IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Stock Market Analysis Using Machine Learning

Helly Patel[1], Dr. Vikas Tulshyan[2] , Prof. Naimish Patel[3]
[1]Comp Dept., Silver Oak College of Engg. & Technology, Silver Oak University, Ahmedabad
[2] IT Department, Silver Oak College of Engineering & Technology, Silver Oak University, Ahmedabad
[3] IT Department, Silver Oak College of Engineering & Technology, Silver Oak University, Ahmedabad

Abstract **:** Analysis of the stock market is crucial for investors and financial institutions to make informed decisions. As historical stock market data and advances in machine learning algorithms increase, the interest in using machine learning in stock analysis is growing. This study provides an in-depth analysis of stock market analysis using machine learning, focusing on the application of various machine learning techniques and methods. Research begins with data collection, where historical stock market data is collected from sources such as financial databases, APIs, and online surveys. The data are pre-processed to handle missing values and outliers and to generate relevant features for analysis. Feature selection and dimensionality reduction techniques are used to reduce the complexity of the dataset. Next, various machine learning algorithms are applied to the preprocessed data, including linear regression, decision trees, random forests, support vector machines, and neural networks. These algorithms are trained and evaluated using metrics such as mean squared error (MSE), accuracy and F1 scores to assess their effectiveness in predicting stock prices and trends. The study also explores the use of advanced machine learning techniques, such as deep learning, including long-term memory (LSTM) networks to analyze stock markets.

Index Terms:- Machine learning , ARIMA model , LSTM method, Stock Market Analysis, Market Forecasting

1.INTRODUCTION

An essential component of financial decision-making is stock market analysis, which involves assessing investment opportunities, market trends, and stock prices. With its sophisticated methods for analyzing and forecasting changes in stock prices, machine learning (ML) has become a potent instrument in the analysis of the stock market. Large data sets may be processed by ML algorithms, which can also spot intricate patterns and produce accurate predictions. These capabilities help analysts and investors make well-informed decisions. The ARIMA model is a useful tool in stock market analysis because of its ease of use, interpretability, and capacity to identify time-series trends. Although it might not always be the most precise model, it offers a strong basis for comprehending and forecasting changes in stock prices. Because LSTM networks can record.

1.1 Stock Market Analysis Using ML: A Comprehensive Look

Stock market forecasting is an attractive area of research for many researchers. Accurate stock price forecasts allow investors to make more informed decisions about buying and selling stocks at the right time. In recent years, many researchers have studied the predictability of stock prices using machine learning algorithms. The main purpose of this research is to identify the right machine learning techniques that can be used to predict stock prices. The methodology used to achieve this goal is a review of 12 research papers on the use of machine learning and deep learning algorithms to predict stock prices. To distinguish this study from the current study, instead of selecting random articles, it focuses on two main sectors, which are banking and healthcare. The results of this study show that long-term short-term memory and limited repetition unit techniques produce the best results for most of the selected books, regardless of the industry in which they are used. important limitations of the study..[1] As the popularity of stock trading increases, individuals and financial entities such as investment firms, hedge funds, and private investors actively participate in the stock market to earn profits. Many strategies have been developed and implemented, from traditional methods that use fundamental and technical analysis to modern approaches that use cutting-edge technology. However, determining the optimal method remains difficult. Designing an effective strategy in a complex and dynamic stock market environment creates serious difficulties. Therefore, the purpose of this article is to provide an overview of machine learning applications in the stock market and find out the most used

machine learning models or methods for market forecasting. In addition, the study aims to identify the strategy that achieves the highest accuracy in forecasting stock prices. Based on a systematic literature review that included quantitative and qualitative analyses, Support Vector Machine (SVM) was found to be the most popular machine learning technique for predicting stock prices. However, Long Short-Term Memory (LSTM) stands out as the machine learning technique that shows the highest accuracy, achieving an impressive accuracy of 99.58 percent. These findings highlight the effectiveness of LSTM in predicting stock value with exceptional accuracy. The results of this study contribute to a comprehensive understanding of machine learning in the stock market and provide valuable information for developing effective trading strategies. Investors and investors can use LSTM and SVM techniques to improve their stock market forecasts and make informed investment decisions. [2]

