Microsoft Stock Price Prediction
Time Series Analysis
Data Source: kaggle.com Dataset

DSC680 – Applied Data Science
Project No. 2 - Milestone 3 – Final Paper
Winter 2021
Manish Kalkar

Final Paper

1. Abstract

Traditionally most machine learning (ML) models use as input features some observations (samples / examples) but there is no time dimension in the data. Time-series forecasting models are the models that are capable to predict future values based on previously observed values. Time-series forecasting is widely used for non-stationary data. Non-stationary data are called the data whose statistical properties e.g., the mean and standard deviation are not constant over time but instead, these metrics vary over time. These non-stationary input data (used as input to these models) are usually called time-series. Some examples of time-series include the temperature values over time, stock price over time, price of a house over time etc. So, the input is a signal (time-series) that is defined by observations taken sequentially in time.

Stock market prediction is a major challenge owing to non-stationary, blaring, and chaotic data, and thus, the prediction becomes challenging among the investors to invest the money for making profits. Several techniques are devised in the existing techniques to predict the stock market trends. Accurate stock price prediction is very challenging because of multiple (macro and micro) factors, such as politics, global economic conditions, unexpected events, a company's financial performance, and so on. But all of this also means that there's a lot of data to find patterns in.

2. Background

Stock market prediction and analysis are some of the most difficult jobs to complete. There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision. However, with the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data. We will perform the Exploratory Data Analysis (EDA) of the stock price of Microsoft Corporation (MSFT) as reported by the NASDAQ. It should be noted the Open, High, Low values of the stock are recorded on the same day, as well as the Closing Value at the end of the day, are all indicated for each date.

3. Business Problem

Predicting how the stock market will perform is very difficult. There are many influential factors involved in the prediction, for example psychological, rational, and irrational behavior etc. All these factors make stock prices very difficult to predict with a high accuracy. Stock market technical analysis includes reading the charts and using statistical figures to identify the trends in the stock market.

Our objective will be to perform Time Series Analysis by leveraging machine learning techniques, which have the potential to look into patterns of data to make accurate predictions. We will be using the dataset of Microsoft stock prices from April 2015 to April 2021 to build a model to predict the Microsoft closing stock price.

4. **The Dataset**

Source: kaggle.com

Dataset: https://www.kaggle.com/vijayvvenkitesh/microsoft-stock-time-series-analysis

The file Microsoft_Stock.csv contains 1511 records of the stock information of Microsoft from 04/01/2015 to 04/01/2021. This data was acquired in google sheets using the command 'GOOGLEFINANCE'.

5. **Exploratory Data Analysis**

In this dataset, the Close column contains the values whose future values we want to predict. So let's have a closer look at the historical close prices of Microsoft's stock price by loading the data into the pandas data frame, and plot the overall performance of the Microsoft Stock Price based on the Close price and comparing it to Open price, comparing high vs low prices and average monthly and yearly volume. We will also analyze the performance of the stock price by calculating daily percentage return over the period of time and between the specific time intervals and the risk value analysis at the end.

Feature Selection

List of relevant features available for the analysis:

1. Date - DateTime

2. Open - Opening Price - Decimal

3. High - Highest value of the day - Decimal

4. Low - Lowest value of the day - Decimal

5. Close - Closing price - Decimal

6. Volume - Number of shares traded on that day - Integers
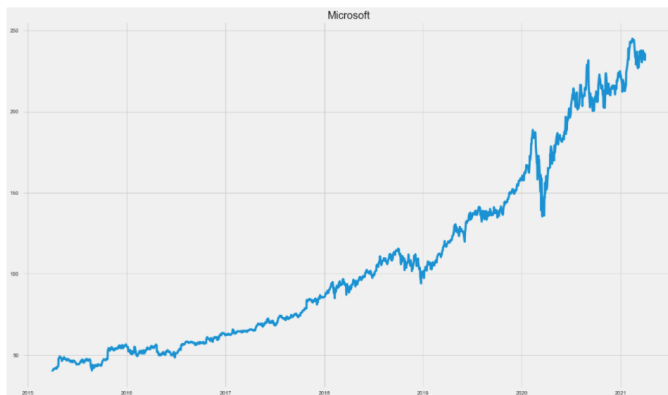
Feature Engineering

To calculate how much, you gained or lost per day for a stock, subtract the opening price from the closing price. Then, multiply the result by the number of shares you own in the company. The additional feature will be as below:

