**DSBA 6211: Group Project Final Report**

on

**Stock Market Analysis for Tech Stocks (Microsoft)**

by

**Group 5**

Hridita Rubyat

Jahnavi Patnala

Koushik Meesala

Shreya Singh Joshi

Vishaal Sai Devisetty

1. **Project Summary**
   1. **Introduction**

Non-professional investors often try to find an interesting stock among those in an index. They need only one company, the best, and they don't want to fail (perform poorly). In practical terms, people like to find the best stocks to buy from an index and wait a few days hoping to get an increase in the price of this investment.

* 1. **Objective**

Analyzing historical stock data for Microsoft and predicting the short term (daily, weekly, fortnightly) future stock price or market trends.

We will compare the performance of different models/approaches for the prediction of stock price.

1. **Strategy**

We have used multiple techniques for the analysis and predictions of the Microsoft stock price using

1. **Non-Time Series modelling with Feature Engineering:**

* **Random Forest:** We added lag 1 differencing as a feature and used Open, High, Low, Volume and lag1 (representing trend) to run Random Forest to create a model similar to the linear model and square of trend model in time series

1. **Time Series Models:**

* Linear, Square, 3rd Power and 4th Power Trend
* SARIMA
* LSTM

1. **Exploratory Data Analysis**
2. **Trend and Seasonality: In** a time series, the trend component represents the long-term movement or pattern in the data. It captures the underlying direction or tendency of the data over time, ignoring short-term fluctuations, seasonality, and noise. Microsoft stock exhibits an overall increasing trend except for large dips in Covid period and the year 2023. There is no seasonality exhibited.

A graph with blue lines

Description automatically generated

Figure 1: Stock price from year 2019 to 2024

1. **Volatility:** Volatility is a measure of the variability or dispersion of returns, indicating the degree of fluctuation in the stock price. Microsoft stock exhibits high volatility or significant trends over short periods, such as two years as shown in Figure 2 below which shows daily deviation or fluctuation in the stock price.

A blue line graph with numbers and text

Description automatically generated

Figure 2: Standard Deviation of Daily Returns (Closing Price)

From the above plots we see that the sharp decline in the price trend and the big dip in the daily stock value in the year 2020 due to Covid-19 is an anomaly. Since a pandemic is a very rare scenario, we have chosen to focus on analyzing the stock price from the beginning of 2021 to 2024 as the daily stock price fluctuation shows stationarity and no seasonality.

1. **Models:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S. No.** | **Independent Features** | **Independent Features/ Parameters** | **MAE** | **RMSE** | **Others** |
| Non Time Series | Random Forest | Open, High, Low, Volume, Lag1 | 9.76 | 12.33 |  |
| Time Series | Linear model | Trend | 76.67 | 77.04 | F-statistic = 378.2  P(F-stat) = 5.70e-69  AIC = 7852  BIC = 7862  Adjusted R-sq = 0.326 |
| Trend Square | 41.17 | 41.67 | F-statistic = 289.9  P(F-stat) = 8.21e-95  AIC = 7728  BIC = 7742  Adjusted R-sq = 0.427 |
| Trend ^ 3 | 21.5 | 22.13 | F-statistic = 379.0  P(F-stat) = 1.19e-115  AIC = 7632  BIC = 7646  Adjusted R-sq = 0.492 |
| Trend ^ 4 | 5.99 | 6.95 | F-statistic = 454.4  P(F-stat) = 1.85e-131  AIC = 7559  BIC = 7573  Adjusted R-sq = 0.538 |
| SARIMA | p=1, d=1, q=1  Seasonal Order = (0,0,0,0) | 9.17 | 10.98 |  |
| LSTM |  | 6.50 | 7.98 |  |

1. **Conclusion**

We have implemented multiple models to compare their performance to predict time-series stocks data. The findings so far with the models completed are:

1. Random Forest or ensemble models did not perform well with the data.
2. LSTM results show low mean square errors and predicted values following the actual values very closely.
3. **Next Steps**

We are working on implementing traditional time series analysis on this data and evaluating the performance. We will next use moving averages, trend and seasonality to run regression time series analysis. We will also try to explore and run Arima model for analyzing this time series.

**Figures:**

1. **LSTM (Completed):**

**Target Variable:** Close (closing price for the given day)

**Goal:** To predict the closing price for the given day looking at past 60 days)

We scaled the data and used a sequence length of 60 i.e., used 60 days data to process at once. The results are as below:

A close up of a message

Description automatically generated

**A graph showing a green line

Description automatically generated**