TECHNOCOLABS DATA SCIENCE INTERNSHIP
PROJECT REPORT

MACHINE LEARNING IN INTRA DAY STOCK

TRADING (SVM POLY ,LINEAR based model)

TEAM MEMBERS

Rahul

Abhishek

Swetha

Anukriti

Navya

Yuvraj

Gaurav

1. AIM:

The aim of our project is to build a Machine Learning model which aims to predict whether the price in the next minute will go up or down, using the time series data of stock price, technical-analysis indicators, and trading volume in the Stock market.

2. ABSTRACT:

The ability to precisely predict the price movement of stocks is the key to profitability in trading. In light of the increasing availability of financial data, prediction of price movement in the financial market with machine learning has become a topic of interest for both investors and researchers..

3. INTRODUCTION:

Stock price forecasting is a popular and important topic in financial and academic studies. Share market is a volatile place for predicting since there are no significant rules to estimate or predict the price of a share in the share market. Many methods like technical analysis, fundamental analysis, time series analysis and statistical analysis etc. are used to predict the price in tie share market but none of these methods are proved as a consistently acceptable prediction tool. In this project we have implemented SVM Regression(LINEAR, POLY) and trained the model.

The LINEAR, POLY is like a long short-term memory (SVM) with a forget gate, but has fewer parameters than SVM, as it lacks an output gate. LINEAR, POLY aims to solve the vanishing gradient problem which comes with a standard scaler. Model deployment using streamlit and flask. Streamlit is an open-source Python library that makes it easy to create and share beautiful, custom web apps for machine learning and data science. Flask is a popular Python web framework, meaning it is a third-party Python library used for developing web applications.

4. OVERVIEW:

- Exploratory data analysis
- Data Cleaning and Data Visualization
- Training the model using SVM
- Deployment using streamlit and flask

5. DATASETS:

As we are provided with two Datasets NIFTY and MSFT. Which holds the stock market data in sometimes respectively. SPDR S&P 500 trust (NYSE: SPY) 1-minute interval data from March 1st until May 24th 2019.

6. SOURCE: IEX Trading.

6.1 NIFTY Dataset:

A brief explanation of every column in the dataset is as follows:

- Date - date of observation
- Time- time of observation
- Open - open value of the index on that day
- High - highest value of the index on that day
- Low - lowest value of the index on that day
- Close - closing value of the index on that day

6.2 MSFT Dataset:

A brief explanation of every column in the dataset is as follows:

• <u>Date</u> - date of observation

• <u>Open</u> - open value of the index on that day

• <u>High</u> - highest value of the index on that day

• <u>Low</u> - lowest value of the index on that day

• <u>Close</u> - closing value of the index on that day

• <u>Adj Close</u> - adjusted closing price amends a stock's closing price to reflect that stock's value after accounting for any corporate actions. The closing price is the raw price, which is just the cash value of the last transacted price before the market closes.

• <u>Volume</u> - volume of transaction

7. EXPLORATORY DATA ANALYSIS:

Exploratory Data Analysis (**EDA**) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to:

1. maximize insight into a data set;
2. uncover underlying structure;
3. extract important variables;
4. detect outliers and anomalies;
5. test underlying assumptions;
6. develop parsimonious models; and
7. determine optimal factor settings.

The particular graphical techniques employed in EDA are often quite simple, consisting of various techniques of:

• Plotting the raw data (such as data traces, histograms, bi-histograms, probability plots, lag plots, block plots.

• Plotting simple statistics such as mean plots, standard deviation plots, box plots, and main effects plots of the raw data.

• Positioning such plots so as to maximize our natural pattern-recognition abilities, such as using multiple plots per page.