7. 1day % return

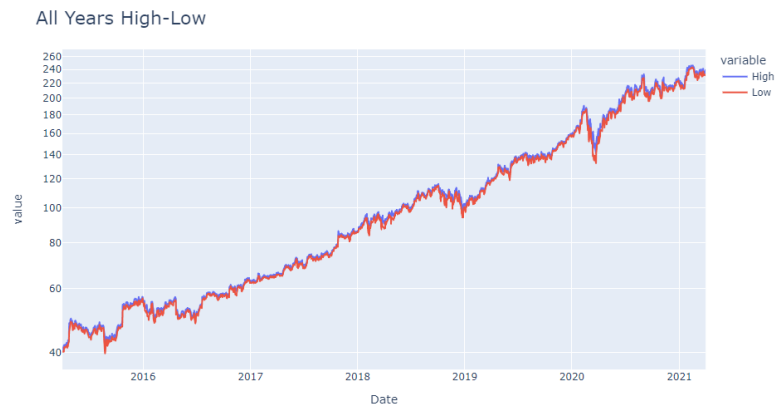**6. Graphical Analysis**

A. Microsoft Stock Price – Overall Performance
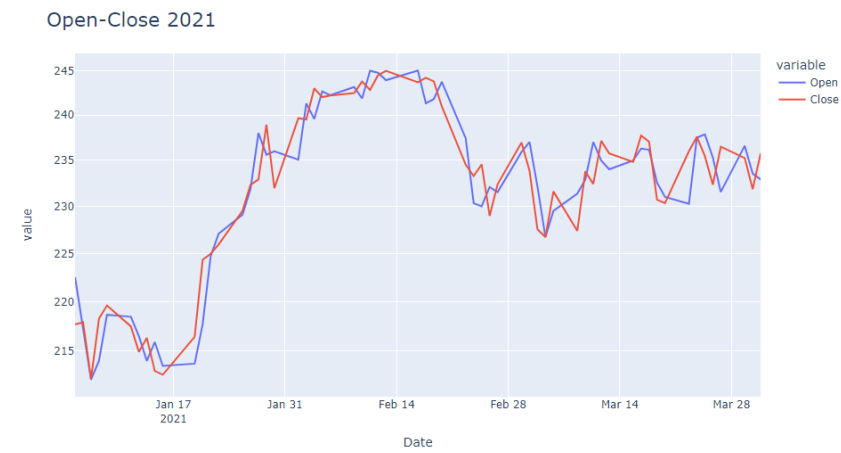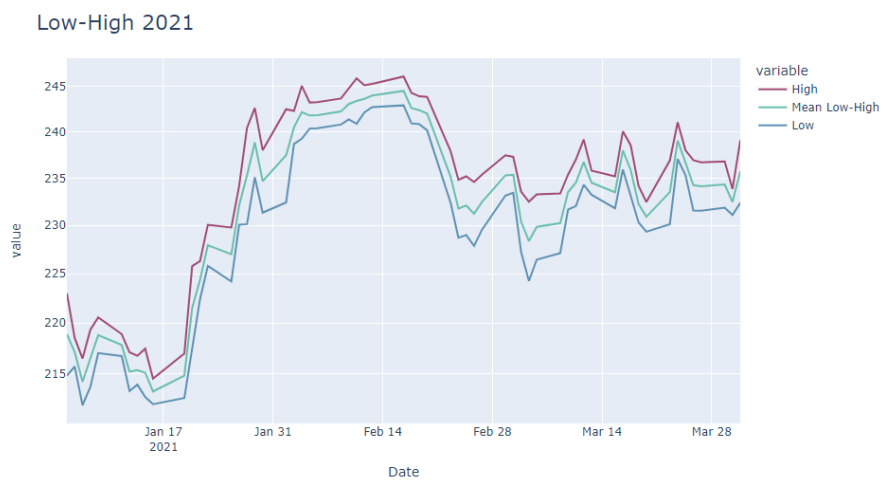
All Years stock price at Close

