Report on

STOCK MARKET PREDICTION

Semester IV

M.Sc. Computer Science (Data Analytics)

(2021-2023)

SUBMITTED BY

AMAL JOHN

2117004

UNDER THE GUIDANCE OF

Mr. SHIJU THOMAS M.Y

JANUARY 2023

DEPARTMENT OF COMPUTER SCIENCE

RAJAGIRI COLLEGE OF SOCIAL SCIENCES
(AUTONOMOUS)

KALAMASSERY- 683104

CERTIFICATE

*This is to certify that the project titled "**Stock Market Prediction**" is a bonafide work carried out by **Amal John (Reg No:2117004)** in partial fulfillment of therequirements for the award of the Master's degree in Computer Science (Data Analytics) of Rajagiri College of Social Sciences (Autonomous), affiliated to Mahatma Gandhi University,during the year 2021- 2023. This project report has been approved as it satisfies the academic requirement of project work prescribed for the M.Sc. Computer Science (Data Analytics).*

Mr. Shiju Thomas M.Y
Project Guide

Dr. Bindiya M Varghese
Dean- Computer Science

Examiner –I **(Seal)** **Examiner -II**

Place:

Date:

DECLARATION

I hereby declare that this project work and the report submitted to the Department of Computer Science, Rajagiri College of Social Sciences (Autonomous), Kalamassery in partial fulfilment of the award of degree of M.Sc. Computer Science (Data Analytics) is an outcome of my own work.

To the best of my knowledge this project work or parts there, does not form a part of any other project work or thesis on the basis of which a degree or award was conferred on an earlier occasion.

Date Amal John

Place: (Reg No: 2117004)

ACKNOWLEDGEMENT

I consider it as privilege to express my sincere gratitude and respect to all those who guided and inspired me in the successful completion of this internship work. I convey my reverential salutation to Almighty God for enabling me to take up and complete the internship.

I would like to express my sincere thanks to Dr. Binoy Joseph, Principal of Rajagiri College of Social Sciences (Autonomous), for providing the necessary infrastructure and support for the completion of this internship work.

I would like to express my sincere thanks to Dr. Bindiya M Varghese, Dean - Department of Computer Science, Rajagiri College of Social Sciences (Autonomous) for her valuable advice and support which have helped me greatly in the accomplishment of the internship.

I would like to express my gratitude to my external guide, Mrs. Siri siripurapu, Project Manager and CEO, of ICE Data Services, for her highly valuable advice, support, and inspiration during the internship.

I sincerely thank my internship guide and coordinator, Mr. Shiju Thomas, Department of Computer Science, Rajagiri College of Social Sciences (Autonomous), for his consistent guidance and inspiration throughout the period for the completion of this internship.

I would like to thank all the teaching and non-teaching staff of Rajagiri College of Social Sciences (Autonomous) for their valuable guidance and suggestions rendered during the internship work.

Finally, I thank my parents and all my friends for their help, encouragement, and moral support during this work. Once again, thank you all.

Abstract

Stock market plays an important role in a country's economy by providing capital to companies, generating returns for investors, reflecting market valuation, and supporting fiscal policy. A healthy and well-performing stock market can contribute to the overall economic growth and development of a country. The stock market provides a platform for companies to raise capital by issuing shares to investors, thus resulting the capital formation for a company. The stock market reflects the value of publicly traded companies, which provides an important measure of the health and performance of the economy. When the stock market is performing well, it can signal positive economic growth and investor confidence, which can attract foreign investment and boost the country's economic prospects. Overall, predicting the stock market is an important tool for investors, policymakers, and companies, as it can provide valuable insights into the direction of the economy. This project aims to forecast various aspects of Indian and US stock market using different methods. It involves understanding the relationship between micro economic features and the in the market, predict the market using various algorithms and to classify the market using news paper headlines.

Table of Contents

Table of Figures

Table of Tables

Chapter 1 Introduction

1.1 Introduction

Stock market is an organized financial market where brokers and traders buy and sell stocks, bonds and securities. There are two stock exchanges in India i.e. National Stock Exchange (N.S.E.) and Bombay Stock Exchange (B.S.E.). National Stock exchanges incorporated 1992 and Bombay Stock exchange in 1875. Major trading is done in NSE and thus it is in the leading position. The US stock market is made up of several major exchanges, including the New York Stock Exchange (NYSE), the Nasdaq Stock Market, and the Chicago Stock Exchange. These exchanges operate as regulated markets, where buyers and sellers meet to trade stocks through brokers who represent them. The aim of the paper is to predict the Indian and US Stock Market using ML and AI algorithms. In this present situation of Indian Stock market each and every investors are required to watch closely to the ongoing activities in both the national and international markets. It is very important that each and every investors must be aware about the major factors responsible for the up and downs of the market.

1.2 Organization Profile

ICE Data Services is a global financial technology and data provider with a presence in Hyderabad, India. The Hyderabad office is part of ICE Data Services' broader global network of offices, which includes locations in New York, London, Chicago, and other major financial centers.

Intercontinental Exchange (ICE) is an American company founded in 2000 that operates worldwide financial exchanges and clearing houses, providing mortgage technology, data, and listing services. With its presence on the Fortune 500, S&P 500, and Russell 1000 lists, ICE owns exchanges for financial and commodity markets and runs 12 regulated exchanges and marketplaces. These include ICE futures exchanges in the US, Canada, and Europe, the Life futures exchanges in Europe, the New York Stock Exchange, equity options exchanges, and OTC energy, credit, and equity markets.

