<u>Stock Price Prediction using Time Series Forecasting</u>

Abstract:
The term "stock price" describes the current market value of a publicly traded company's share of stock. It is decided by the open market's supply and demand for the particular stock. In the financial markets, stock price prediction is essential since it helps traders, businesses, and investors make wise choices. It is an essential tool for businesses to assess market sentiment and plan their course of action, and it assists investors in efficiently managing risk, optimizing returns, and diversifying their holdings. Accurate predictions promote market stability, impact capital allocation, and bolster economic analysis by offering perspectives on industry developments and the state of the economy. The main aim of this study is to predict future stock price of Microsoft company by using Time series forecasting. Time series forecasting is highly valuable in stock price prediction as it enables investors and traders to leverage historical price data to make informed decisions about future market movements. Stock price predictions can be generated by preprocessing and evaluating the time series data, choosing a suitable forecasting model (such ARIMA, SARIMAX or machine learning methods like Random Forest, XGBoost) or LSTM, and training the model using past data. After the implementation of all the models, a comparison analysis can be done by using all this models. The model with the highest accuracy i.e. accuracy greater than 80% is chosen and is used for further predictions.

Keywords: Stock price prediction, Time series forecasting, ARIMA, Random Forest

1. Introduction

The term "stock price" describes the current market value of a publicly traded company's share of stock. It is decided by the open market's supply and demand for the particular stock. In the financial markets, stock price prediction is essential for helping traders, businesses, and investors make wise choices. Not only does it assist investors in risk management, return optimization, and portfolio diversification, but it is also an essential tool for businesses to assess market sentiment and plan their course of action. Precise forecasts contribute to market stability, impact capital allocation, and facilitate economic analysis by offering perspectives on industry developments and the state of the economy. Additionally, algorithmic trading systems are informed by stock price forecast, which enhances market efficiency and liquidity. But it's crucial to understand that stock price predictions are inherently unpredictable, and wise decision-making should combine these projections with a more comprehensive investing approach. Due to its ability to increase savings and stimulate the economy, the stock market, with its high risk and reward structure, has grown in popularity as a place to put money. Risk factors are present in this dynamic and complicated system, though. Expert investors utilize technology to forecast stock values in order to lower risk.

In the financial markets, stock price prediction is essential since it helps traders, businesses, and investors make wise choices. It is an essential tool for businesses to assess market sentiment and plan their course of action, and it assists investors in efficiently managing risk, optimizing returns, and diversifying their holdings. In order to predict potential movements in stock prices, time series forecasting analyzes previous price data to find trends, seasonality, and autocorrelation patterns. It helps investors to efficiently manage risk, including placing stop loss orders and allocating their portfolios optimally, and to develop short-term projections. Though time series forecasting improves prediction accuracy, it should only be used as one component of a larger investment plan that takes into account both quantitative and qualitative analysis. It is crucial to recognize that a variety of factors impact stock markets. The outcomes of past performance may not be indicative of future success when using historical data. Predictions of stock prices should be utilized in conjunction with a comprehensive investment plan and risk management techniques.

There are two types of time series forecasting. They are Univariate forecast and Multivariate forecast. As the name implies, a univariate time series consists of only one time-dependent variable. Multiple time-dependent variables combine to form a multivariate time series. Each variable has some dependence on other variables in addition to its historical values. Future values are forecast using this dependency. Time series forecasting is majorly classified into classical or statistical models, Machine learning models and Deep Learning models. Statistical models include Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Seasonal AutoRegressive Integrated Moving Average (SARIMA) etc. Linear Regression, XGBoost, Random Forest etc are some of the Machine learning models. Long short-term memory (LSTM) network, recurrent neural network (RNN) etc are Deep Learning models.

Data Collection:

This study forecasts stock values using real-time financial news and comments data of Microsoft company from Yahoo Finance. Over 40,000 stock quotes from various exchange marketplaces are supported by the data, which is received from the (United States Securities and Exchange Commision) SEC. The paper provides a dataframe for each stock quote and focuses on historical price data for a sample of equities for a period of last five years from 2018 to 2023 of Microsoft stock price. The information covers the volume of shares traded as well as the daily starting, maximum, minimum, and closing values. The objective is to use this data to forecast future stock values.

2. Project Description

The present research is an advanced exploration into the dynamic field of financial forecasting, with a particular focus on Microsoft stock price prediction. The project uses a wide range of forecasting models, including machine learning algorithms like Random Forest and XGBOOST and traditional time series models like ARIMA (AutoRegressive Integrated Moving Average) and SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors), all of which are leveraged from a rich dataset of historical Microsoft stock price data. Furthermore, the research employs a deep learning methodology by utilizing LSTM (Long Short-Term Memory), a recurrent neural network version. To enable the use of these models, the dataset is carefully preprocessed to organize it into a logical time series. The project's goal is to reliably estimate Microsoft's stock values through this thorough process, while also offering insights into the relative advantages and disadvantages of each forecasting method. The project aims to make a significant contribution to the field of financial data analysis by using Microsoft's stock as a case study. This will demonstrate how different forecasting models may be adapted to capture the intricate dynamics of stock market activity.

The project addresses challenges with financial forecasting, especially with regard to identifying intricate patterns and trends in changes in stock prices. Seasonality and anomalies in financial data are difficult for conventional time series models like ARIMA and SARIMAX to adjust to. It is difficult for machine learning models such as Random Forest and XGBOOST to identify complex associations. Because of its large parameter space, the integration of LSTM, a deep learning model, presents difficulties for training optimization and hyperparameter tuning. The study presents new approaches that use ensemble learning with XGBOOST and Random Forest to capture complex non-linear patterns. Sequential financial data with long-term dependencies is better captured when LSTM is included. The research assesses and contrasts several forecasting models, highlighting the advantages and disadvantages of each.

3. Background

Several papers have been reviewed in context of stock price prediction using Time series forecasting in order to find out different existing methods and technologies. Some of the previous work has been presented below.

