

* **AI in Bio Medicine Crash Course -**

**Stocks Price Predictor**

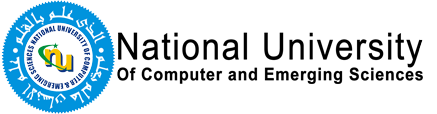
Documentation

Contributors

Ahmad Abdullah (21L-5330)

Ahsan Naveed (21L-5316)

Sajeel Haider (21L-5302)



1. **Introduction**

This project is based entirely on Machine Learning. A dataset is involved from yahoo that involves the closing, opening, upper and lower prices of the stock from 1957 - 2017. Through this data and machine learning we are able to predict future stock prices and trends.

* 1. **Prediction with Machine Learning**

Developing a stock price prediction system involves leveraging machine learning techniques to analyze historical stock data and forecast future prices. The initial step encompasses collecting historical stock data, which typically includes information such as Open, High, Low, Close, and Volume. Following data collection, preprocessing steps involve handling missing values, converting date information, and creating additional features for enhanced predictive capabilities. Subsequently, the dataset is split into training, validation, and testing sets to train and evaluate the selected machine learning model. Various models, ranging from traditional regression models like Linear Regression to more sophisticated approaches like Decision Trees or LSTM for deep learning, can be employed. The model is trained on historical data, evaluated on a validation set, and then used to predict future stock prices on a test set. The process is finalized by visualizing and interpreting the model's predictions against actual stock prices. This project requires a comprehensive understanding of data science and machine learning principles, allowing for the development of robust and accurate prediction models for financial markets.

1. **Goal**

The goal of this project is to predict the closing prices of the upcoming or future stocks

1. **Dataset**

Yahoo S&P 500 stocks data from 1957 - 2017

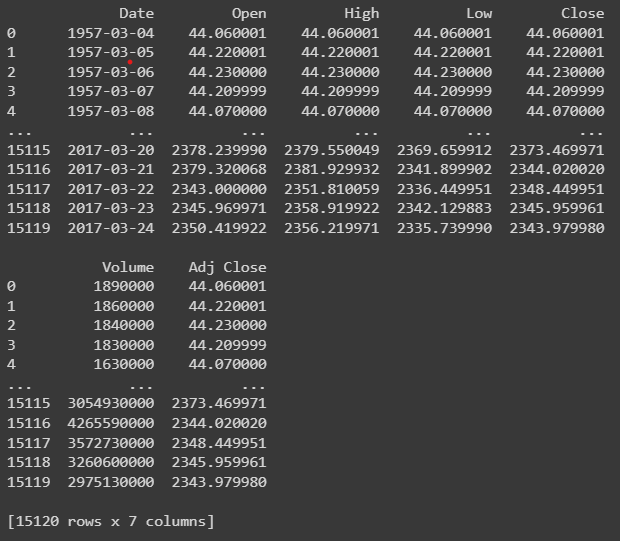
* 1. **What is S&P 500**

The S&P 500 is a market-weighted index representing 500 leading U.S. companies. It's a vital gauge for the U.S. economy's health, used by investors and analysts to track market trends and guide investment decisions.

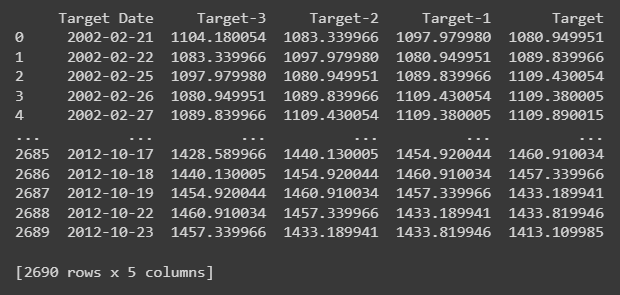
1. **Models used**
   1. **LSTM with hidden layers 20 activation is sigmoid**

The Long Short-Term Memory (LSTM) neural network architecture, configured with hidden layers set to 20 and activation functions using the sigmoid function, represents a powerful and sophisticated model for sequential data analysis. LSTMs are a type of recurrent neural network (RNN) designed to capture and remember patterns over extended sequences, making them particularly effective for time-series prediction and natural language processing tasks. The choice of 20 hidden layers allows the network to learn intricate dependencies and intricate patterns within the data. The sigmoid activation function in each layer introduces non-linearity, enabling the model to capture complex relationships and long-term dependencies in the input sequences. This configuration is often employed in applications where understanding and predicting sequential data patterns are critical, showcasing the versatility and adaptability of LSTM networks in various domains.