Overall, ICE Data Services' Hyderabad office plays an important role in the company's global operations, helping to provide clients with the financial data and tools they need to make informed decisions and stay ahead in an ever-changing financial landscape.

Project Scope

Stock market prediction is used for a variety of reasons, including:

Investment decision making: Investors use stock market predictions to make informed investment decisions. If they can accurately predict the future movement of the market, they can buy or sell stocks at the right time to maximize profits.

Risk management: Stock market predictions can help investors manage their risks by identifying potential market downturns or identifying sectors that are likely to perform well.

Portfolio management: Stock market predictions can help portfolio managers optimize their investment portfolios by adjusting the allocation of assets based on market predictions.

Economic analysis: Stock market predictions can provide insights into the overall health of the economy. By analyzing trends and patterns in the stock market, economists can make predictions about the future direction of the economy.

Trading strategies: Traders use stock market predictions to develop trading strategies, such as momentum trading or contrarian trading, that are designed to take advantage of market trends.

Overall, stock market prediction is a useful tool for investors, portfolio managers, traders, and economists to make informed decisions about investments and manage their risks.

1.3 Project Objectives

The objectives of the project are:

Section 1

This examines the microeconomic factors that have the strongest correlation with the stock market. Specifically, it considers the impact of oil prices, bond prices, repo rates, USD to INR exchange rates, inflation rates, and gold prices to determine which factors have the most significant influence on the market.

Section 2

Second objective aims is to use various methods to predict the stock market with precision and identify the most effective model.

Section 3

Third objective aims to predict the stocks by analyzing the daily after market reports.

Chapter 2 Literature Review

2.1 Related Works

The proposed project is in the field of Stock market prediction.

The research paper, **Determinants of Equity Share Prices in India: A Panel Data Approach**

By P. Srinivasan: The purpose was to analyze the factors affecting the Indian stock market. Here six major sectors in the market were taken under consideration, it provided how the fundamental factors affect the share price. For this study, multiple Regression models are used to determine the stock prices. The article is concluded by discussing the impact of fundamental factors on the stock pieces, It was found that dividend has influenced the market price significantly and the earnings per share and price-earnings ratio are crucial determinants of share prices.

The research paper titled, **Factors Affecting Stock Market in India**.

By Dr. Perways Alam: the article aims to identify the major factors responsible for the volatility of the stock market in India. It speaks about various marketable and non-marketable investments.

The research paper titled, **Relevance of Macro Economic factors for the Indian stock market**

By Urvin Shah: The purpose of the article is to identify the Relevance of Macro Economic factors in the Indian stock market. The research attempts to establish a relationship between changes in macroeconomic factors and stock market returns in both long-term and short-term investments, Here error correction mechanism is used to examine the relationship. The paper concludes by saying that Industrial Production and interest rate have a significant effect on the stock market.

The research paper titled, **Factors Affecting Stock Prices in India: A Time Series Analysis.**

By Arpit Bhargava, Ankush Bhargava, and Surbhi Jain: The research was done to understand the relationship between macro variables and Stock prices using time series

regression models and it was found that only exchange rate and oil prices and Inflation have a significant impact on the stock market prices

The research paper titled, **STOCK TREND PREDICTION USING NEWS SENTIMENT ANALYSIS**.

By Kalyani Joshi, Prof. Bharathi H.N, and Prof. Jyothi Rao: This paper aims to predict future stock trends using news articles and financial reports with the assumption that there is a relation between stock price and news. Here the researchers have created different classification models which depict the polarity of the news article being positive or negative.

The research paper titled, **Stock market prediction using machine learning classifiers and social media, news**

By Wasiat Khanq, Mustansar Ali Ghazanfar, Muhammad Awais Azam, Amin Karami, Khaled H. Alyoubi, Ahmed S. Alfakeeh: This paper aims to predict the stock market based on social media news. In this study, they propose a system where stock prediction is done based on social media, financial news, or both. Here the sentimental analysis of processed tweets and financial news is performed using Sandford sentimental analysis also twelve machine learning classifiers are selected and compared in terms of prediction performance. These classifiers are trained and tested on the final data set to get the feature stock market trends. The results show that the highest prediction accuracy of 80.53% and 75.16% are achieved using social media and financial news respectively.

2.2 Stock market Terms and Definitions

Stock Market is a vast and specialized domain, so it is imperative to know the definitionsof terms related with it. This section describes the terms related to stock Market and economic domain which will be used throughout this report.

Nifty 50 – Nifty 50, also known as the National Stock Exchange Fifty, is an index of 50 large-cap Indian companies listed on the National Stock Exchange (NSE). It is one of the two main stock indices in India, the other being the BSE Sensex. The Nifty 50 index represents the performance of some of the most prominent and financially stable companies across various sectors, including banking, finance, energy, information technology, consumer goods, and services.

NYSE – The New York Stock Exchange (NYSE) is the world's largest stock exchange by market capitalization and is located on Wall Street in New York City, United States. The NYSE was founded in 1792 and is one of the oldest stock exchanges in the world. The NYSE lists over 2,400 companies, including many of the largest and most well-known companies in the world, such as Apple, Microsoft, Coca-Cola, and ExxonMobil.