The study by Wong et al [1] presents a novel method for high-frequency (fifteen-minute interval) stock price prediction using properly chosen exogenous variables. The stock price data of Tesla is used in this study. Support vector regression, multilayer perceptron, random forest, and XGBoost are four basic machine learning algorithms that were used to verify the suitability and precision of employing these features for stock price predictions.

In order to forecast and comprehend the dynamics of future stock prices, the study by S. Addagalla et al [2] focuses on the examination of real-time stock market data. Using daily TATA datasets from Yahoo Finance, the study compares the performance of LSTM and RNN in stock market prediction, producing noticeably better prediction accuracy. The study comes to the conclusion that machine learning approaches have the potential to improve the efficiency and accuracy of stock market predictions.

The study by M. Hirey et al [3] uses a hybrid modeling approach to improve stock price prediction by fusing different machine learning and deep learning models. They have included RIL index data from the National Stock Exchange of India for a given period of time in their dataset. To evaluate each regression model's predictive accuracy, the authors build and contrast three different models: LSTM, Auto-ARIMA, and Linear Regression. The findings indicate that the LSTM-based univariate model performs better than the others, especially when it comes to predicting the closing prices of Reliance Industries Limited (RIL) over the following 25 days.

The study by D. S. A. Elminaam et al [4] focuses on the challenging task of forecasting stock market returns, which is frequently made more difficult by the volatile financial markets and the complex, non-linear structure of stock prices. The study uses multiple predictive models, such as K-Nearest Neighbors (KNN), Random Forests (RF), Linear Regression (LR), and Gradient Boosting (GB), to forecast the closing prices of three companies from different industries the following day: Bank of New York Mellon Corp., HP Inc., and Pfizer. By utilizing artificial intelligence and increased computational capabilities, the study is able to forecast these closing prices.

In the study by K. Prakhar et al [5] predicted the future stock price trends of National Stock Exchange – India (NSE) using Time series forecasting. Initially, a comparison analysis was made among the time series forecasting models ARIMA, Facebook Prophet Model and the ETS models. Among all the models, it was found that that the Facebook Prophet model works best to predict the stock price trends for a short-term basis.

The author S. Sarvesh et al [6] presents a novel stock price prediction model of National Stock Exchange (NSE) using Long short-term memory (LSTM) and Convolutional Neural Network (CNN) in his study. A hybrid model using deep learning models LSTM and CNN was constructed. It was found that this model has good accuracy and can be successfully implemented to predict stock prices.

The study by Daryl et al [7] emphasizes the growing demand of stock ownership as an investment choice, with a focus on Apple's stock. The "SARIMA" (Seasonal Autoregressive Integrated Moving Average) model, a statistical technique frequently employed for stock market forecasts due to its capacity to handle the dynamic character of stock prices, is the subject of the study. Three distinct "SARIMA" models are developed in the study, and their effectiveness is assessed. All things considered, the "SARIMA" model is considered a useful statistical tool for data that exhibits distinct patterns, but it might not be the best choice for forecasting the erratic character of stock markets, like Apple's.

The present study by Y. Wang et al [8] endeavors to improve prediction accuracy by utilizing a new hybrid model called DWT-ARIMA-GSXGB to tackle the crucial problem of stock price forecasting. By applying the discrete wavelet transform, the method divides the stock data set into approximation and error parts. To handle the incorrect data, an enhanced XGBoost model (GSXGB) is utilized. The ARIMA models (0, 1, 1), (1, 1, 0), (2, 1, 1), and (3, 1, 0) are used to process the approximate data. Particularly in fitting the opening stock index price, the simulation results validate the hybrid model's high approximation and generalization capabilities.

The study by G. W. R. I. Wijesinghe et al [9] highlights the critical role of time series forecasting in decision-making across diverse scientific fields. Employing K-means clustering and Principal Component Analysis, the research clusters a dataset based on central points and Euclidean distance measurement, ultimately identifying the most significant contribution sector in the Colombo Stock Exchange (CSE) during the 2008-2017 period. The study's focus on forecasting high volatility stock price indexes in CSE represents a relatively unexplored area in financial literature.

The author M. Goyal et al [10] predicted the stock prices of two energy companies using Time series forecasting. The author mainly implemented the deep learning model LSTM recurrent neural network to predict the stock prices. It had been found that this model had given best results with low error.

From the previous works, it can be concluded that stock price prediction can be done by using time series forecasting models very effectively. Different models like ARIMA, SARIMA, Random Forest, XGBoost, Support Vector Regression, LSTM, RNN, CNN, hybrid models etc were implemented on different stock price datasets like Apple, Tesla, TATA, National Stock Exchange (NSE) etc. All these are accurate for precise stock price predictions. In all of these papers, a clear comparison analysis among some time series forecasting models like ARIMA, SARIMA, Random Forest, XGBoost, LSTM is not done particularly on stock price data of Microsoft. Thus, this study mainly aims to extend the previous works stock price prediction on Microsoft data using Time series forecasting models ARIMA, SARIMA, Random Forest, XGBoost and LSTM. After a detailed comparison analysis, the model with highest accuracy will be chosen for further predictions.

The project makes use of scikit-learn for machine learning models like Random Forest and XGBOOST, TensorFlow or PyTorch for deep learning models like LSTM, and Python libraries like statsmodels and pmdarima for traditional time series models ARIMA and SARIMAX. Tools for managing and visualizing data, such as pandas and matplotlib, are utilized to ensure a thorough and effective implementation of different forecasting models.

4. Problem Definition

The project addresses several formal mathematical problems intrinsic to the realm of financial forecasting. One key problem involves defining the time series forecasting task itself, wherein given a sequence of historical Microsoft stock prices P_t, the objective is to predict the future stock price $P_{\{t+k\}}$ at a certain time horizon k. This is formally expressed as the minimization of the prediction error

$$E_k = P_{t+k} – hat(P_{t+k})$$

where $hat(P_{t+k})$ denotes the predicted stock price.