8. NIFTY DATASET:

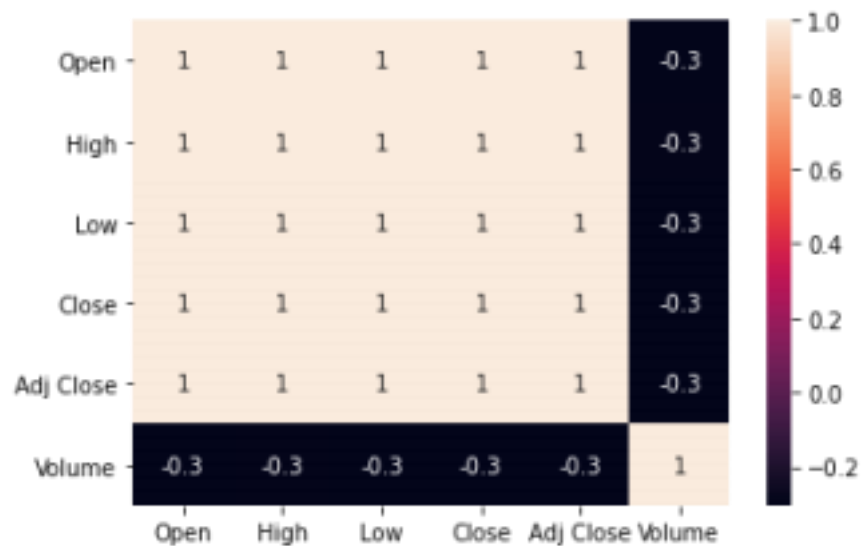1. Graphical representation of data that uses a system of color-coding to represent different values.