DJIA – DJIA stands for Dow Jones Industrial Average, which is one of the oldest and most widely used stock market indices in the United States. It was created in 1896 by Charles Dow, the founder of Dow Jones & Company, and it consists of 30 large-cap US stocks across a range of industries. The DJIA includes some of the most well-known and financially significant companies in the world, such as Apple, Microsoft, Coca-Cola, and Goldman Sachs.

Open – In the context of the stock market, the term "open value" typically refers to the opening price of a stock or other financial instrument at the start of a trading session. This is the price at which the first transaction occurs for that day, and it is determined by a combination of factors, including the previous day's closing price, any after-hours trading activity, news or announcements that may affect the stock, and overall market conditions.

Close - In the context of the stock market, the term "close value" typically refers to the closing price of a stock or other financial instrument at the end of a trading session. This is the final price at which the stock is traded for the day, and it is an important indicator of the stock's performance for that trading day.

Volume – In the context of stock market data, the term "volume" refers to the total number of shares of a particular stock that were traded during a specified period of time, such as a trading session or a trading day. Volume is an important metric for investors and traders, as it provides insight into the level of demand for a particular stock.

Turnover – In the context of stock market data, the term "turnover" refers to the total value of all shares of a particular stock that were traded during a specified period of time, typically a trading session or a trading day. Turnover is calculated by multiplying the total number of shares traded by the price of the stock at the time of the trades.

Crude oil rate – Crude oil rate refers to the price of crude oil in the global market. Crude oil is a major commodity that is used in many industries, including transportation, manufacturing, and energy production. As a result, changes in the price of crude oil can have a significant impact on the global economy and financial markets, including the stock market.

Inflation rate – Inflation rate refers to the rate at which prices for goods and services in an economy are increasing over time. Inflation is typically measured using an index such as the Consumer Price Index (CPI) or the Producer Price Index (PPI), which track the prices of a basket of goods and services over time. Inflation can have a significant impact on the stock market. When inflation is high, the purchasing power of consumers decreases, as the prices of goods and services increase. This can lead to a decrease in consumer spending, which can negatively impact corporate earnings and stock prices.

Gold Rate – Gold rate refers to the price of gold, The price of gold can have a significant impact on the stock market, particularly on sectors such as mining, precious metals, and jewelry. When the price of gold is high, companies in these sectors typically benefit, as they can charge higher prices for their products and services.

USD vs INR – USD vs INR refers to the exchange rate between the US dollar and the Indian rupee. The exchange rate between two currencies is determined by a variety of factors, including global supply and demand for the currencies, interest rate differentials, inflation rates, and geopolitical events. Changes in the USD vs INR exchange rate can have a significant impact on the Indian stock market, particularly on sectors that are heavily dependent on exports or imports. A weaker rupee, which means that it takes more rupees to purchase a US dollar, can make Indian exports more

competitive in global markets, which can benefit companies that rely on exports for their revenues. Conversely, a stronger rupee can make Indian imports cheaper, which can benefit companies that rely on imports for their raw materials or other inputs.

Bond Price – A bond price refers to the market value of a bond, which is a debt instrument that is issued by companies, governments, or other organizations to raise capital. When an organization issues a bond, it promises to pay the bondholder a fixed amount of interest over a specified period of time, and to repay the principal amount of the bond at maturity. Changes in bond prices can have a significant impact on the stock market, particularly on sectors that are sensitive to interest rate changes, such as financials, utilities, and real estate. When bond prices rise, it typically means that interest rates have fallen. This can benefit companies that borrow money to finance their operations, as they can access capital at a lower cost. Lower interest rates can also stimulate consumer spending and investment, which can benefit the overall economy and stock market.

Repo Rate – Repo rate, short for repurchase rate, is the interest rate at which the central bank of a country lends money to commercial banks. In India, the repo rate is set by the Reserve Bank of India (RBI). The repo rate can have a significant impact on the stock market, as it influences the cost of borrowing for banks and other financial institutions. When the RBI lowers the repo rate, it becomes cheaper for banks to borrow money from the central bank, and they may pass on these lower borrowing costs to consumers and businesses. This can stimulate borrowing and spending, which can lead to increased demand for goods and services, and potentially higher revenues and profits for companies. In turn, this can drive up stock prices, particularly for sectors that are sensitive to changes in interest rates, such as financials and real estate.

High - In stock data, a "high value" typically refers to the highest price that a particular stock has reached within a specified time period, such as a day, week, month, or year. This information is often provided along with other data points such as the stock's opening price, closing price, and volume of shares traded. High values are important for investors and analysts as they can provide insight into the performance of a company's stock over time and can help identify trends and patterns in the market.

Low - In stock data, a "low value" typically refers to the lowest price that a particular stock has reached within a specified time period, such as a day, week, month, or year. This information is often provided along with other data points such as the stock's opening price, closing price, and volume of shares traded. Low values are important for investors and analysts as they can provide insight into the performance of a company's stock over time and can help identify trends and patterns in the market.

Chapter 3 Data Analytics

Experiment 1

3.1 Dataset and Data Preprocessing for Experiment 1

3.1.1 Problem Statement

The problem statement of the project is to examines the microeconomic factors that have the strongest correlation with the stock market. Specifically, it considers the impact of oil prices, bond prices, repo rates, USD to INR exchange rates, inflation rates, and gold prices to determine which factors have the most significant influence on the market. . The problem statement can be broken down into following steps as in Figure 3.1.