For the ARIMA and SARIMAX models, the formal definition includes establishing the autoregressive *(AR)*, differencing *(I)* and moving average *(MA)* components, along with the identification of seasonality factors in SARIMAX. In the context of Random Forest and XGBOOST, the problem involves formulating an ensemble of decision trees $F(x)$ that collectively minimize the mean squared error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_{t+i} - F(x_i))^2$$

where x_i represents the features associated with the stock prices. Lastly, for LSTM, the problem is formalized as the optimization of the neural network parameters to minimize the loss function

$$L = \sum_{i=1}^{n}(P_{t+i} - hat(P_{t+i}))^2$$

where $hat(P_{t+i})$ is the LSTM-predicted stock price. The formalization of these problems lays the groundwork for the subsequent application of mathematical and computational techniques within each forecasting model, guiding the optimization processes toward accurate stock price predictions.

Given the volatility of the financial markets, the project must employ machine learning models, hyperparameter tuning, training optimization, and traditional time series models to tackle mathematical difficulties. In order to predict stock prices effectively, one must also strike a balance between model complexity and interpretability, which calls for a sophisticated comprehension of both market dynamics and mathematical formalizations.

The project forecasts stock values in volatile financial markets using a variety of models, including as LSTM, Random Forest, SARIMAX, ARIMA, and XGBOOST. In order to identify long-term connections in sequential financial data, these models use deep learning with LSTM, navigate non-linear interactions, and capture temporal patterns. An in-depth analysis of these models offers valuable insights into tactics for predicting stock prices that work well in dynamic market environments.

5. The Proposed Techniques

The project uses a time series forecasting models to forecast future stock prices based on historical data. It uses a variety of models, such as LSTM for deep learning, Random Forest and XGBOOST for machine learning-based ensemble approaches, and ARIMA and SARIMAX for conventional time series analysis. Every model is customized to tackle issues such as long-term dependencies, seasonality, and non-linearity. Because of the framework's versatility, several approaches may be thoroughly evaluated, allowing the most successful solutions for stock price prediction in dynamic financial markets to be identified.

The project forecasts stock prices using a variety of methods and models. ARIMA and SARIMAX are used in traditional time series analysis to capture seasonal components and temporal trends. For identifying non-linear correlations, machine learning models such as Random Forest and XGBOOST are utilized. Complex neural networks are constructed using deep learning techniques such as LSTM. Within the project's extensive framework, each technique provides a comprehensive approach to stock price prediction through data preprocessing, model training, and evaluation.

1. ARIMA

The AutoRegressive Integrated Moving Average, or ARIMA, is a fundamental time series forecasting model that is frequently used because of its capacity to identify and anticipate temporal trends in sequential data. Moving average (MA), auto-regression (AR), and differencing (I) are the three main parts of the model. In the auto-regressive component, the relationship between an observation and its historical values is modeled. By deducting the prior observation from the present observation, differencing resolves non-stationarity and stabilizes the time series mean. Past forecast mistakes are weighted averaged and represented by the moving average component. ARIMA is a flexible framework that may be used to capture anomalies, trends, and seasonality in a variety of time series data. This flexibility is what makes ARIMA so useful. approaches. The ARIMA model is a helpful tool in the resources of time series forecasting techniques since its parameters are optimized to maximize its prediction accuracy.

It finds trends and dependencies in historical data by capturing temporal patterns. The moving average component of the model records previous forecast errors, while the autoregressive and integrated components deal with non-stationarity. Because of its adaptability, ARIMA can handle a variety of time series data types, which improves its accuracy for predicting stock values in volatile financial markets.

2. SARIMAX

A time series forecasting model called SARIMAX uses exogenous variables to take into consideration outside influences on market movements. In the stock market, where news, events, and economic indicators can affect prices, it is especially helpful. SARIMAX captures trends, patterns, and stationarity across time by keeping the moving average, autoregressive, and differencing components of ARIMA. In order to address recurrent patterns in stock prices, it also incorporates seasonality variables. SARIMAX provides a flexible and reliable framework for stock price prediction by taking into account both the intrinsic time series structure and outside influences. This enables analysts to incorporate a wider range of data into the forecasting process.

3. Random Forest

In the field of time series forecasting, Random Forest—a flexible ensemble learning technique—emerges as a potent stock price prediction tool. Random Forest uses the combined knowledge of several decision trees to reduce overfitting and enhance generalization when predicting stock prices. Because each tree is trained on a different sample of the data, the models become more diverse and are better able to identify subtle patterns in the movements of stock prices. Random Forest is an excellent tool for deciphering complex patterns in stock prices since it can handle non-linear interactions and deep dependencies inside financial data. Randomness is incorporated into the training process to guarantee robustness and adaptability—two essential qualities for stock price prediction, which is influenced by a multitude of uncontrollable factors. Random Forest offers a comprehensive and robust method of stock price prediction by combining the predictions from individual trees. This makes it a useful substitute for conventional time series forecasting models.

4. XGBoost

One of the best models for stock price prediction in the field of time series forecasting is XGBoost, an advanced gradient boosting technique. With regard to stock price prediction, XGBoost is particularly good at identifying complex patterns and relationships found in financial data. Through an iterative process that emphasizes the correct prediction of cases that were difficult for prior trees, XGBoost combines the advantages of decision trees and boosting algorithms to create a powerful predictive model. It excels at handling feature interactions and non-linear correlations, which makes it especially well-suited to the intricate dynamics of stock markets. Regularization techniques are used into XGBoost in order to improve the model's robustness and reduce overfitting, which is an important consideration when working with noisy and volatile financial data. Furthermore, XGBoost facilitates the integration of external characteristics, offering adaptability in incorporating pertinent data that extends outside the temporal framework of the time series. XGBoost is a powerful instrument for reliable and accurate stock price predictions because of its versatility, precision, and capacity to handle a wide range of data features.