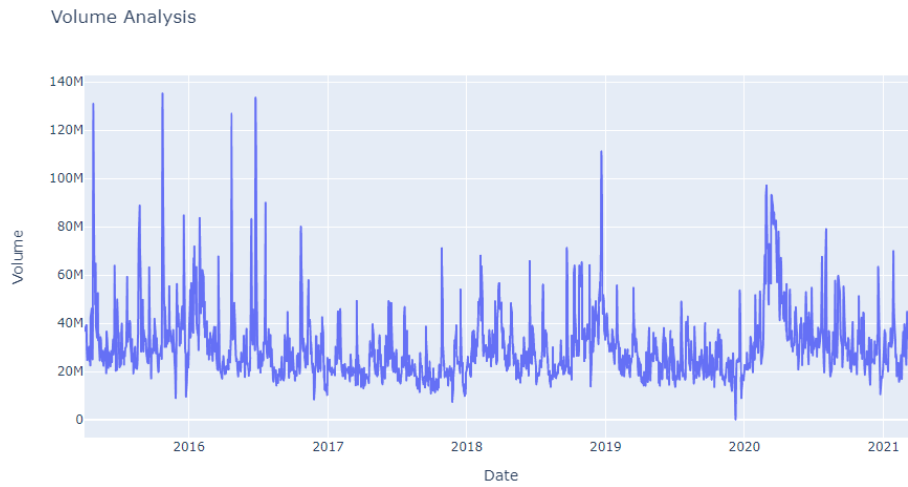
All Years Open vs Close

All Years High vs Low

All Years High-Low

2021 Open vs Close
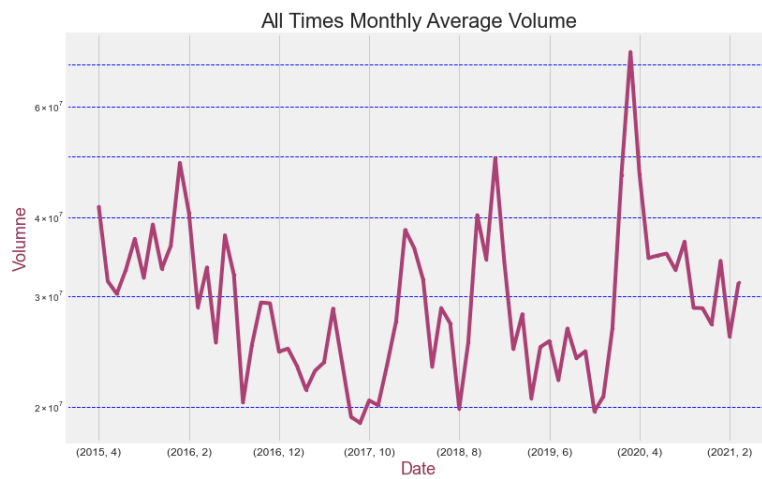
Open-Close 2021

2021 High vs Low
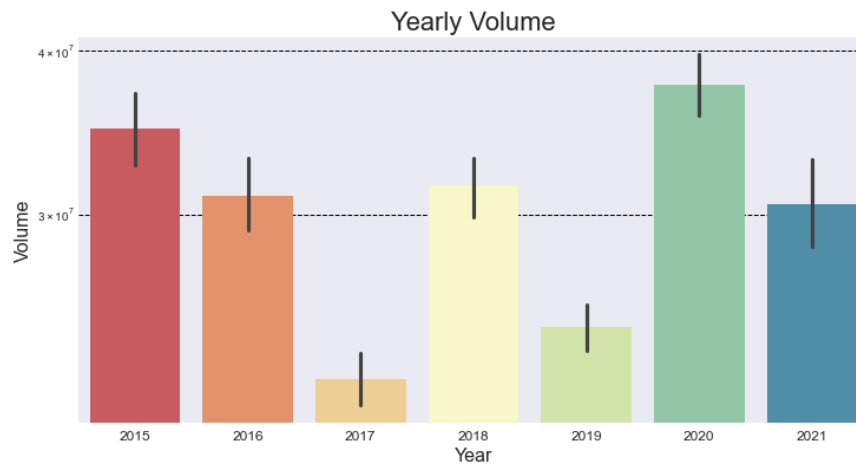