Figure 1 Flowchart of Experiment 1

1. Initially, a comprehensive analysis was carried out on the Indian stock market to identify the various factors that influence it. Subsequently, multiple consultations were held with market experts to gain further insights into the microeconomic factors that impact the Indian stock market.
2. Various websites and other resources were utilized to collect data, with the majority of the data being obtained from Yahoo Finance, RBI (Reserve Bank of India), and NSE (National Stock Exchange).
3. The data was initially collected from various websites and consolidated into a single dataset. The collected data was then preprocessed to prepare it for further analysis. This step is essential to ensure that the data is of high quality and can be used to generate meaningful insights. Once the data was preprocessed, it was ready for further analysis.
4. To gather insights from the data, Granger causality was applied to the dataset, with the price of Nifty 50 being used as the target variable. Each factor was evaluated

against Nifty 50 to draw conclusions about their impact. This analysis helped to identify causal relationships between the various factors and the performance of Nifty 50.

5. For visualizing the results, Looker Studio was utilized, which provided the ability to filter the data based on dates. This allowed for the results to be displayed graphically, which made it easier to understand and interpret the findings.

3.1.2 Dataset

The datasets used for the analysis is extracted from the Yahoo Finance Website(https://finance.yahoo.com/).

It is a network that provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content.

Following are the data used for analysis:

1. ***NIFTY 50*** **-** The Nifty 50 is an index of the top 50 publicly traded companies on the National Stock Exchange of India (NSE). It serves as a benchmark index for the Indian stock market, providing a measure of the overall performance of the largest companies traded on the exchange.

Table 1 Nifty 50 dataset

Date	Open	High	Low	Close
1/11/2018	12060.79981	12178.79981	12045.4502	12103.09961
1/12/2018	12144.79981	12146.75	12004.40039	12099.2002
1/15/2018	12103.5	12214.65039	11971.9502	11992.79981
1/16/2018	12036.54981	12487.25	12027.29981	12430.9502
1/17/2018	12473.04981	12653.25	12432.75	12531.40039
1/18/2018	12565.84961	12605.15039	12431.79981	12570.7002
1/19/2018	12577.25	12668.54981	12485.29981	12633.84961
1/22/2018	12544.59961	12924.59961	12521.15039	12894.9502
1/23/2018	12919.29981	13084.2002	12898.25	13048.90039

2. ***OIL PRICE*** - The price of oil represents the market value at which oil is bought and sold internationally. As a key component of the global supply chain, fluctuations in oil prices can significantly impact supply chain management and have a notable influence on the broader market.

Table 2 Oil price dataset

Date	Open_crude_oil_price	High_crude_oil_price	Low_crude_oil_price	Close_crude_oil_price
2-Jan-18	60.2	60.74	60.1	60.37
3-Jan-18	60.39	61.97	60.28	61.63
4-Jan-18	61.96	62.21	61.59	62.01
5-Jan-18	61.9	62.04	61.09	61.44
8-Jan-18	61.61	61.97	61.34	61.73
9-Jan-18	61.92	63.48	61.8	62.96
10-Jan-18	63.41	63.67	63.09	63.57
11-Jan-18	63.5	64.77	63.43	63.8
12-Jan-18	63.57	64.5	63.06	64.3

3. ***BOND PRICE*** **-** The S&P BSE Bond Price Index is a stock market index that tracks the performance of bond securities listed on the Bombay Stock Exchange (BSE) in India. Bond price is directly related to the interest rates in the market and thus it has an indirect relation with the inflation rates.

Table 3 Bond price dataset

Date	S&P BSE India Bond Index
2-Jan-18	146.12
3-Jan-18	146.16
4-Jan-18	146.26
5-Jan-18	146.45
8-Jan-18	146.87
9-Jan-18	146.72
10-Jan-18	146.36
11-Jan-18	146.4
12-Jan-18	146.29

4. ***REPO RATE*** **-** It is the rate at which commercial banks borrow money from the central bank of a country. The central bank uses this tool as a means to control the money supply and control inflation by raising or lowering the repo rate.

Table 4 RBI Repo Rate dataset

Date	RBI Repo Rate
1/2/2018	6
1/3/2018	6
1/4/2018	6
1/5/2018	6
1/8/2018	6
1/9/2018	6
1/10/2018	6
1/11/2018	6
1/12/2018	6

5. ***EXCHANGE RATE -*** The exchange rate between the United States dollar (USD) and the Indian Rupee (INR) is the rate at which one currency can be exchanged for another. The value of the USD/INR currency pair fluctuates based on the supply and demand of the two currencies in the foreign exchange market.

Table 5 Exchange Rate dataset

Date	Open_usd_vs_inr	High_usd_vs_inr	Low_usd_vs_inr	Close_usd_vs_inr
1-Jan-18	63.840801	63.869999	63.66	63.840801
2-Jan-18	63.867599	63.867599	63.419998	63.867599
3-Jan-18	63.380001	63.580002	63.380001	63.459999
4-Jan-18	63.419102	63.59	63.330002	63.419102
5-Jan-18	63.369598	63.419998	63.209999	63.369598
8-Jan-18	63.264999	63.529999	63.23	63.264999
9-Jan-18	63.451801	63.73	63.419998	63.452
10-Jan-18	63.635601	63.84	63.568001	63.636101
11-Jan-18	63.759998	63.84	63.634998	63.759998

6. ***GOLD PRICE -*** The price of gold refers to the cost of one ounce of gold in the global market.