5. LSTM

Recurrent neural network (RNN) based model Long Short-Term Memory (LSTM) becomes a powerful and popular model for time series forecasting stock price prediction. LSTM is very effective at identifying the complex dynamics included in stock price fluctuations, and it is particularly well-suited for capturing long-term dependencies and patterns in sequential data. Because of its architecture, which includes memory cells and gates, LSTM may selectively recall or forget information, which makes it possible to simulate complex temporal correlations in the context of stock price prediction. In the financial markets, where past trends and minute patterns can have a big impact on future stock values, this talent is essential. The efficiency of LSTM is attributed to its resistance to vanishing gradient difficulties, context sensitivity, and adaptation to changing time lags. Moreover, the model's capacity for long-term learning and memory makes it a useful instrument for forecasting stock values over a range of time periods. Long short-term memory (LSTM) is a major player in the field of time series forecasting for financial markets because, despite its complexity, it has shown to be a reliable and adaptable method for capturing the fine details of stock price movements.

6. Visual Applications

The following flowchart is used to implement the suggested methods. The following describes the steps involved in implementing the suggested techniques:

Data Collection:

Microsoft's historical stock price data from January 1, 2018, to January 1, 2023, is sourced from a trustworthy source and includes relevant information like the date and closing price.

Data Preprocessing:

An imputed or removed approach is put into place to deal with missing data. Feature engineering is used to design additional features like technical indicators, moving averages, and other financial metrics. To guarantee comparability across all attributes, the data is scaled or normalized.

Feature Selection:

Based on the analysis, features important for stock price prediction are chosen. Aspects including volume, historical pricing, and perhaps outside influences on stock prices are taken into account.

Exploratory Data Analysis (EDA):

In order to comprehend patterns, trends, and seasonality in the past stock prices, basic statistics and data visualizations are analyzed.

Data Splitting:

To make the process of developing and evaluating models easier, the historical data is separated into training and testing sets. The data is separated chronologically in order to preserve the temporal integrity of the information.

Data Modelling:

In this, different time series models like statistical models, machine learning models and deep learning models are employed in order to carry out the stock price prediction.

Statistical Models:

ARIMA (AutoRegressive Integrated Moving Average):

The order (p, d, q) hyperparameter is adjusted to get the best fit for the Microsoft stock price data.

SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors):

By adding exogenous variables that could affect the time series, SARIMAX expands on ARIMA. Exogenous variables that are pertinent are found and added to the model.

Machine learning models:

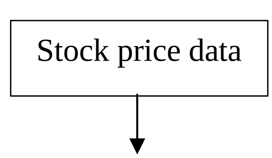
Random Forest and XGBoost:

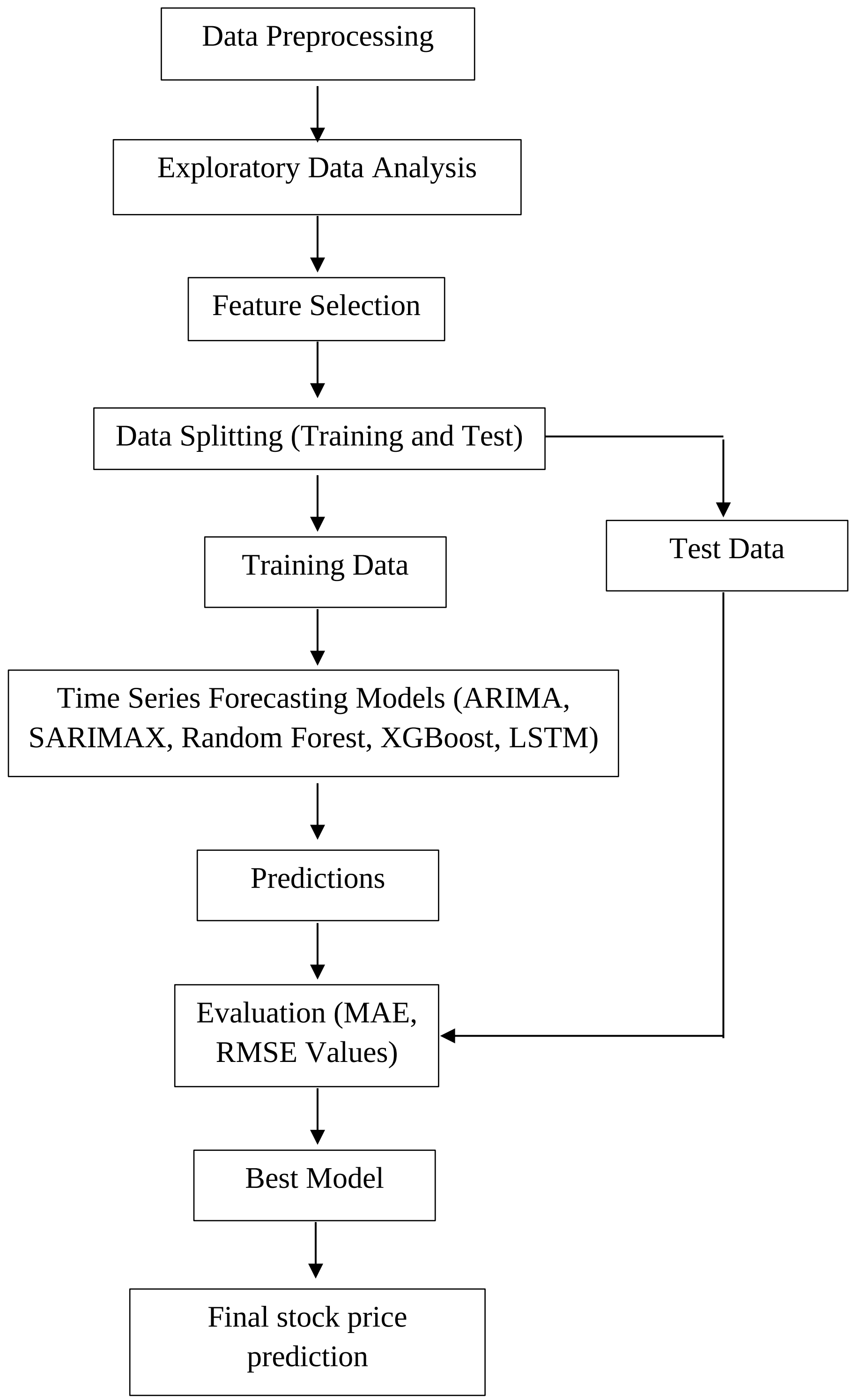
Random Forest and XGBoost are two examples of ensemble learning approaches that aggregate predictions from several base models. Hyperparameters are tuned to improve predictive performance once these models are trained on historical stock price data.