This literature review summarizes existing research on the use of machine learning in stock market forecasting. The review includes research from various sources such as journals, conference proceedings and theses. Methods used to forecast stock markets using machine learning include decision trees, support vector machines, artificial neural networks, and time series analysis. The review also highlights the advantages and limitations of these methods and their applications in the stock market. The results of the review show that machine learning can provide valuable information about the stock market, but there is still room for improvement in terms of accuracy and robustness. The review concludes by suggesting future directions for research in the field..[3] The average person is interested in stock trading because it is a good way to make money if done well. Although this is a difficult task as it requires a lot of information about market shares and trends to forecast stock prices. Stock markets are inherently volatile and dynamic. Observing the behavior of stock prices is very uncertain and difficult. Previous stock forecasting was done using technical, fundamental and econometric models. The sheer volume and complexity of data collected in today's markets makes it impossible for traditional analytical methods to work effectively. Machine learning is becoming an increasingly useful tool for stock valuation. By using sophisticated algorithms to analyze large data sets, machine learning can help identify patterns and trends that may signal future changes in stock prices. Machine learning can also be used to find the correlation between stock prices and the external environment. It can also be used to decide when it would be a good time to buy, sell or hold a stock. Stock price estimation can be done accurately using a machine learning algorithm. It is also used to detect market anomalies such as insider trading and market manipulation. Stock market price forecasting has become a burning research problem among traders, investors and traders. In this forecasting activity, investors need fast and real-time information to make quick and accurate decisions. Recently, most researchers have developed algorithms that predict the average movement of a stock and its price. This research paper provides an overview of various modern forecasting techniques, especially machine learning and sentiment analysis, highlighting the dataset, forecast types and metrics used to predict stock prices. The findings, results, research gap and future scope of all these techniques used are also discussed [4]

Machine learning makes predictions based on current stock market index values by training its past values in a sequential and timely order using an artificial neural network, while deep learning makes predictions based on current stock market index values by training its past values. . . sequential timed ordering using an artificial neural network.[5] Stock market prices are generated en masse and change by the second. The stock market is a complex and difficult system where people either make money or lose their life savings. This research paper attempts to predict the development of the stock market. Predicting stock market movements can be a major disadvantage in the stock market. Social media fully represents people's feelings and opinions on current events. Twitter has played an important role in attracting much attention from researchers to the study of human emotions. The main analysis was public opinion on Twitter, supported by exchange forecasts and associated with alternative social media. To achieve the goal, this research mainly uses machine learning techniques to study various factors related to the stock market. Machine learning approaches have a successful track record of extracting information by building stock market prediction models from stock datasets. Data mining of this data can be useful for stock market forecasting. The aim of the work is to investigate how well existing changes, i.e. increases and decreases in the costs of the company, correlate with public opinions expressed in the tweets of the company.[6]

1.2 METHODOLOGY

1.2.1 Machine Learning

The work focuses on the technical analysis segment, which involves performing statistical analysis of data, understanding charts and identifying trends in the stock market. Two approaches were used in the project to forecast the stock market: a new ARIMA-LSTM hybrid was designed that combines neural networks with time series forecasts to capture the linear and non-linear part of the time series. Another prediction library created by Facebook called Prophet, which handles missing data and outliers using intuitive parameters for optimal predictions, was also used. Finally, the algorithms are compared and a more accurate algorithm with fewer errors is selected to predict the future stock market. The user can enter the name of the company stock for which he wants to receive forecasts. It retrieves real-time stock data, applies a selected machine learning algorithm, trains on past data, and finally predicts the expected future trend and stock values with user visualization. The proposed method includes collecting real-time stock market data using the yfinance API, developing a new ARIMA-LSTM hybrid, applying Prophet to time series data, comparing the two algorithms to find the optimal solution, and finally applying it to stock market forecasting. system [1].