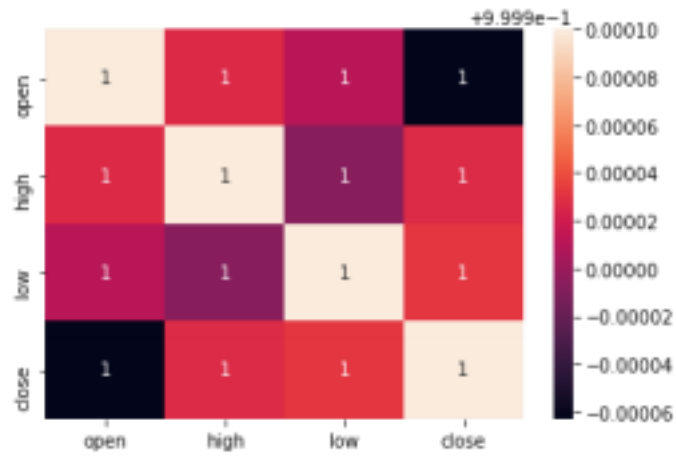
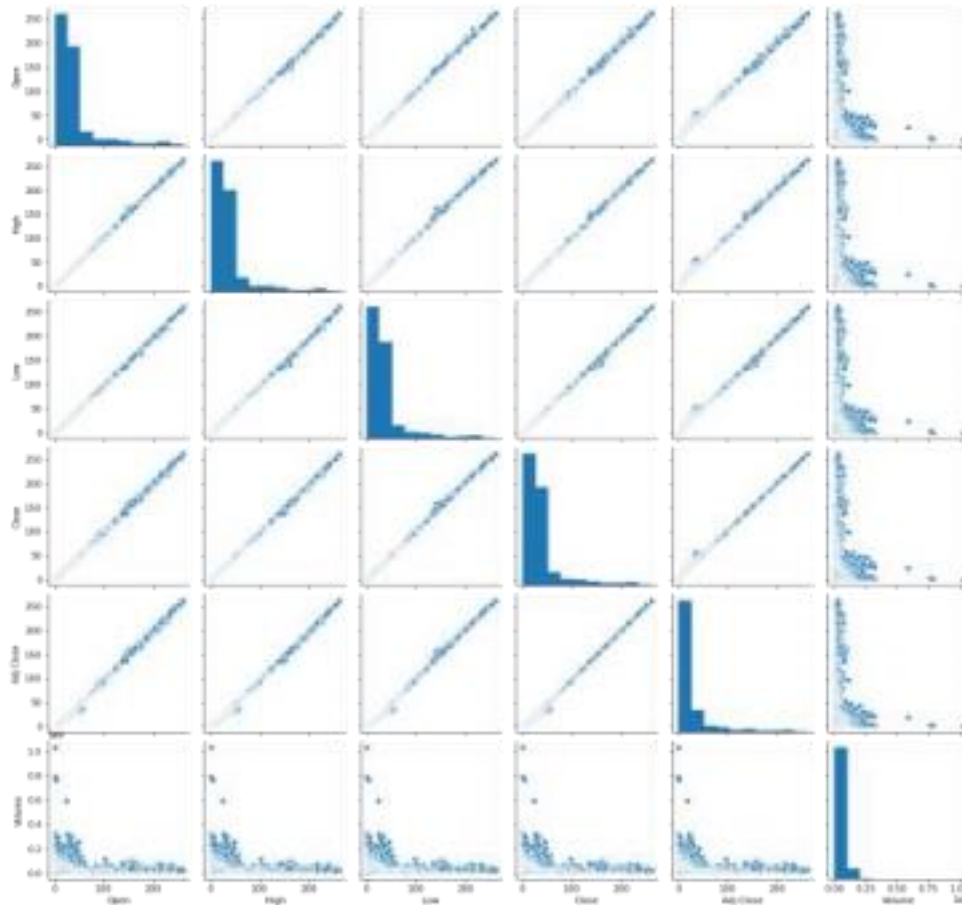2. Representation of pair wise relationships in a dataset.
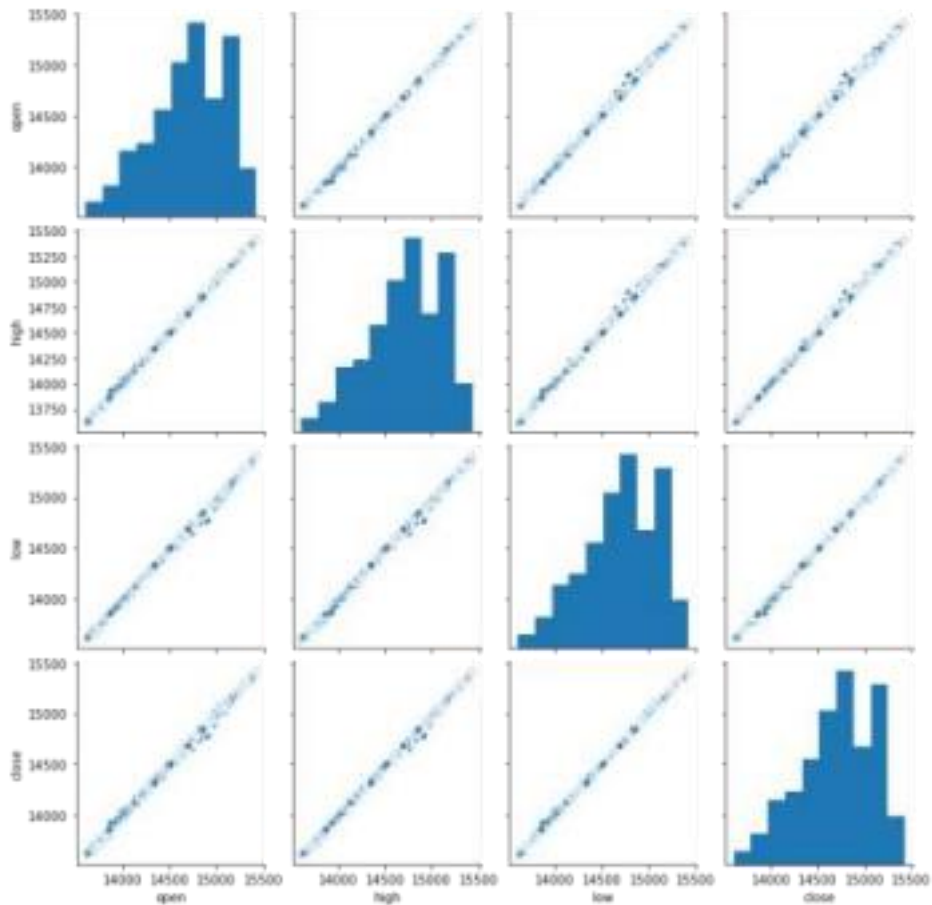
9. MSFT DATASET:

1. Graphical representation of data that uses a system of color-coding to represent different values.

2. Representation of pair wise relationships in a dataset.

10.TRAINING THE MODEL USING SVM:

Support Vector Machine: SVC with Linear, Polynomial (degree 3), Sigmoid, and Radial Basis Function kernel. The variables are tuned by adjusting the cost of constraint violation as well as the constant of regularization term in Lagrange formulation

10.1 NIFTY DATA:

1. Model Loss is a parsimonious mathematical description of the behaviour of a collection of risks constituting an insurance portfolio.
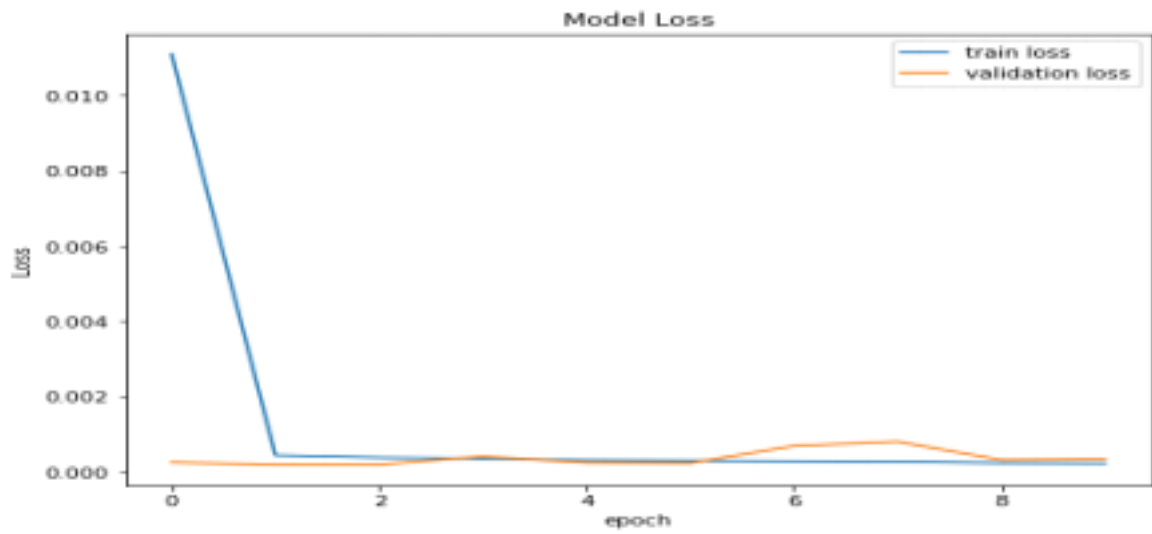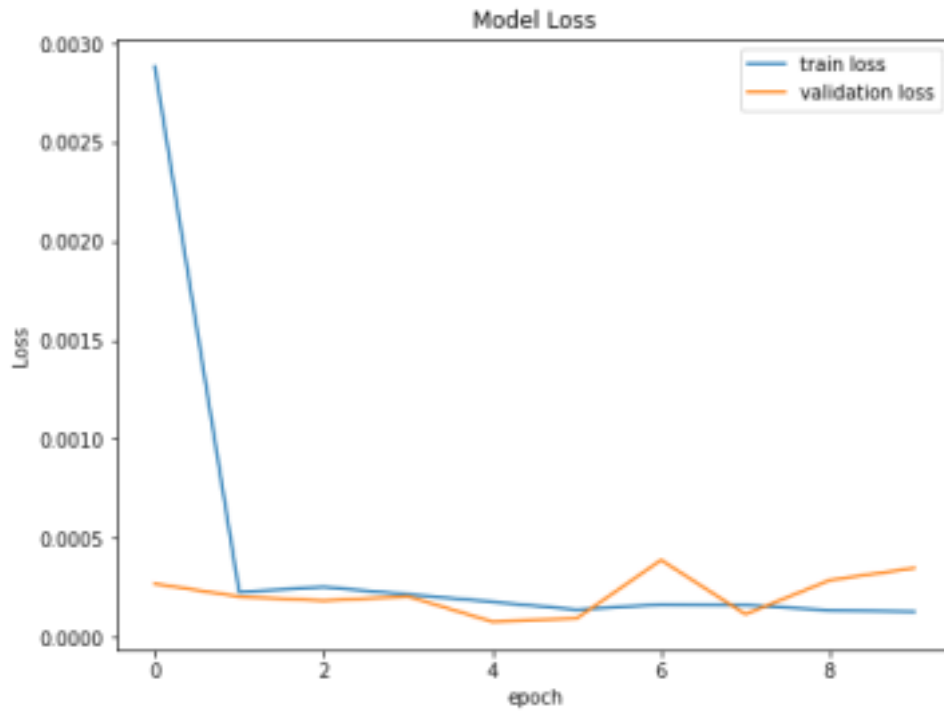
2. Predicted vs Original data with date
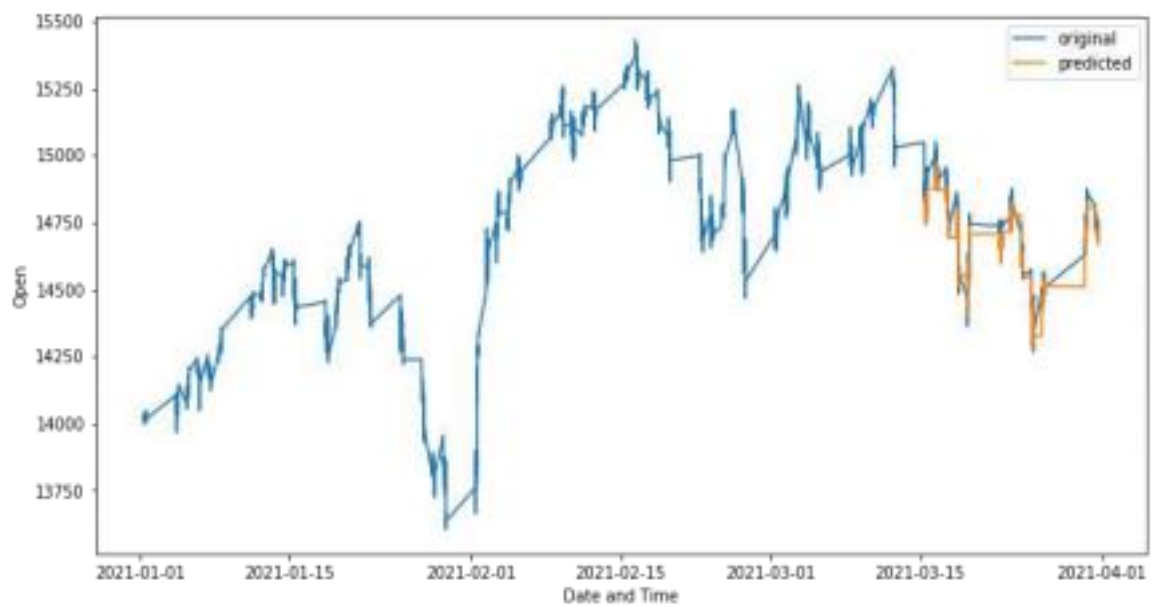
3. Accuracy:

Nifty Data	99.8170

```
print("Accuracy :", 100-(100*(abs(c.Original-c.Predicted)/c.Original)).mean())
```
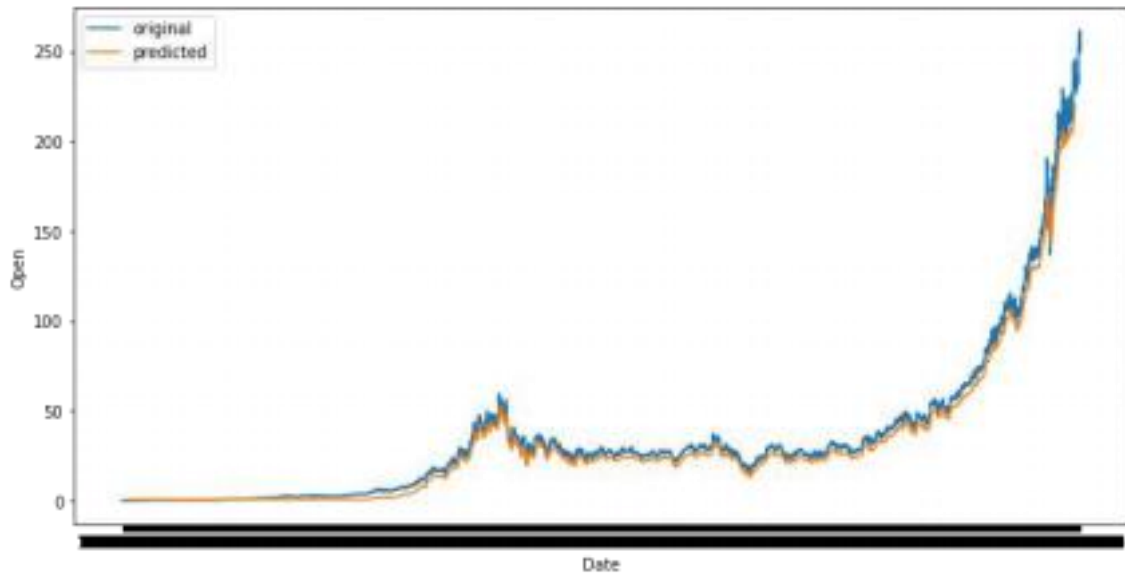
```
Accuracy : 99.81704374948961
```

10.2 MSFT DATA:

1. Model Loss is a parsimonious mathematical description of the behaviour of a collection of risks constituting an insurance portfolio.