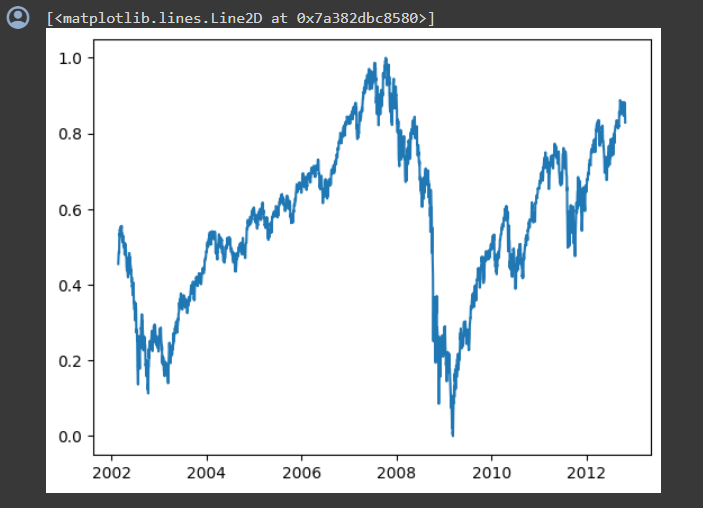
1. **Infographics**
   1. **Model 1**



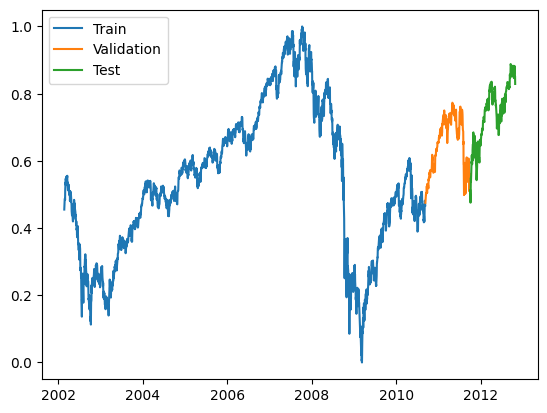
Here is the dataset which is relying on the above attributes



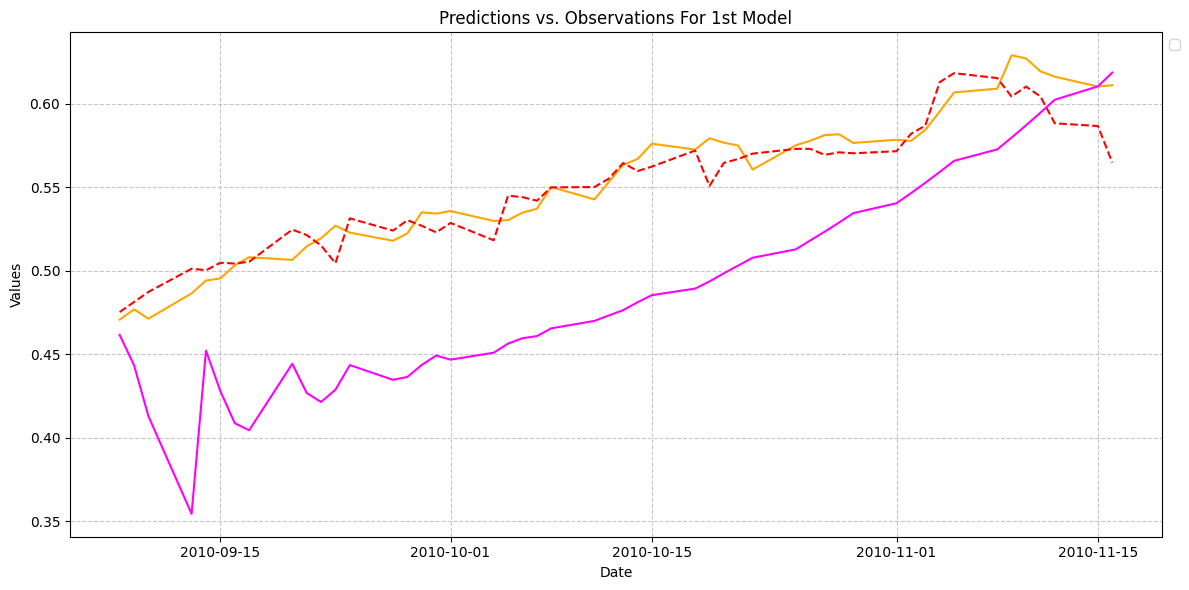
Data extracted after shortening the window