1.2.2 Model Processing

The paper will focus on the technical analysis segment that includes performing a statistical analysis of the data, understanding the charts and identifying the trends in the stock market. Two approaches have been used in the project for stock market prediction: a novel ARIMA- LSTM hybrid has been designed which combines neural networks with time forecasting series for capturing the linear and non- linear portion of the time series. Another forecasting library called Prophet designed by Facebook has also been used that handles the missing data and outliers uses intuitive parameters for optimal predictions. Finally, the algorithms are compared and the more accurate and less error prone algorithm is selected for future stock market prediction. The user can input the company's stock name whose predictions he wants to get. It will retrieve the stock's live data, apply the chosen machine learning algorithm, train the previous data and finally predict the expected future trend and stock values along with visualizations to the user. The proposed methodology involves capturing the live stock market data using the yfinance API, developing the novel ARIMA-LSTM hybrid, applying Prophet on the time series data, comparing the two algorithms to find an optimal solution and finally deploy it for the stock market prediction system.

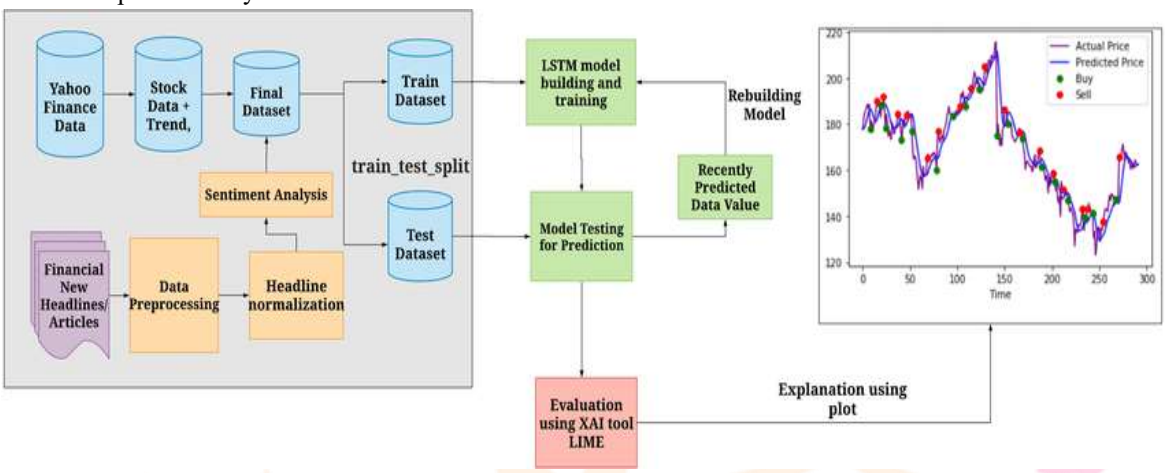


Figure: 1 Stock Market Prediction System

Analysis of the stock market is essential for investors to make wise choices. Because machine learning algorithms can identify intricate patterns and relationships in data, they have been employed more and more in the analysis of stock market data in recent years. This paper investigates the use of two well-liked machine learning techniques for stock market analysis: Long Short-Term Memory (LSTM) neural networks and Autoregressive Integrated Moving Average (ARIMA). The main goal of the research is to forecast stock prices using past price, volume, and technical indicator data from the stock market. A classic technique for time series analysis, the ARIMA model works well for identifying linear correlations in the data. Conversely, long-term dependencies in sequential data can be captured by LSTM, a kind of recurrent neural network.

The evaluation is based on metrics such as mean absolute error (MAE), mean squared error (MSE) and R-squared (R2) to assess the accuracy of forecasts. The results show that both ARIMA and LSTM models can effectively predict stock prices, and LSTM outperforms ARIMA in forecasting accuracy. Overall, this study demonstrates the effectiveness of machine learning methods, especially ARIMA and LSTM, in stock market analysis and provides valuable insights. for investors and financial analysts. Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) are two popular methods used in time series forecasting, including stock market analysis. Let's dive into each method: Automatic Integrated Moving Average (ARIMA): Auto Regressive (AR) Component: ARIMA models use an autoregressive component that models the relationship between an observation and several lagged observations (ie its past values). It reflects the linear relationship between the current value and its historical values. Integrated (I) component: The integrated component of ARIMA refers to the separation of raw observations so that the time series can be stationary. Stationarity is important because many time series forecasting methods assume that the underlying data are stationary. Moving Average (MA) Component: The moving average component of ARIMAN models the relationship between the current value and a linear combination of past forecast errors. This helps smooth out data noise and catch short-term trends.