Deep Learning Model:

LSTM (Long Short-Term Memory):

Recurrent neural networks (RNNs) of the long-term dependency type (LSTM) are used to extract long-term dependencies from the Microsoft stock price time series.





Model Evaluation:

Metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and others are used to assess each model's performance. The model predictions are contrasted, and the advantages and disadvantages of each strategy are noted.

7. Experimental Evaluation

The following provides an explanation of the thorough exploratory data analysis of Microsoft's stock price data during a five-year period, from 01-01-2018 to 01-01-2023:

Exploratory Data Analysis (EDA):

The figures below display the first and last five rows of the Microsoft stock data.

	Open	High	Low	Close	Adj Close	Volume
Date						
2018-01-02	86.13	86.31	85.50	85.95	80.229	22483800
2018-01-03	86.06	86.51	85.97	86.35	80.602	26061400
2018-01-04	86.59	87.66	86.57	87.11	81.312	21912000
2018-01-05	87.66	88.41	87.43	88.19	82.320	23407100
2018-01-08	88.20	88.58	87.60	88.28	82.404	22113000

	Open	High	Low	Close	Adj Close	Volume
Date						
2022-12-23	236.11	238.87	233.94	238.73	236.632	21207000
2022-12-27	238.70	238.93	235.83	236.96	234.877	16688600
2022-12-28	236.89	239.72	234.17	234.53	232.469	17457100
2022-12-29	235.65	241.92	235.65	241.01	238.892	19770700
2022-12-30	238.21	239.96	236.66	239.82	237.712	21938500

The dataset consists of 1259 observations with six columns

Open price: The price of the stock at the beginning of the trading day.

High price: The highest price of the stock during the trading day.

Low price: The lowest price of the stock during the trading day.

Close Price: The final price of the stock at the end of the trading day.

Adjusted Closing price: The closing price adjusted for dividends and stock splits.

Volume: The total number of shares traded during the trading day.

The daily open, close, high, low, and volume of equities traded each day are the input variables. Exogenous variables that could also affect stock prices include sentiment in the news, economic data, and industry-specific variables. The target variable for prediction is the Close Stock Price.

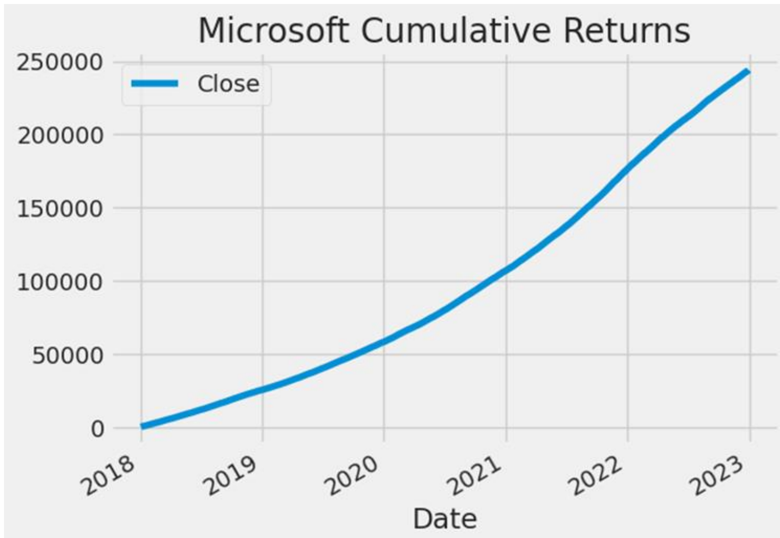
```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1259 entries, 2018-01-02 to 2022-12-30
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Open       1259 non-null   float64
 1   High       1259 non-null   float64
 2   Low        1259 non-null   float64
 3   Close      1259 non-null   float64
 4   Adj Close  1259 non-null   float64
 5   Volume     1259 non-null   int64
 6   Date       1259 non-null   datetime64[ns]
dtypes: datetime64[ns](1), float64(5), int64(1)
memory usage: 78.7 KB
```

From the above figure, it can be observed that the dataset has no missing values. Hence, this dataset can be used for further steps.

The graph below shows how the closing price changed annually during a five-year period, from 2018 to 2022:

According to the graph, the closing price increased until the end of 2021 and then decreased in 2022. This could be attributed to the COVID-19 situation, which shows that unexpected events, geopolitical developments, and economic factors can all easily affect stock price data and raise the uncertainty rate.