Low-High 2021

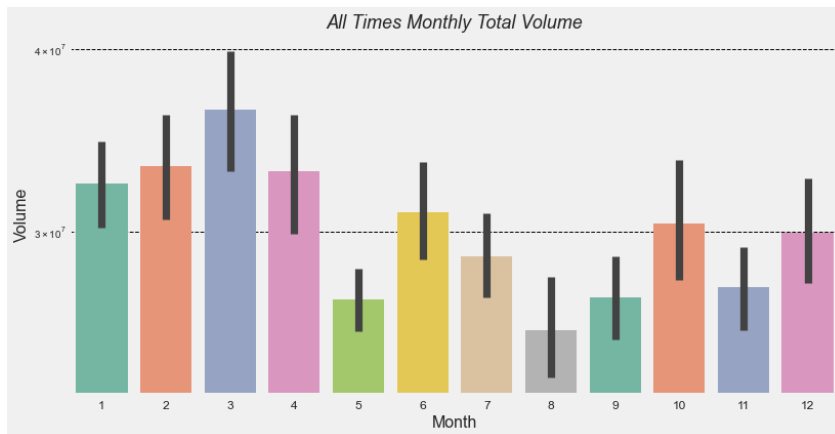## B. Microsoft Stock – Overall Volume
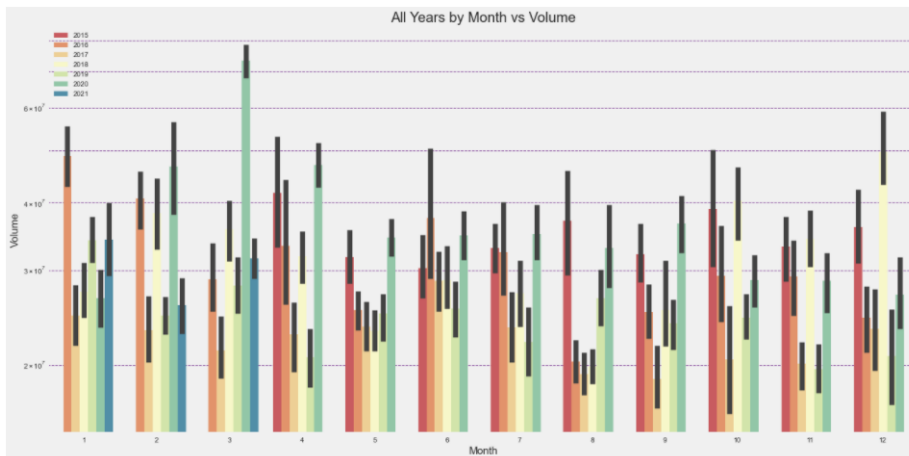

All Times Monthly Average Volume

Yearly Volume
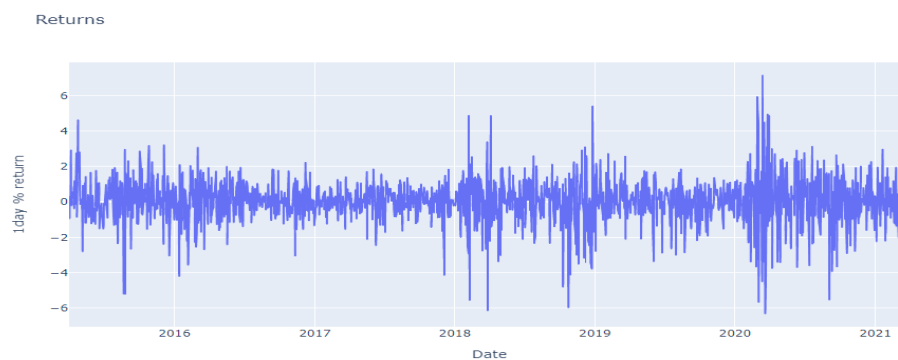
All Times Monthly Total Volume