Long-Term Short-Term Memory (LSTM): Memory cells: LSTM networks have memory cells that can store information for long periods of time. This is achieved by a system of gates (input gate, forgotten gate, output gate) that direct the flow of data into and out of the memory cell. Sequence prediction: LSTM networks are well suited for sequence prediction, such as stock market analysis, because they can learn patterns and relationships in sequential data. This makes them particularly effective at capturing long-term dependencies in time series data. Variable-length series: LSTM networks can handle variable-length series, which is useful for modeling stock market data where the number of historical data points used for forecasting can vary. . Based on stock market analysis, ARIMA models are often used for short-term forecasting of stock prices, especially when there are linear patterns in the data. On the other hand, LSTM networks are better suited to capture complex non-linear relationships in stock price movements, making them effective for long-term forecasting and capturing trends in data. Both ARIMA and LSTM have their strengths and weaknesses, and it

is possible to choose between them. they depend on the characteristics of the data and the forecasting task at hand. In practice, it is often useful to try both methods and choose the one that gives the best performance for a given data set.

Stock analysis using machine learning involves a systematic approach that includes data collection, preprocessing, feature design, model selection, training, evaluation, deployment and continuous improvement. Initially, historical stock data including prices, volumes and various financial indicators are collected from various sources such as APIs, databases and news platforms. Next, preprocessing steps address data quality issues such as missing values and outliers while ensuring smooth scaling. Feature engineering augments the dataset with derivative functions, such as moving averages and news sentiment analysis, that enrich the forecasting capabilities. Model selection involves selecting appropriate algorithms using ensemble methods such as random forests to improve performance. After training and validating the dataset, model evaluation evaluates performance metrics blindly to test data that guides deployment decisions. In production, the model is monitored, updated and retrained to adapt to changing market conditions. However, it is important to recognize the inherent uncertainty of financial markets and to interpret model forecasts in the broader context of market dynamics and with caution when making investment decisions.[2]