Table 6 Gold price Dataset

Date	Open	High	Low	Close
29-Dec-22	1,805.80	1,819.50	1,805.80	1,819.50
28-Dec-22	1,803.20	1,807.90	1,803.10	1,807.90
27-Dec-22	1,803.40	1,826.30	1,803.40	1,814.80
23-Dec-22	1,794.30	1,802.80	1,794.10	1,795.90
22-Dec-22	1,818.10	1,818.10	1,785.00	1,787.00
21-Dec-22	1,817.40	1,821.40	1,814.40	1,815.90
20-Dec-22	1,786.50	1,821.40	1,786.50	1,815.90
19-Dec-22	1,791.60	1,796.30	1,785.60	1,787.70
16-Dec-22	1,777.90	1,791.40	1,777.90	1,790.00
15-Dec-22	1,801.00	1,801.00	1,774.80	1,777.20

7. ***INFLATION RATE -*** Inflation rate refers to the rate at which the general level of prices for goods and services in an economy is increasing over a period of time.

Table 7 Inflation Rate

Date	inflation
1/1/2018	3.02
2/1/2018	2.74
3/1/2018	2.74
4/1/2018	3.62
5/1/2018	4.78
6/1/2018	5.68
7/1/2018	5.27
8/1/2018	4.62
9/1/2018	5.22

3.2 Data Modeling and Analysis for Experiment 1

3.2.1 Data Exploration and Analysis

Exploring historic stock market data of Nifty 50 can provide valuable insights into the behavior and performance of the Nifty 50 index, which is comprised of the top 50 companies listed on the National Stock Exchange (NSE) in India. Here are some of the uses of historic stock data exploration of Nifty 50:

1. Analyzing Trends and Patterns: By analyzing historic stock data of Nifty 50, investors can identify trends and patterns in the market, such as which sectors or industries are performing well or which stocks are showing signs of volatility. This can help investors make informed investment decisions.

2. Conducting Market Research: Historic stock data of Nifty 50 can be used for market research, allowing investors to study past trends and patterns to gain insights into future market behavior. This can help investors identify potential investment opportunities and make more informed investment decisions.

3. Assessing Risk: Historic stock data of Nifty 50 can also be used to assess risk. By studying the performance of the index during past market downturns, investors can better understand its risk profile and make more informed investment decisions.

Microeconomic factors can have a significant impact on the performance of individual companies, which in turn can affect the broader stock market index like Nifty 50. Nifty 50 is composed of the top 50 companies listed on the National Stock Exchange of India (NSE), so any significant changes in the performance of these companies can affect the index as a whole. Therefore, any changes in these microeconomic factors can have a direct impact on the performance of individual companies and, in turn, the broader Nifty 50 index.

Analyzing the dataset: The dataset for objective one contains 1022 entries with seven attributes, where one attribute ('Date') is of type 'datetime64[ns]' and the remaining attributes are of type 'float64'

.

#	Column	Non-Null Count	Dtype
0	Date	1022 non-null	datetime64[ns]
1	Close_nifty_50	1022 non-null	float64
2	gold_price	1022 non-null	float64
3	Close_usd_vs_inr	1021 non-null	float64
4	bond_index.S&P BSE India Bond Index	1022 non-null	float64
5	ripo.RBI Repo Rate	1022 non-null	float64
6	oil.oil_price.Close_crude_oil_price	1022 non-null	float64

Figure 9 Data information

visualization of attributes of the data: This graph displays the relationship between the Nifty 50 index and several other factors, including the crude oil rate, bond index, repo rate, USD vs INR exchange rate, and gold price..