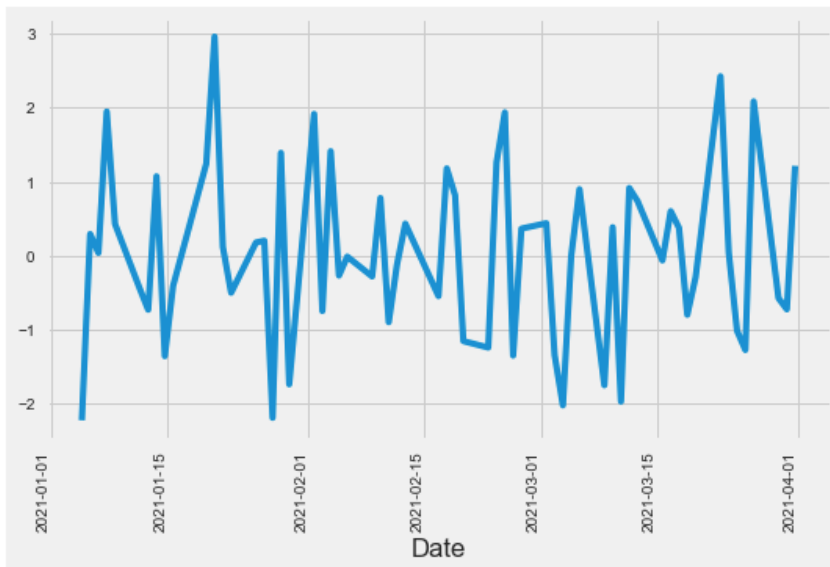**All Years by Month vs Volume**

C. Analyzing Daily price change in stock
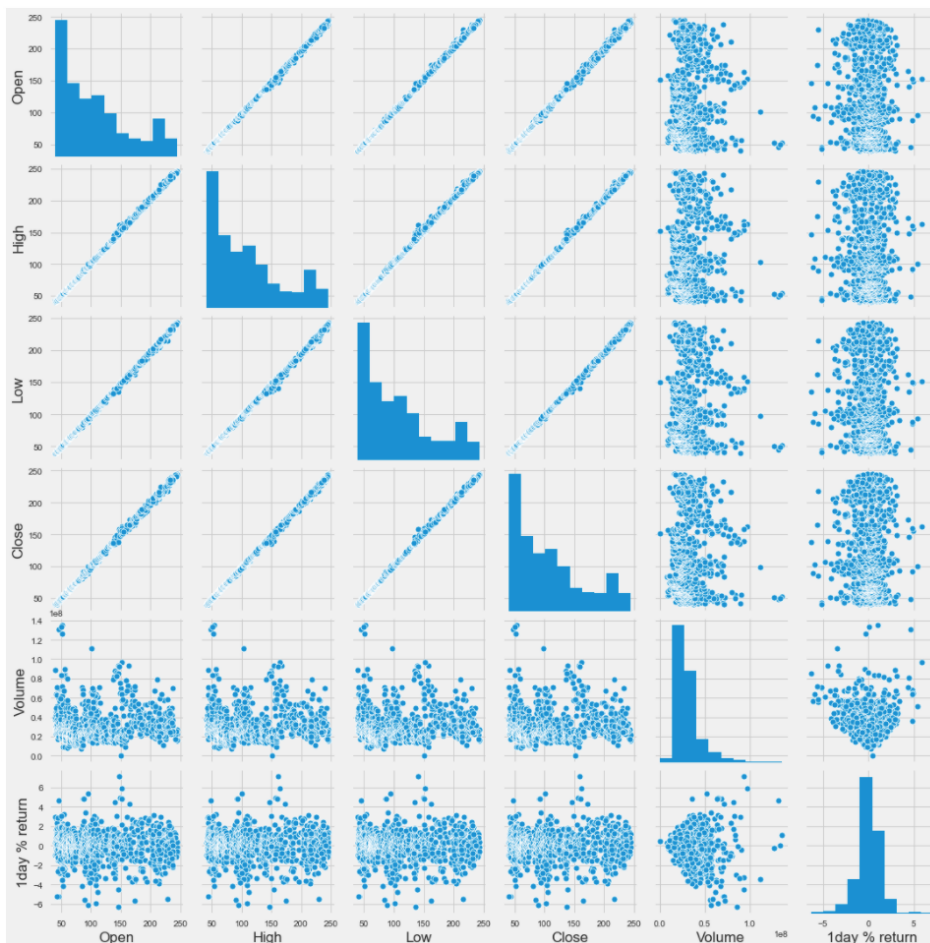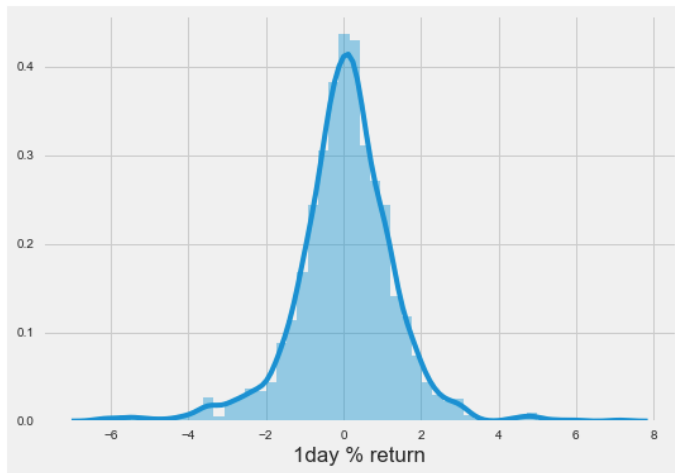
Percentage Returns

Percentage return between interval

Seaborn Pairplot

D. Value at risk analysis

6. Modeling

Model Selection

Long Short-Term Memory (LSTM)

We will use the Long Short-Term Memory (LSTM) method to create a Machine Learning model to forecast Microsoft Corporation stock values. They are used to make minor changes to the information by multiplying and adding. Long-term memory (LSTM) is a deep learning artificial recurrent neural network (RNN) architecture.
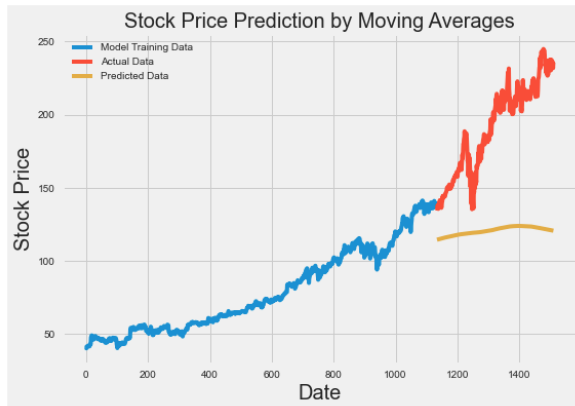
Decision Tree Regression Model

Use the Decision Tree Regression Algorithm to train the Microsoft Stock Price prediction model and have a look at the predicted stock prices for the next 5 days