2.Research Elaborating

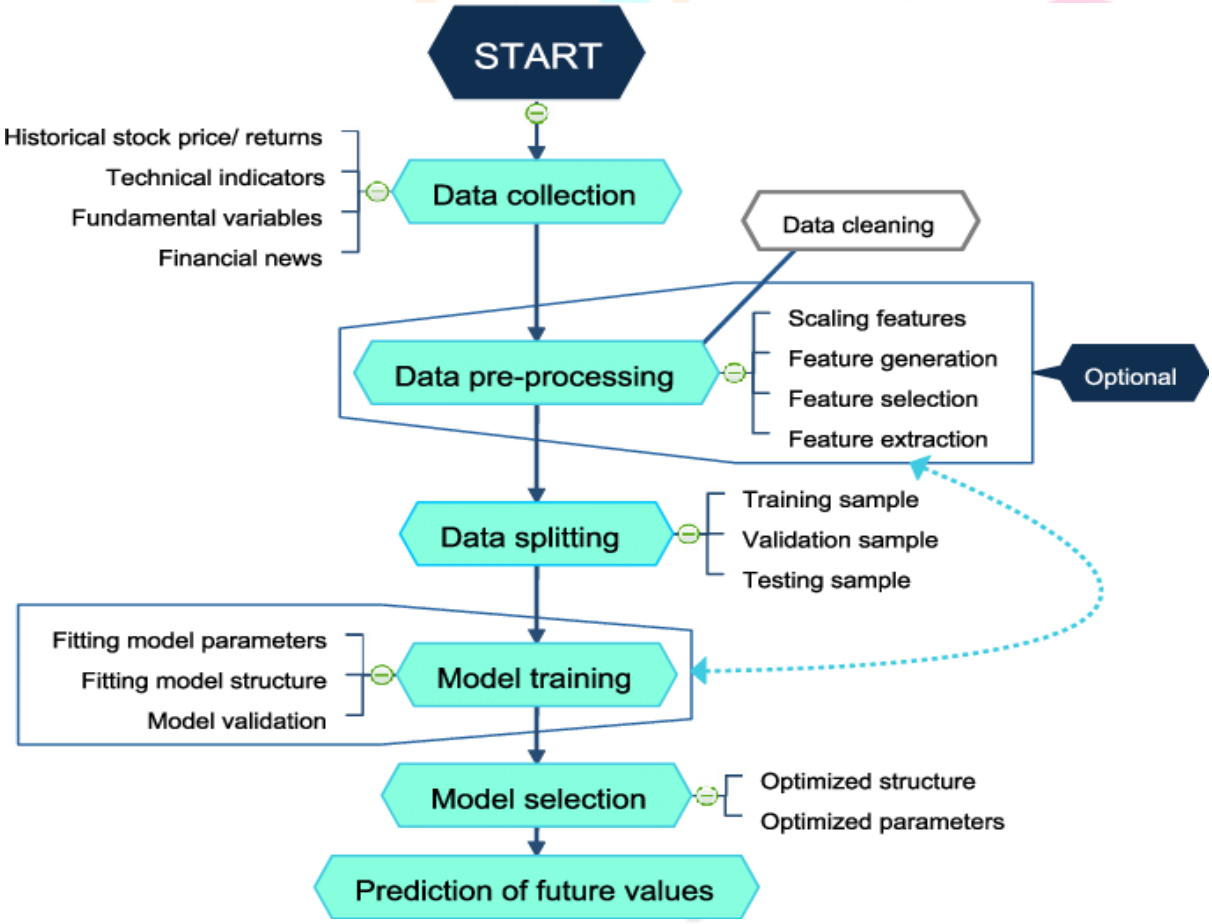


Figure 2 Workflow of a stock market prediction model[1]

```
1 ticker
```

Date	Open	High	Low	Close	Adj Close	Volume
1996-01-01 00:00:00+05:30	5.550000	5.600000	5.533333	5.583333	3.323906	985500
1996-01-02 00:00:00+05:30	5.466666	5.566666	5.288888	5.372222	3.198225	7470000
1996-01-03 00:00:00+05:30	5.133333	5.254444	5.101111	5.200000	3.095698	15160500
1996-01-04 00:00:00+05:30	5.200000	5.332222	5.144444	5.297777	3.153908	12397500
1996-01-05 00:00:00+05:30	5.297777	5.277777	5.188888	5.202222	3.097020	5008500
...	...	...	...	...	...	...
2023-02-27 00:00:00+05:30	385.100006	386.799988	378.100006	382.200012	382.200012	8142319
2023-02-28 00:00:00+05:30	382.500000	384.250000	375.000000	376.700012	376.700012	16367477
2023-03-01 00:00:00+05:30	374.250000	380.399994	373.600006	378.700012	378.700012	7766667
2023-03-02 00:00:00+05:30	379.500000	379.750000	374.700012	375.549988	375.549988	5019603
2023-03-03 00:00:00+05:30	377.000000	385.399994	376.100006	384.950012	384.950012	8180709

Snapshot 1 Training Data

'pandas_datareader' library to download stock data from Yahoo Finance using the 'yfinance' API. It prompts the user to enter a stock ticker symbol and downloads the data for that stock. The downloaded data is stored in a pandas DataFrame named 'ticker'.

```
1 nse-fo-futures-and-options.html#:~:text=NSE%20F%26O%20Stock%20List%3A%20%20%20%20SL,%20%201000%20%2052%20more%20rows%20')[0]
```

```
1 tickers = tickers.SYMBOL.to_list()
```

```
1 tickers
```

```
['FINNIFTY',
 'MIDCPNIFTY',
 'BANKNIFTY',
 'NIFTY',
 'ABBOTINDIA',
 'ABCAPITAL',
 'ACC',
 'ADANIPORTS',
 'ALKEM',
 'AMARAJABAT',
 'AMBUJACEM',
 'APOLLOTYRE',
 'ATUL',
 'AXISBANK',
 'BAJAJ-AUTO',
 'AARTIIND',
 'BALRAMCHIN',
 'BANKBARODA',
 'BEL',
```