Here is the plotting of the data without feature engineering or you can say data after scaling



Data is split into three parts Train, Validation and Test



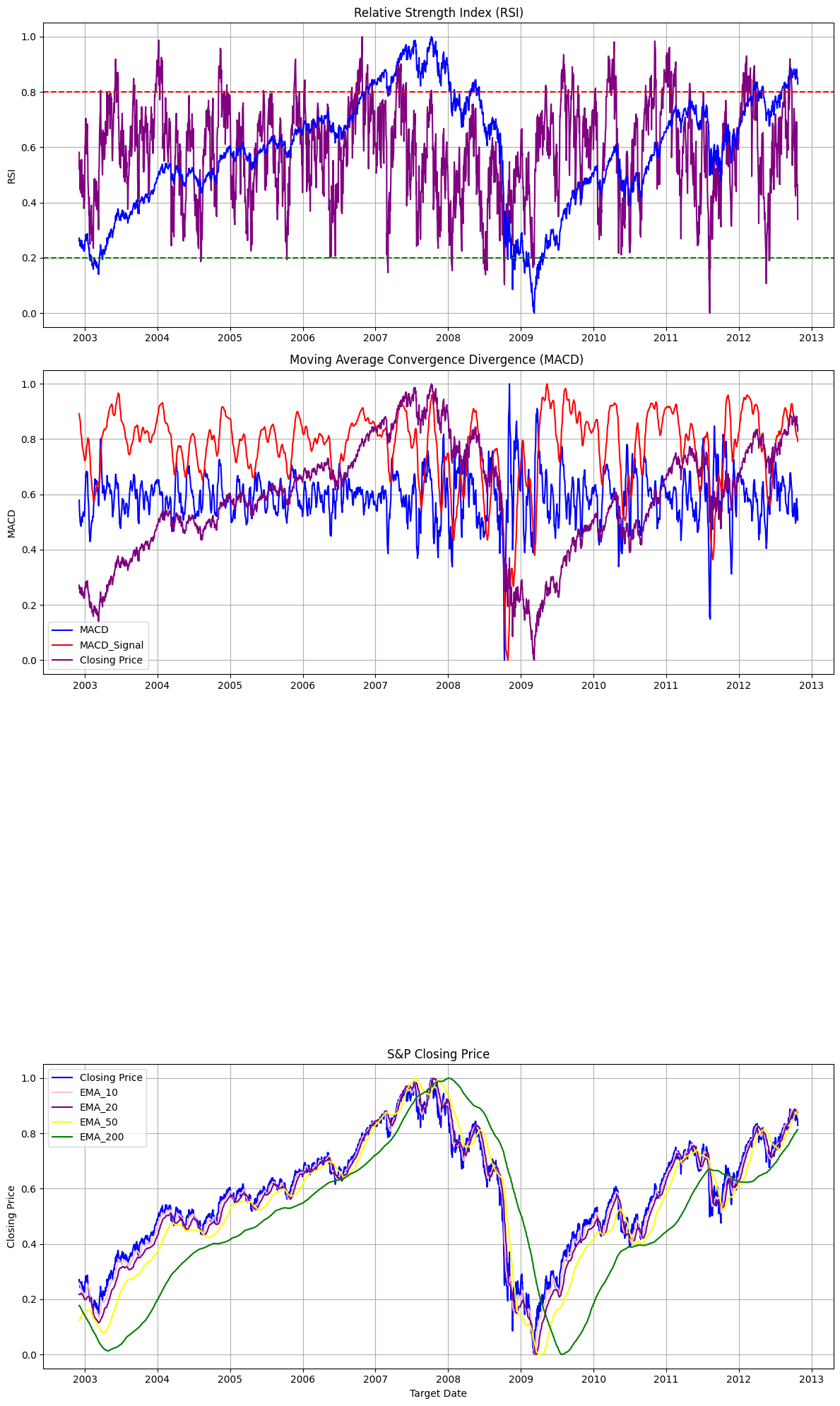
The pink shows the prediction of the 50 upcoming days while the red dotted line shows the actual price