Additionally, we will also develop algorithms for below listed models:
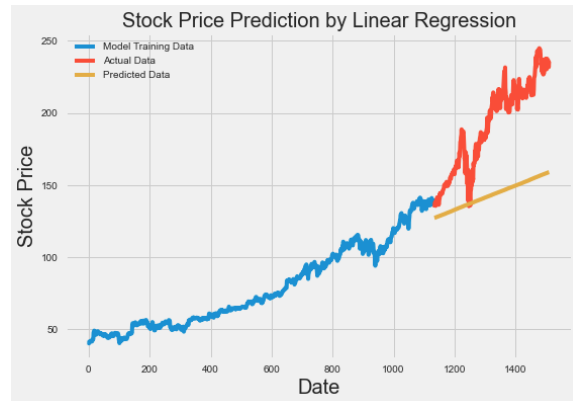
- Moving Average
- Linear Regression
- K-Nearest Neighbors
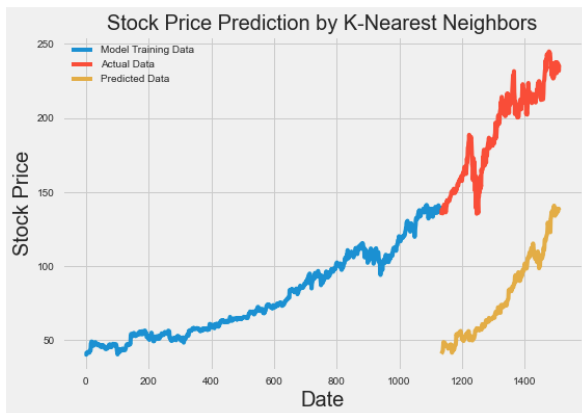- Auto ARIMA

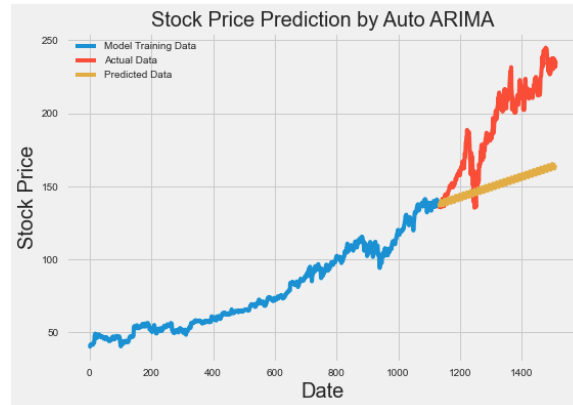# 7.   Model Evaluation – Results / Outcome

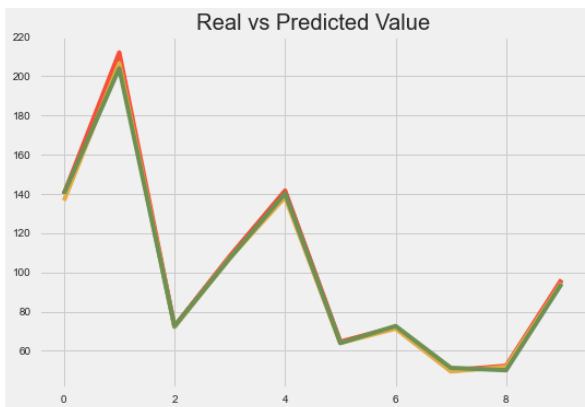## Moving Average

Linear Regression
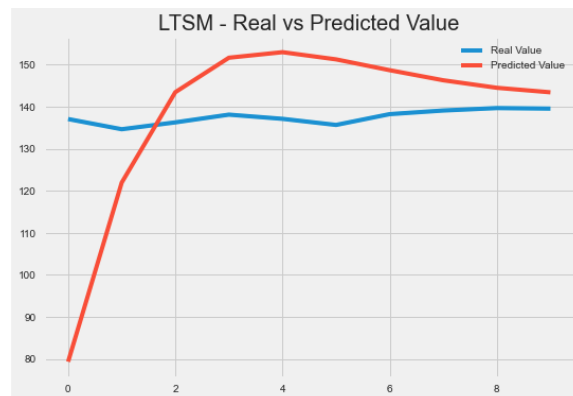

K-Nearest Neighbors

Auto ARIMA

Decision Tree Regression Model

Long Short-Term Memory (LSTM)

8. **Assumptions**

- Analysis, Data Preparation and building Machine Learning Algorithms are solely based on the historical data available as part of the dataset obtained from Kaggle data source.