Snapshot 2 Tickers

Importing Indian stocks
And printing it

Date	Open	High	Low	Close	Adj Close	Volume	long_MA	short_MA	crosszero	position	buy	sell
2023-03-02 00:00:00+05:30	1499.900024	1646.000000	1407.900024	1607.250000	1607.250000	26970925	2990.145002	1631.014692	1.0	0.0	NaN	NaN
2023-03-03 00:00:00+05:30	1688.849976	1905.949951	1685.500000	1879.500000	1879.500000	22888891	2988.637001	1614.264692	1.0	0.0	NaN	NaN
2023-03-06 00:00:00+05:30	1966.699951	2135.000000	1855.000000	1982.900024	1982.900024	27180456	2987.977751	1617.629402	1.0	0.0	NaN	NaN
2023-03-08 00:00:00+05:30	2040.000000	2088.000000	1941.000000	2039.650024	2039.650024	13568243	2987.078251	1628.964700	1.0	0.0	NaN	NaN
2023-03-09 00:00:00+05:30	2049.800049	2068.850098	1905.349976	1953.150024	1953.150024	12627794	2985.990752	1642.817641	1.0	0.0	NaN	NaN
2023-03-10 00:00:00+05:30	1870.000000	1940.000000	1820.599976	1896.199951	1896.199951	9419942	2984.670752	1651.435288	1.0	0.0	NaN	NaN
2023-03-13 00:00:00+05:30	1917.000000	1985.000000	1857.400024	1874.400024	1874.400024	7572604	2983.672003	1657.041174	1.0	0.0	NaN	NaN
2023-03-14 00:00:00+05:30	1874.000000	1874.849976	1651.349976	1738.199951	1738.199951	12679663	2982.131002	1653.605878	1.0	0.0	NaN	NaN
2023-03-15 00:00:00+05:30	1760.900024	1891.449951	1728.099976	1839.000000	1839.000000	12296703	2980.925253	1660.447057	1.0	0.0	NaN	NaN
2023-03-16 00:00:00+05:30	1861.000000	1875.000000	1795.000000	1800.349976	1800.349976	1411173	2979.098753	1670.970588	1.0	-1.0	NaN	1800.35

Snapshot 3 Moving Data

Calculating Moving Average[3]
And from that generating Buy and sell signal

ARIMA Model Implementation

```
1  import pandas as pd
2  import numpy as np
3  import os
4  import matplotlib.pyplot as plt
5  import yfinance as yf
6  from statsmodels.tsa.arima_model import ARIMA
7  from pandas_datareader import data as pdr
8  from datetime import date
9  import yfinance as yf
10 import statsmodels.api as smapi
11
```

Snapshot 4 ARIMA Model Implementation

Importing required Libraries and modules
This code is using the ARIMA model from the statsmodels library to forecast the future values of the daily high prices of a stock using the past data.
The yfinance library is used to download the stock price data from Yahoo Finance and save it as a CSV file. The CSV file is then loaded into a Pandas DataFrame[4].
The ARIMA model is then initialized with the order (p,d,q) of (1,1,2), which means that the model will be fit with first-order differencing (d=1), one autoregressive term (p=1), and two moving average terms (q=2).

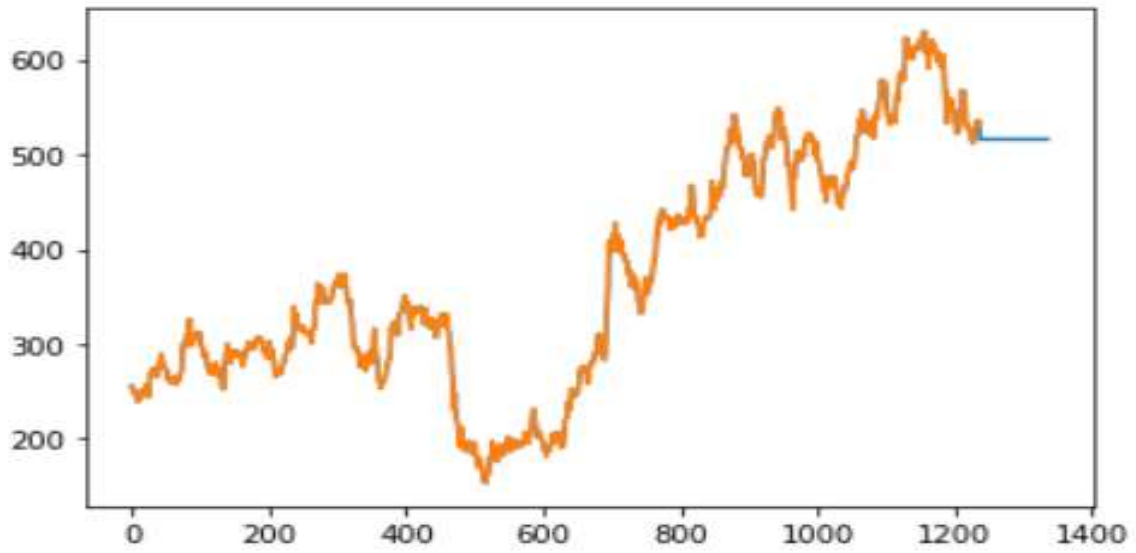
```
step = daysToPredict
#fc , se, conf =
fc=result.forecast(step)
#mse = np.mean(se)
#rmse = np.sqrt(mse)
print(fc)
```

```
988       515.705675
989       515.866501
990       515.781534
991       515.826423
992       515.802708
              ...
1083      515.810906
1084      515.810906
1085      515.810906
1086      515.810906
1087      515.810906
Name: predicted_mean, Length: 100, dtype: float64
```