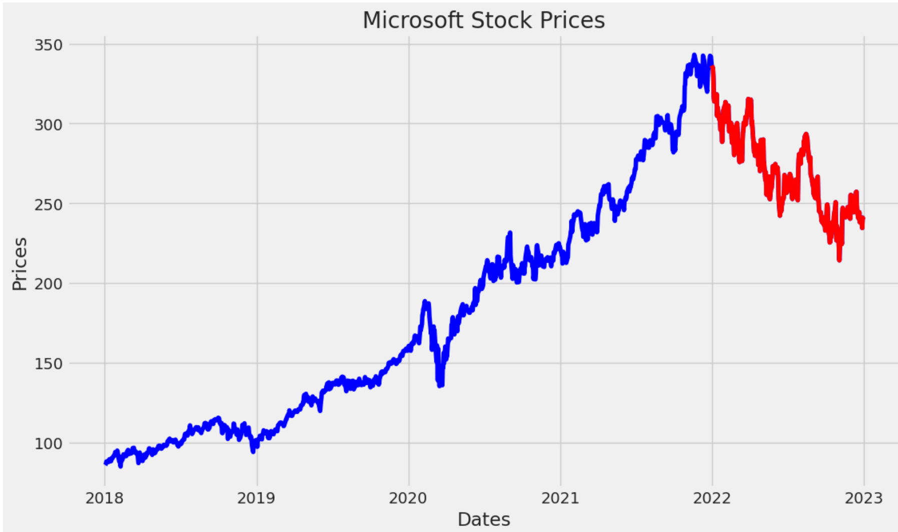
The graph below shows the trend of Microsoft's stock price data:



From the graphs, it can be observed that the trend of the close price for Microsoft stock data has positive trend. It can be seen that the stock price is continuously increasing from 2018 to 2022.

Data Splitting:

In this project for data modelling, about 80% of the data is considered training for training the model. Remaining 20% of the data is taken as tests data, in order to perform validation of the predictions.



In this above graph, the blue color data represents the training data and the red color data represents the test data.

Results:

The results obtained by all the time series models are explained below as follows:

1. ARIMA

Stationarity test:

In order to evaluate the stationarity of time series data—a critical component of time series modeling techniques—stationarity tests are required. Tests like the Phillips-Perron (PP) exam, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, and Augmented Dickey-Fuller (ADF) test are frequently used. PP permits several deterministic trends specifications, KPSS identifies trends, and the ADF test assesses non-stationarity. In order to select a model and guarantee the reliability of the analysis, the test option is determined by the properties of the data and the analysis assumptions.

In this, the stationarity test is performed using Augmented Dickey-Fuller (ADF) test. Time series data stationarity can be assessed statistically using the Augmented Dickey-Fuller (ADF) test. It identifies non-stationarity in a univariate time series by looking for the presence of a unit root. The alternative hypothesis for the test proposes stationarity, while the null hypothesis for the test presumes a unit root. The ADF test is used by researchers to verify model appropriateness and statistical inference reliability in econometrics and time series analysis. The results of the stationarity test are presented below:

```
Test Statistic: -1.3093219635176894
P-Value: 0.624896785991395
Non-Stationary
```

The results show that the data is not stationary. In order to employ ARIMA model, the data should be converted into stationary which can be obtained by taking the difference of the values.

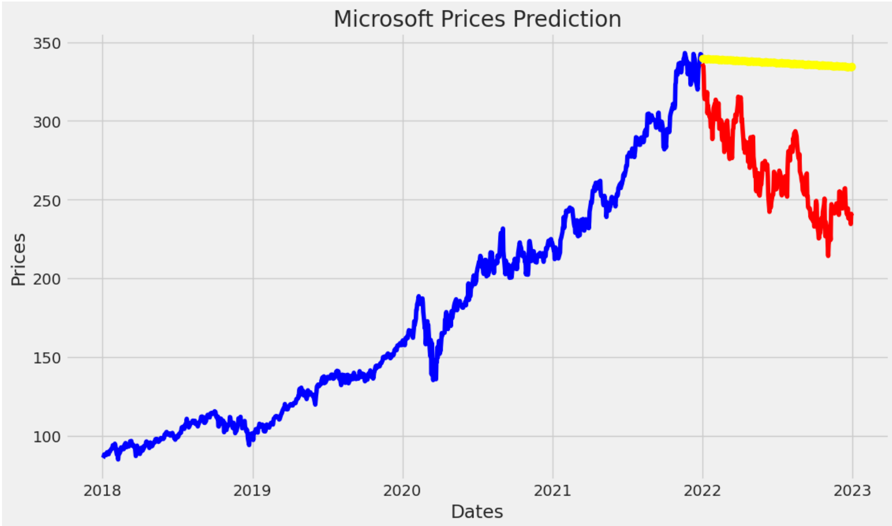
```
Test Statistic: -11.975767092935117
P-Value: 3.8051651890473555e-22
Stationary
```

Thus, the data is made stationary. Hence, the ARIMA model can be successfully employed.

The Mean Absolute Error (MAE) and Root Mean Square (RMSE) values obtained by ARIMA model are 66.23 and 70.41 respectively. The predictions obtained by this model can be seen in this below figure:

2. SARIMAX

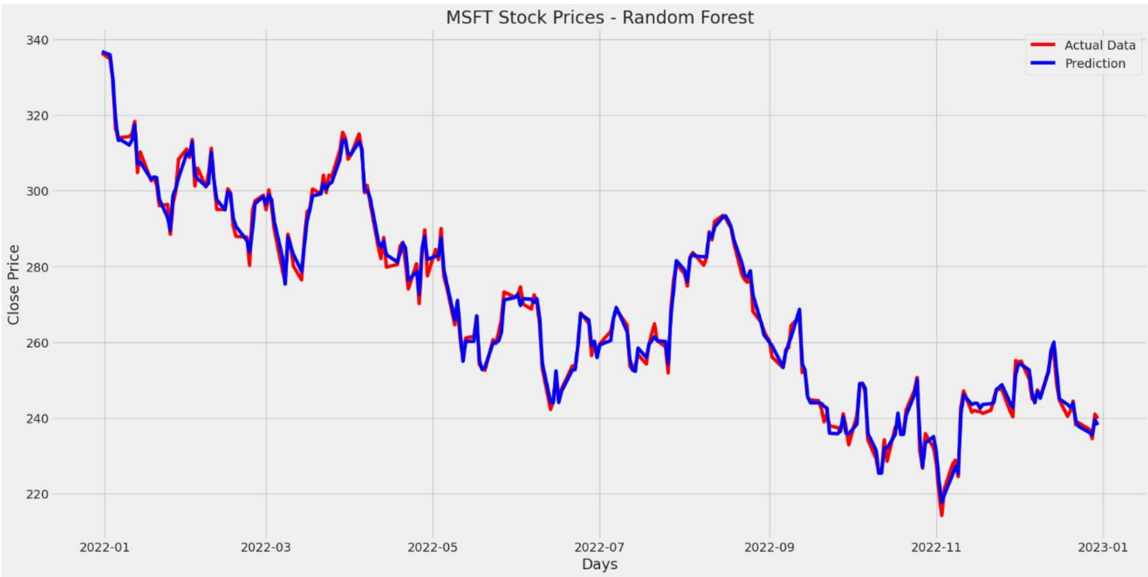
The Mean Absolute Error (MAE) and Root Mean Square (RMSE) values obtained by SARIMAX model are 67.80 and 72.18 respectively. The predictions obtained by this model can be seen in this below figure:



Here the blue line represents the training data, red line represents the actual test values and the yellow line represents the predictions.

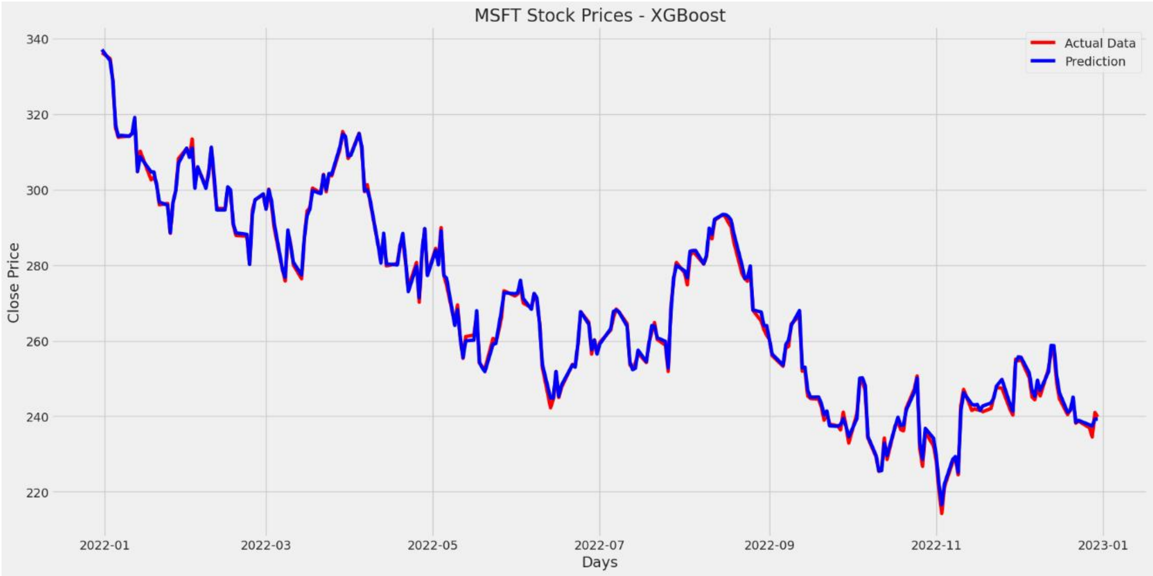
3. Random Forest

The Mean Absolute Error (MAE) and Root Mean Square (RMSE) values obtained by Random Forest model are 1.21 and 1.71 respectively. The predictions obtained by this model can be seen in this below figure:



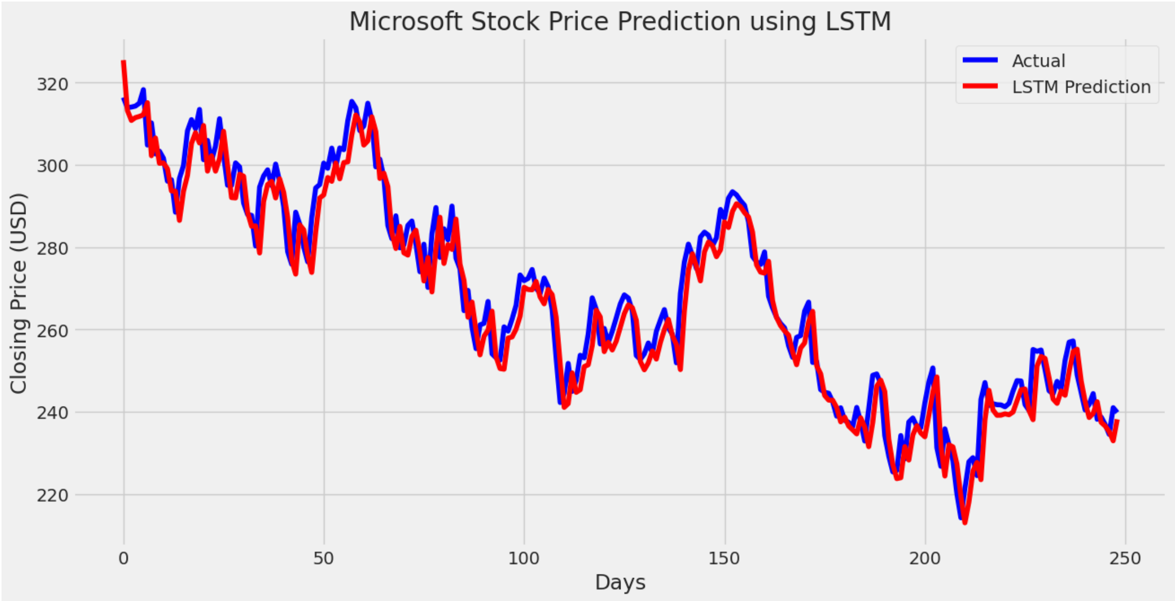
4. XGBoost

The Mean Absolute Error (MAE) and Root Mean Square (RMSE) values obtained by XGBoost model are 0.49 and 0.69 respectively. The predictions obtained by this model can be seen in this below figure:



5. LSTM

The Mean Absolute Error (MAE) and Root Mean Square (RMSE) values obtained by LSTM model are 4.93 and 6.24 respectively. The predictions obtained by this model can be seen in this below figure:



Comparison analysis of all the models:

The table below displays the MAE and RMSE values that each model produced:

S. No.	Model	MAE Value	RMSE Value
1	ARIMA	66.23	70.41
2	SARIMAX	67.80	72.18

3	Random Forest	1.21	1.71
4	XGBoost	0.49	0.69
5	LSTM	4.93	6.24

From the table, it can be observed that the model XGBoost has given the lowest MAE and RMSE values. The R2 score of XGBoost model is obtained as 0.999. Thus, XGBoost model is the best model.

The predicted values of Microsoft data given by the XGBoost model are shown in the below table.

Date	Actual values	Predicted values
2022-12-23	238.73	238.45
2022-12-27	236.96	238.92
2022-12-28	234.53	237.74
2022-12-29	241.01	237.45
2022-12-30	239.82	239.24

From the above table, it can be observed that the actual and predicted values by XGBoost model are very close. Thus, the XGBoost model can be used for stock price prediction of Microsoft.

Conclusions:

Finally, by utilizing a variety of models such as ARIMA, SARIMAX, Random Forest, XGBoost, and LSTM, the project on time series forecasting for stock price prediction produced insightful information about the predictive power of each approach. The results of a comprehensive analysis and comparison showed that the XGBoost model was the best at predicting the price of Microsoft shares. This result implies that the ensemble learning strategy used by XGBoost performed better at capturing the intricacies of the stock price time series by integrating the advantages of distinct base models. The XGBoost model outperformed previous approaches in part because of its capacity to handle non-linear interactions and capture complex patterns. Predictive models should be utilized cautiously in real-world trading scenarios because to the inherent uncertainties introduced by the dynamic nature of financial markets. The project's overall conclusions highlight how crucial it is to use modeling strategies that are suited for the features of the financial time series data.

8. Future Work

In the project on time series forecasting for stock price prediction, there are a number of directions to pursue for future study, with an emphasis on the XGBoost model as the top performance. The XGBoost model's capacity to identify minute patterns and correlations in the stock price data might first be improved by further developing feature engineering. It might also be possible to increase the XGBoost model's forecast accuracy by investigating the effects of adding exogenous variables—that is, outside variables—like market trends, news mood, and economic indicators. A more reliable and varied forecasting framework might be created by looking at model ensembling techniques, combining the benefits of XGBoost with other algorithms, or even investigating neural network topologies. Furthermore, ongoing observation and retraining of the model with new data may facilitate adaptation to changing market conditions. The project might also be expanded to include a thorough examination of feature significance and model interpretability in order to give stakeholders more useful information. The stock price prediction system will continue to be improved and more reliable with continued exploration and iteration in these directions.

9. References

1. Wong, J. Figini, A. Raheem, G. Hains, Y. Khmelevsky and P. C. Chu, "Forecasting of Stock Prices Using Machine Learning Models," 2023 IEEE International Systems Conference (SysCon), Vancouver, BC, Canada, 2023, pp. 1-7, doi: 10.1109/SysCon53073.2023.10131091.
2. S. Addagalla, S. Koppuravuri, R. Krosuri, M. S. Kunapareddy, S. Reddy Mallu and M. Rashmi, "Stock Market Price Prediction Using Machine Learning Techniques," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-6, doi: 10.1109/INCET57972.2023.10170222.
3. M. Hirey, J. Unagar, K. Prabhu and R. Desai, "Analysis of Stock Price Prediction using Machine Learning Algorithms," 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, 2022, pp. 1-4, doi: 10.1109/ICONAT53423.2022.9725888.
4. D. S. A. Elminaam, A. E. Tanany, M. A. Salam and M. A. E. Fattah, "CPSMP_ML: Closing price Prediction of Stock Market using Machine Learning Models," 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 2022, pp. 251-255, doi: 10.1109/MIUCC55081.2022.9781756.
5. K. Prakhar, S. S, S. E, K. M and S. K. B, "Effective Stock Price Prediction using Time Series Forecasting," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 1636-1640, doi: 10.1109/ICOEI53556.2022.9776830.
6. S. Sarvesh, R. V. Sidharth, V. Vaishnav, J. Thangakumar and S. Sathyalakshmi, "A Hybrid Model for Stock Price Prediction using Machine Learning Techniques with CNN," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-6, doi: 10.1109/ISCON52037.2021.9702382.

7. Daryl, A. Winata, S. Kumara and D. Suhartono, "Predicting Stock Market Prices using Time Series SARIMA," 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI), Jakarta, Indonesia, 2021, pp. 92-99, doi: 10.1109/ICCSAI53272.2021.9609720.
8. Y. Wang and Y. Guo, "Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost," in China Communications, vol. 17, no. 3, pp. 205-221, March 2020, doi: 10.23919/JCC.2020.03.017.
9. G. W. R. I. Wijesinghe and R. M. K. T. Rathnayaka, "ARIMA and ANN Approach for forecasting daily stock price fluctuations of industries in Colombo Stock Exchange, Sri Lanka," 2020 5th International Conference on Information Technology Research (ICITR), Moratuwa, Sri Lanka, 2020, pp. 1-7, doi: 10.1109/ICITR51448.2020.9310826.
10. M. Goyal, "Financial Time Series Stock Price Prediction using Deep Learning," 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2020, pp. 0378-0383, doi: 10.1109/IEMCON51383.2020.9284879