* + 1. **Shortcomings**

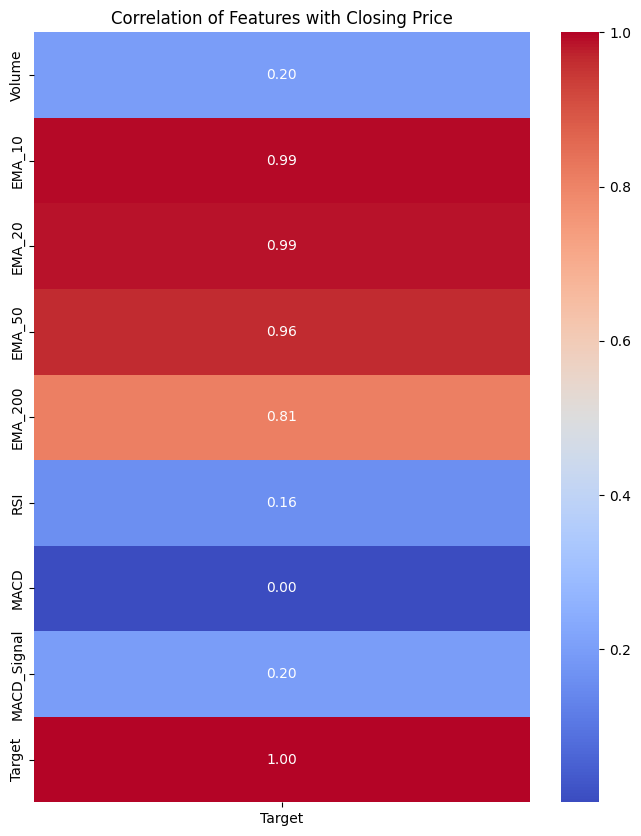
The model was too simple to predict such complex data so there was a gap which needed to be fulfilled so another model was involved. Therefore we moved to model 2 where we introduced new features with the help of feature engineering by technical analysis

* 1. **Model 2 ( Neural Networks LSTM)**
     1. **Grid Search and Kiras Classifier**

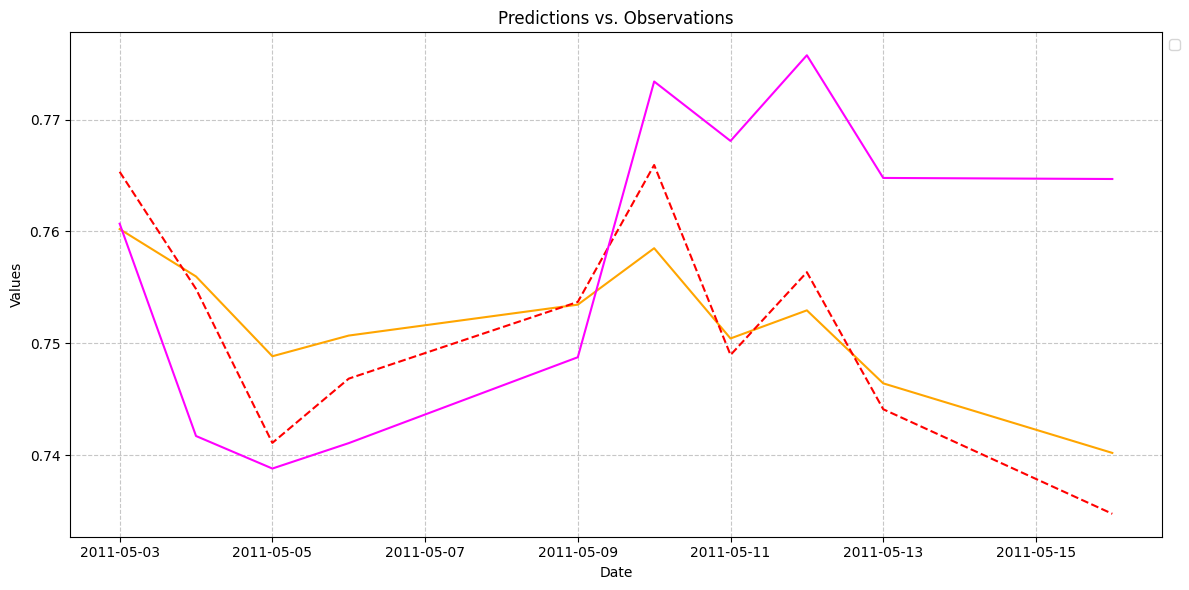
We used these two to get optimum parameters for LSTM model



These are the technical indicator which we engineered using knowledge and concepts of technical analysis