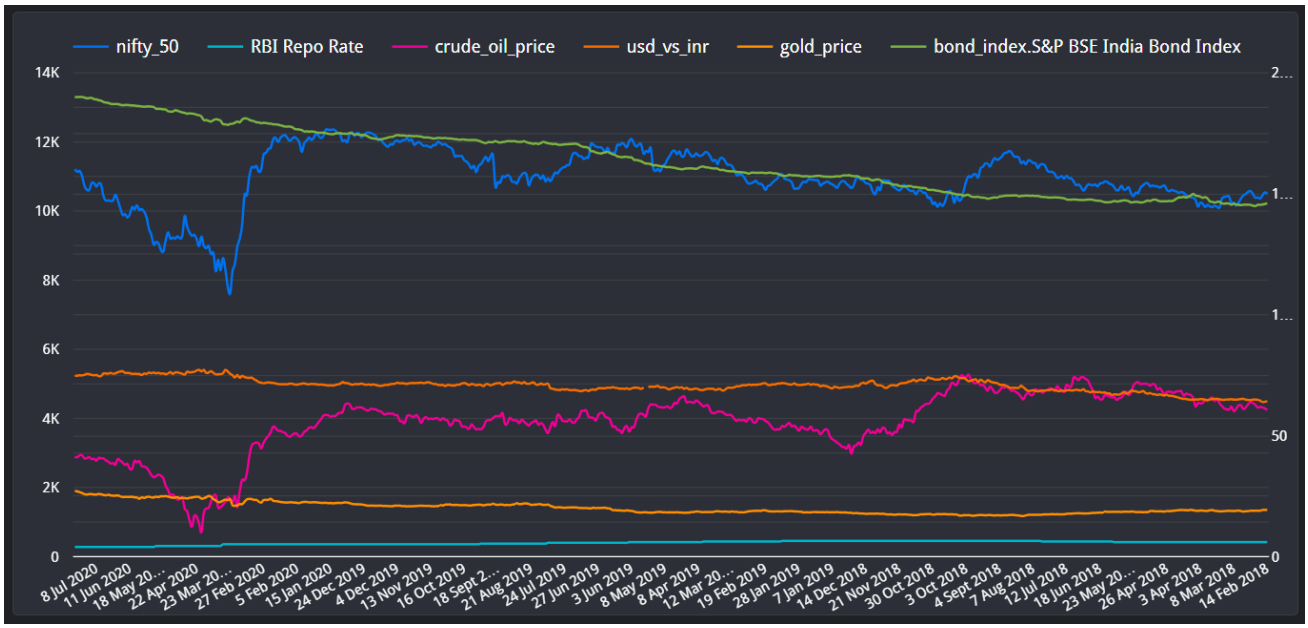


Figure 10 Visualizing of attributes

Data Modeling

The analysis of the Nifty 50 and the factors, crude oil price, Exchange rate of usd vs inr, gold price, bond index here, is done on the basis of 'Close' attribute. It is because all the other attributes except the 'Close' value can be affected or fluctuated due to certain factors such as After Market Orders(AMO),option traders, day traders etc whereas the 'Close' value does not.

Hypothesis testing - Granger causality Testing

Granger causality is a technique used in statistics to assess if one time series is helpful in predicting another. It is based on the principle that if past observations of a time series (X) can be used to make accurate predictions about the current value of another time series (Y), then X is considered to have a causal effect on Y. In simpler terms, it is a method of determining if there is a cause and effect relationship between two variables. This is usually done by utilizing a linear regression model and determining if past values of X have a significant influence on the forecasting of current values of Y.

P-value: The p-value is a measure of the probability that the observed test statistic would occur by chance if the null hypothesis is true. A low p-value (typically less than 0.05) suggests that the null hypothesis should be rejected and that past values of X do have a significant impact on the prediction of current values of Y.

Deployment & Optimization

The Granger Causality Test is done between the close price of Nifty 50 and the Close price of the factors, crude oil price, Exchange rate of usd vs inr, gold price and bond index. The close of Nifty 50 is also tested against the Repo rate.

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=3.9121  , p=0.0482  , df_denom=1017, df_num=1
ssr based chi2 test:   chi2=3.9237  , p=0.0476  , df=1
likelihood ratio test: chi2=3.9162  , p=0.0478  , df=1
parameter F test:         F=3.9121  , p=0.0482  , df_denom=1017, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=5.3455  , p=0.0049  , df_denom=1014, df_num=2
ssr based chi2 test:   chi2=10.7437 , p=0.0046  , df=2
likelihood ratio test: chi2=10.6874 , p=0.0048  , df=2
parameter F test:         F=5.3455  , p=0.0049  , df_denom=1014, df_num=2
```

Figure 11 Test between Nifty 50 and Gold Price

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=3.4608  , p=0.0631  , df_denom=1017, df_num=1
ssr based chi2 test:   chi2=3.4710  , p=0.0625  , df=1
likelihood ratio test: chi2=3.4651  , p=0.0627  , df=1
parameter F test:         F=3.4608  , p=0.0631  , df_denom=1017, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=1.7411  , p=0.1759  , df_denom=1014, df_num=2
ssr based chi2 test:   chi2=3.4993  , p=0.1738  , df=2
likelihood ratio test: chi2=3.4933  , p=0.1744  , df=2
parameter F test:         F=1.7411  , p=0.1759  , df_denom=1014, df_num=2
```

Figure 12 Test between Nifty 50 and Exchange Rate

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=3.1528  , p=0.0761  , df_denom=1017, df_num=1
ssr based chi2 test:   chi2=3.1621  , p=0.0754  , df=1
likelihood ratio test: chi2=3.1572  , p=0.0756  , df=1
parameter F test:         F=3.1528  , p=0.0761  , df_denom=1017, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=1.6084  , p=0.2007  , df_denom=1014, df_num=2
ssr based chi2 test:   chi2=3.2326  , p=0.1986  , df=2
likelihood ratio test: chi2=3.2275  , p=0.1991  , df=2
parameter F test:         F=1.6084  , p=0.2007  , df_denom=1014, df_num=2
```

Figure 13 Test between Nifty 50 and Repo Rate

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=0.0881  , p=0.7667  , df_denom=1017, df_num=1
ssr based chi2 test:   chi2=0.0884  , p=0.7663  , df=1
likelihood ratio test: chi2=0.0884  , p=0.7663  , df=1
parameter F test:         F=0.0881  , p=0.7667  , df_denom=1017, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=7.0811  , p=0.0009  , df_denom=1014, df_num=2
ssr based chi2 test:   chi2=14.2321 , p=0.0008  , df=2
likelihood ratio test: chi2=14.1336 , p=0.0009  , df=2
parameter F test:         F=7.0811  , p=0.0009  , df_denom=1014, df_num=2
```

Figure 14 Test between Nifty 50 and oil Price

From the ‘P - value’ of Nifty 50 and oil Price it is clear that the second lag value is less than 0.05, Hence we could say that Nifty 50 and oil price is highly correlated than any other factors under consideration.

