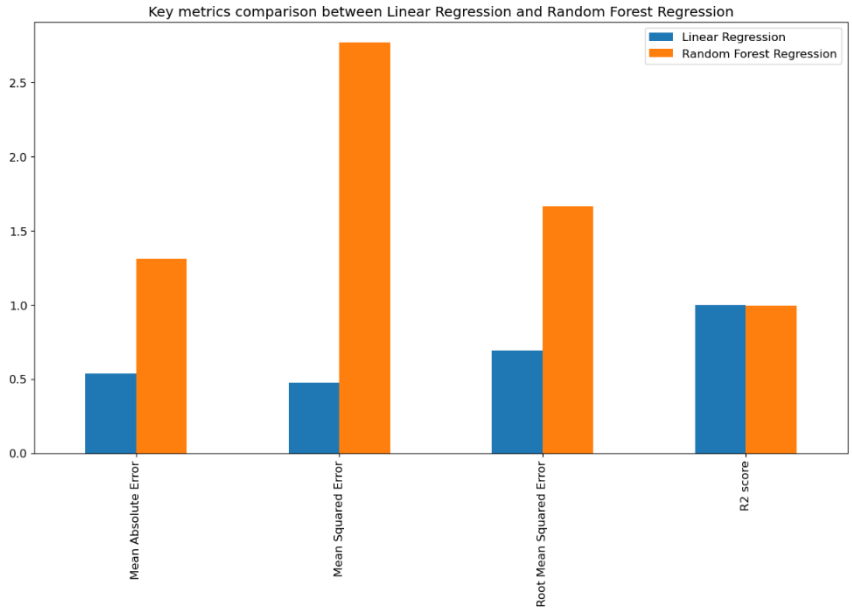
The initial results indicate that regression analysis returned better results than classification.

Classifier algorithms all similar roc-auc scores, they were all around 50%. coin flip.

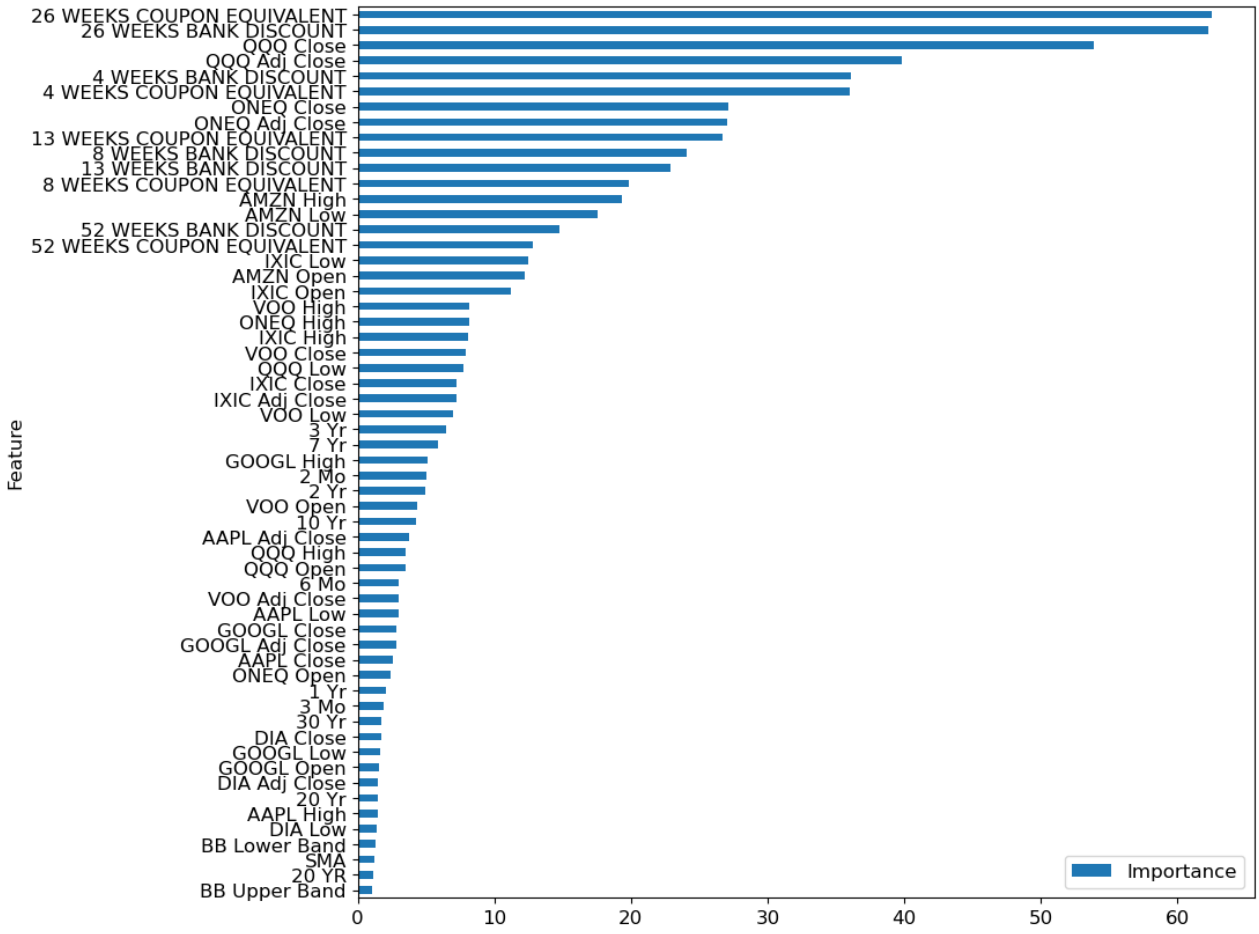


delivered unfortunately Essentially a

Regression algorithms delivered much better performance. Both Linear regression and Random Forest scored much better than the classifier algorithms. However, Linear Regression outperformed Random Forest. Here you can see how each regression algorithm performed on keen scoring metrics. A lower score on the first 3 indicate better performance.



This graph shows what features the linear regression found to be the most important to be able to predict the price of amazon stock.



Future Work

- Recommend creating new data sets for validation with data that is available since the last date of data.
- Use model to purchase stock on stock simulation websites to further test model before using live money.
- Buy and sell Amazon stock according to what the model indicates.