Model Loss

2. Predicted vs Original data with date:

3. Accuracy:

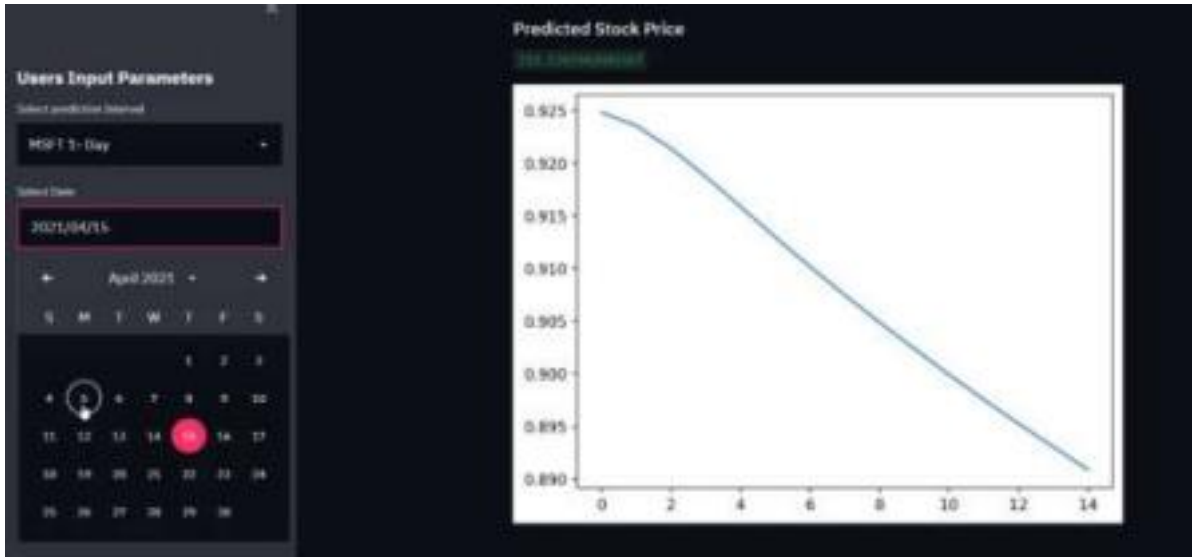| MSFT data	99.04

11.DEPLOYMENT USING STREAMLIT AND FLASK

Streamlit is an open-source Python library that makes it easy to create and share beautiful, custom web apps for machine learning and data science. Flask is a popular Python web framework, meaning it is a third-party Python library used for developing web applications.

After Deployment the model is build such a way that prediction of stock market is done by selecting an Input parameter which has 2 data that is NIFTY and MSFT .

NIFTY DATA: Can select the input parameters as Date, Hour and Minute for the stock market prediction.
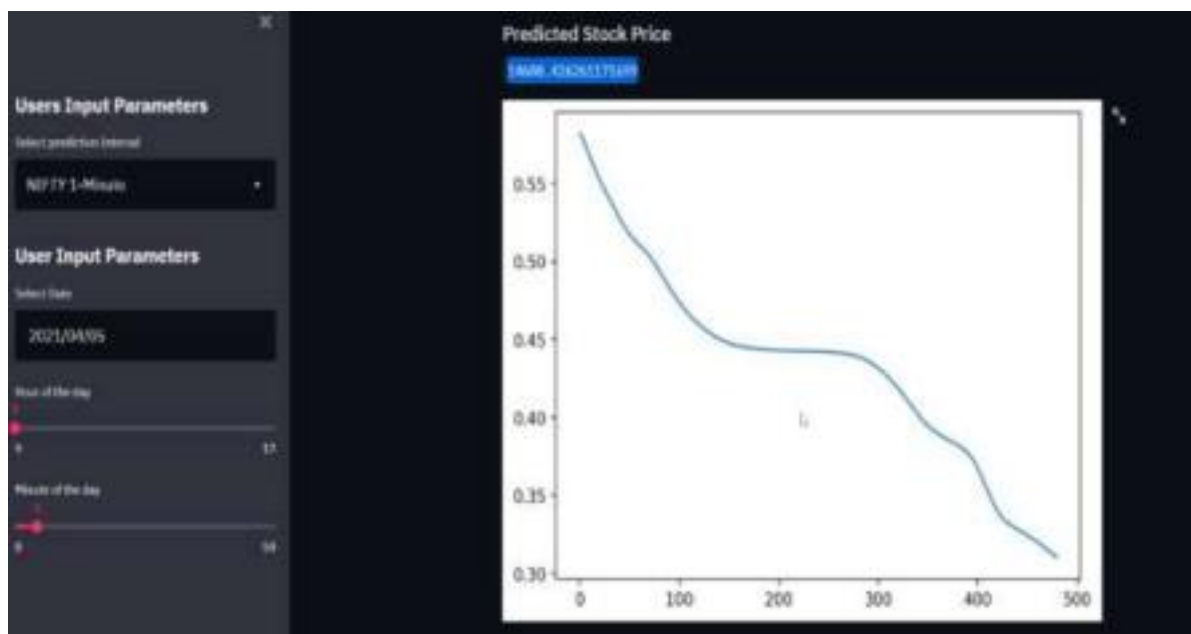
MSFT DATA: Can select the input parameters as Date for the stock market prediction.

|---X--

| **The End**