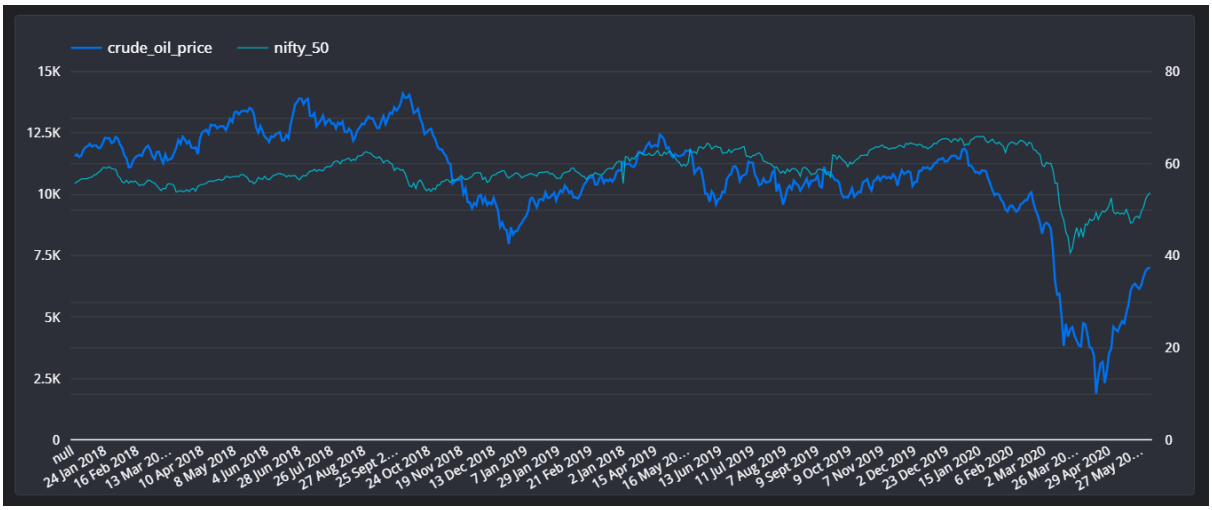


Figure 15 Graph of Nifty 50 and Oil Price

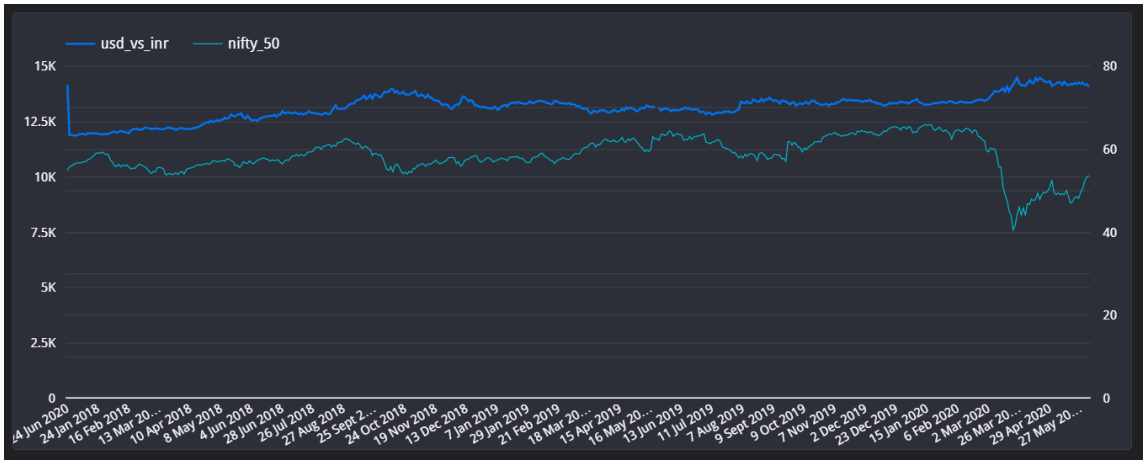


Figure 16 Graph of Nifty 50 and USD vs INR

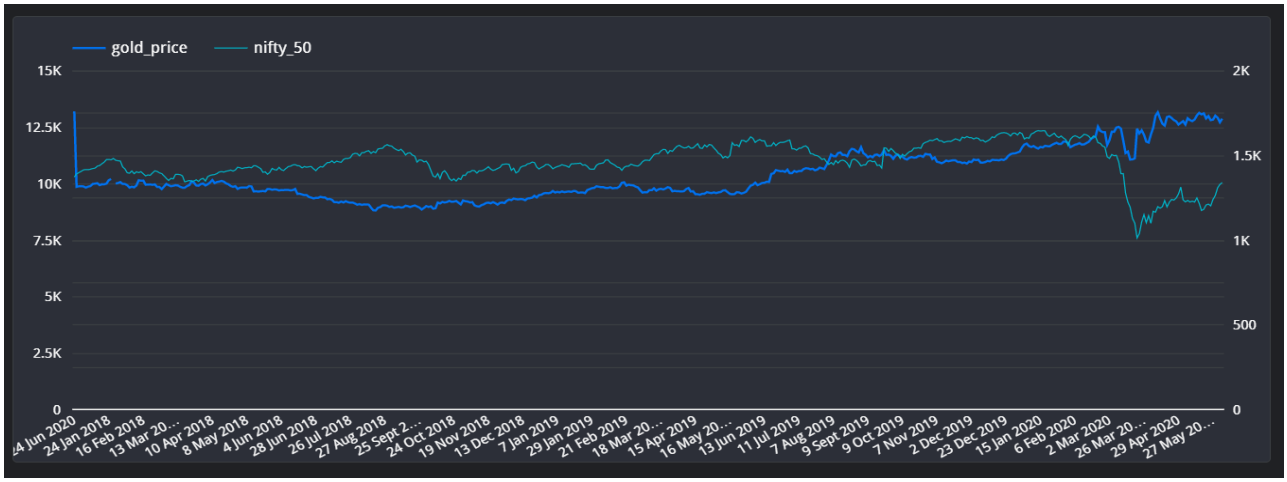


Figure 17 Graph of Nifty 50 and Gold Price

The aim is to determine if certain macroeconomic factors can account for the pricing dynamics of the Indian stock market through the conduct of relevant analyses. The macroeconomic variables are , crude oil price, Exchange rate of USD vs INR, gold price, bond index, inflation rate and Repo rate. The study observes that three out of six factors are relatively more significant and likely to influence the pricing of Indian stock market. The study reveals that the crude oil price, USD vs INR and gold price have significant influence on the close value of Nifty 50.

Experiment 2

3.3 Dataset and Data Preprocessing for Experiment 2

3.3.1 Problem Statement

The objective of the problem statement is to use various methods to predict the stock market with precision and identify the most effective model.

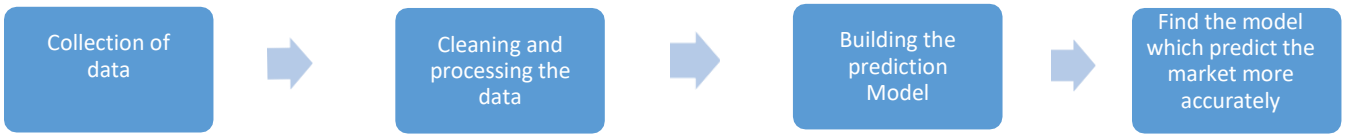


Figure 18 : Flowchart of Experiment 2

1. Extraction of data: To extract the Stock market (Nifty 50) data - utilized the Yahoo Finance website and NES for downloading data in order to extract information on the Nifty 50 stock market.
2. Cleaning and Preprocessing: The extracted data is then cleaned and processed.
3. Several research papers were consulted to identify the optimal model and method for predicting the Nifty 50 stock market. Based on this, a high-accuracy model was constructed for market prediction.
4. The proposed model's accuracy was compared to that of other models to determine which one was more accurate..

Dataset

To extract the Stock market (Nifty 50) data - utilized the Yahoo Finance website and NES for downloading data in order to extract information on the Nifty 50 stock market.

It is a network that provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content.

Following are the data used for analysis:

NIFTY 50 - The Nifty 50 is an index of the top 50 publicly traded companies on the National Stock Exchange of India (NSE). It serves as a benchmark index for the Indian stock market, providing a measure of the overall performance of the largest companies traded on the exchange.

Date	Open	High	Low	Close	Volume	Turnover