- Two main approaches fundamental analysis, and technical analysis have been used to predict the future stock price with past data.

- The Stock Price follows a past trend rather than moving unevenly.

- Stock Price pattern is repetitive. In other words, the past stock price to predict future price trends.

9. **Limitations**

The question of the efficiency of Machine Learning algorithms is always under close scrutiny. The question is - Do ML models really work for the stock market? And if they do, how well do they actually perform?

During the analysis and prediction of Microsoft Stock Price here, below factors could not be taken into consideration:

- Stock Market Volatility

- Decreasing Accuracy

- High Competition

- Training Data Density

- Economic and Political influence

10. Conclusion

There are quite a few ways to design an ML model for stock trading using linear methods, such as moving average, linear regression, k-nearest neighbors, decision trees, etc.

However, given the complexity and multi-factorial dependencies of the problem of stock prices predictions, deep learning obviously fits the challenge better. Deep learning models like CNN, RNN, and especially LSTM outperform linear models because of the nature of neural networks. The inspiration for them came from how the human brain works. Neural networks are capable of transferring the analyzed data from one layer to the next one, thus making the training and analytical process much more efficient than in the case of linear ML models.

11. Challenges

Building an ML algorithm for the stock market has been a challenge that a lot of data scientists and ML engineers have pursued over the years. Empirical evidence suggests that such algorithms can be successful for automated stock trading.

This is naturally due to the highly detailed historical data from a variety of companies and stock exchanges. In itself, such data usually represents an already processed and pre-annotated dataset that can be used to train an ML algorithm. As the algorithm learns to see the hidden patterns in the historical data you feed into it, it can analyze them and offer predictions about how the stock prices will change in the nearest future.

The nature of the debate relies on the foundation of the axiomatic argument that the stock market is inherently unpredictable. The reason behind this is, on the one hand, the multiplicity of variables that contribute to the changing stock prices, starting with the economic, political,

and sociological factors and ending with environmental changes and natural disasters. On the other hand, there's also the irrational behavior of active agents, which makes it virtually impossible to predict the status of the stock market in the long term.

12. Implementation Plan

Below six steps were considered while implementing the effort end to end:

- Step 1: Define Problem Statement

- Step 2: Data Collection

- Step 3: Data Cleaning

- Step 4: Data Analysis and Exploration

- Step 5: Data Modelling

- Step 6: Optimization and Deployment

13. Ethical Considerations – Human Bias

Fraudulent activities such as customer fraud, false statements about a company's current situation or the manipulation of stock prices are detrimental to both, investors, and companies, that need equity. Investors lose trust, which leads to less investments in stocks. This again leads to negative consequences. How stock manipulation works, what it causes to the stock market and especially how it could be prevented. As a consequence, it is necessary to have a look on the ethical point of view.

14. References

- **Dataset:** https://www.kaggle.com/vijayvvenkitesh/microsoft-stock-time-series-analysis

- https://www.kaggle.com/aditirokade08/microsoft-time-series-analysis/notebook

- https://www.kaggle.com/paramarthasengupta/microsoft-stocks-price-prediction/notebook

- https://www.kaggle.com/miraculusik/microsoft-stock-eda-for-datetime

- https://www.kaggle.com/faelk8/microsoft-stocks-price-prediction-lstm/notebook

- https://www.grin.com/document/988695

- https://www.analyticsvidhya.com/blog/2021/10/machine-learning-for-stock-market-prediction-with-step-by-step-implementation/

- https://medium.com/analytics-vidhya/time-series-forecasting-predicting-microsoft-msft-stock-prices-using-arima-model-be1c45961739

- https://thecleverprogrammer.com/2021/06/21/microsoft-stock-price-prediction-with-machine-learning/

- https://www.analytixlabs.co.in/blog/stock-market-prediction-using-machine-learning/#sub5

- https://choiceindia.com/blog/technical-analysis-of-stock-trends/#:~:text=The%20last%20assumption%20of%20technical%20analysis%20states%20that,future%20price%20trends%20by%20using%20the%20chart%20patterns.