Snapshot 5 Forecasting Result

Forecasting result for the given time period

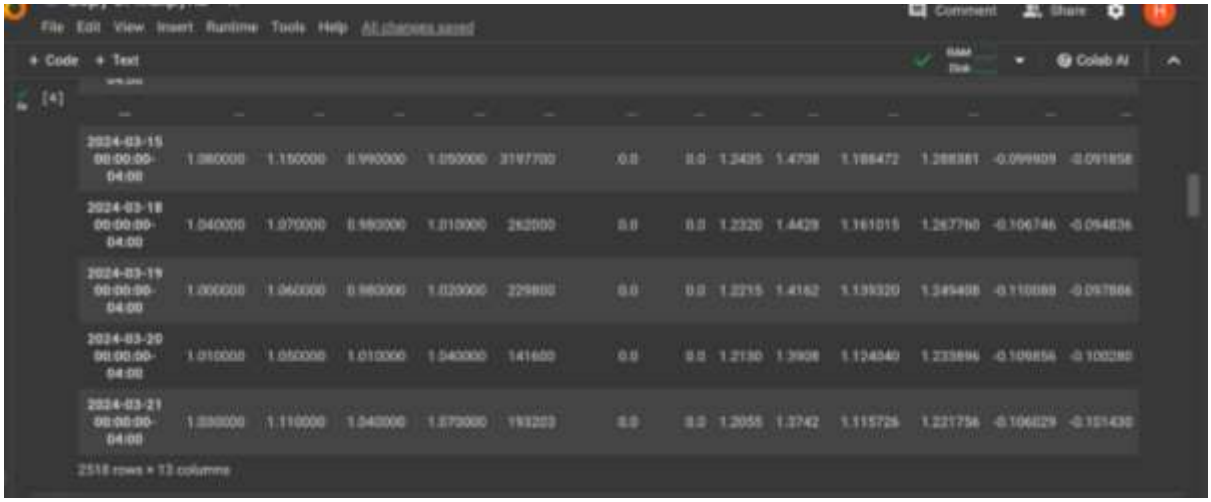
```
df_d=df.to_numpy()
df_d=np.concatenate((df_d,fc))
plt.plot(df_d)
plt.plot(df)
plt.savefig("static/arima/"+quote+".png")
```