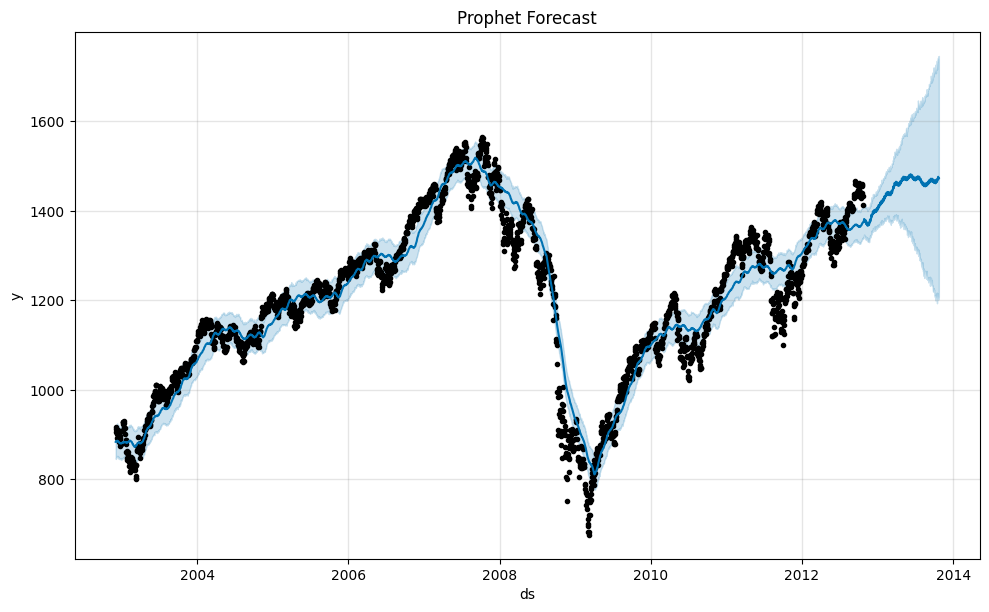


Site of Heatmap

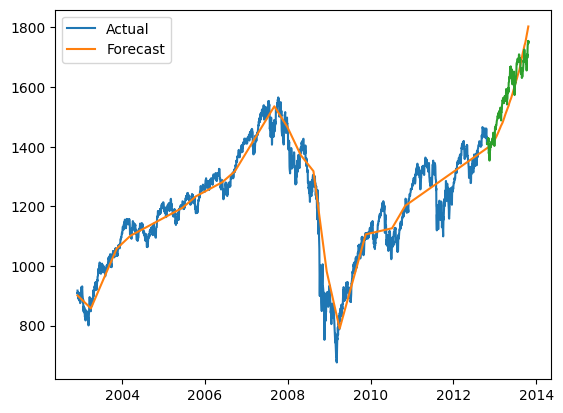


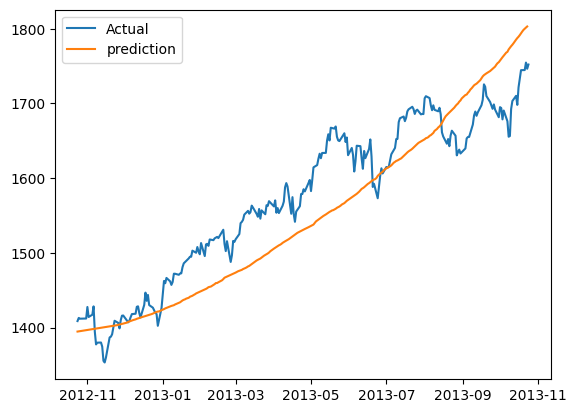
The pink shows the prediction of the 50 upcoming days while the red dotted line shows the actual price

* 1. **Model 3 ( FB Prophet)**



Dots show the actual price and line show our predicted price and at the end of x-axis we have the predictions for the next year





1. **Results**

Our stock price prediction model worked well for a short time, but it's tough to predict stock prices accurately in the long run. This is because stock prices depend on many factors beyond what our model considered. So, while it's helpful for short-term decisions, it's not reliable for long-term investing. To make better investment choices, it's important to look at a wider range of information and risk management strategies

1. **Room for improvement**

We can have better predictions if we can add more relevant features as we that stocks depend heavily on worldly and current affairs, trends and real markets and as you can see that we have no or less data related that is why the prediction we get is a bit low or reliable.

1. **Conclusion**

Our model for predicting stock prices shows promising results for short-term forecasts, but it’s not completely accurate. We’ve found that relying only on a few indicators isn’t enough to predict stock prices precisely, as they are influenced by many factors. Our findings suggest that while the model can be helpful, it’s important to remember its limitations and the complex nature of the stock market