Snapshot 6 Ploting Old Data

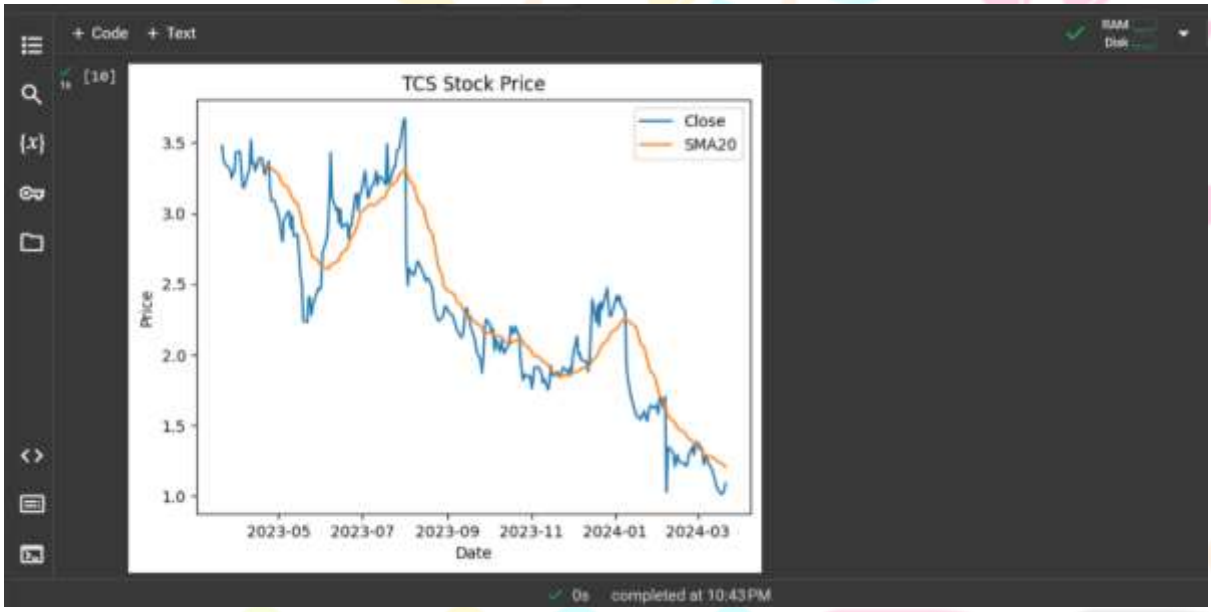
Plotting old data and based on that showing predicted output by the ARIMA model[6]

3. Result



Snapshot 6 Result Of Current TCS stock

Result of current TCS stock price prediction



Snapshot 7 Actual vs Predicted Data

4.Conclusion

By using historical stock data and various technical indicators, we can train our models to make accurate predictions about future market trends. However, it is important to note that the stock market is a complex and dynamic system that is influenced by numerous factors, many of which are unpredictable. The use of the ARIMA (Autoregressive Integrated Moving Average)[7] model to forecast the future values of daily high prices of a stock. The model is fitted to the past data, and then used to forecast the future values for a specified number of days.

5.References

[1] Adel Ismail Al-Alawi et. al., “Stock Market Prediction using Machine Learning Techniques: Literature Review Analysis” [2023]
[2] Jaydip Sen , “Stock Price Prediction Using Machine Learning and Deep Learning Frameworks” [2022]
[3] Bandi Jaswanth et. al., “Stacked LSTM a Deep Learning model to predict Stock market” [2022]
[4] K. Ritwik Reddy et. al. “Stock Market Prediction Using Recurrent Neural Network” [2022]
[5] Pushpendra Singh Sisodia et. al., “Stock Market Analysis and Prediction for Nifty50 using LSTM Deep Learning Approach” ,[2022]
[6] Anshuman Behera et. al., “Stock Price Prediction using Machine Learning” [2022]

[7] Surendra Kumar Shukla et. al., “Stock Market Prediction Using Deep Learning” [2022]
[8] Shoban Dinesh et. al., “Prediction of Trends in Stock Market using Moving Averages and Machine Learning” [2021]
[9] Julee kumari et. al., “Prediction of stock price using machine learning Techniques : A survey” [2021]
[10] Shruti Goswami et. al., “stock market prediction using deep learning LSTM model” [2021]
[11] Mojtaba nabipour et. al., “Predicting stock market trends using machine learning and deep learning via continuous and binary data ; a comparative analysis” [2020]
[12] Obthong et. Al., “A survey on machine learning for stock price prediction: algorithms and techniques” [2020]
[13] Suellen Teixeira et.al., “Comparing Artificial Neural Network Architectures for Brazilian Stock Market Prediction” [2020]
[14] X. Pang et. al., “‘An innovative neural network approach for stock market prediction” [2020]
[15] M. Nabipour et. al., “Deep learning for Stock Market Prediction” [2020]