2016-09-15	8743.85	8751.95	8704.35	8742.55	149889703	7.347340e+10
2016-09-16	8780.85	8847.65	8750.50	8779.85	274039556	1.443063e+11
2016-09-19	8788.45	8824.30	8774.20	8808.40	135814177	7.364640e+10
2016-09-20	8816.10	8816.45	8759.30	8775.90	140718486	7.035070e+10
2016-09-21	8790.30	8826.85	8757.30	8777.15	153441330	7.572410e+10

Figure 19 Nifty 50 Dataset

The dataset contains the 1258 rows with 7 columns .

- **Open** – In the context of the stock market, the term "open value" typically refers to the opening price of a stock or other financial instrument at the start of a trading session. This is the price at which the first transaction occurs for that day, and it is determined by a combination of factors, including the previous day's closing price, any after-hours trading activity, news or announcements that may affect the stock, and overall market conditions.
- **Close** - In the context of the stock market, the term "close value" typically refers to the closing price of a stock or other financial instrument at the end of a trading session. This is the final price at which the stock is traded for the day, and it is an important indicator of the stock's performance for that trading day.
- **Volume** – In the context of stock market data, the term "volume" refers to the total number of shares of a particular stock that were traded during a specified period of time, such as a

trading session or a trading day. Volume is an important metric for investors and traders, as it provides insight into the level of demand for a particular stock.

- **High -** In stock data, a "high value" typically refers to the highest price that a particular stock has reached within a specified time period, such as a day, week, month, or year. This information is often provided along with other data points such as the stock's opening price, closing price, and volume of shares traded. High values are important for investors and analysts as they can provide insight into the performance of a company's stock over time and can help identify trends and patterns in the market.

- **Turnover** – In the context of stock market data, the term "turnover" refers to the total value of all shares of a particular stock that were traded during a specified period of time, typically a trading session or a trading day. Turnover is calculated by multiplying the total number of shares traded by the price of the stock at the time of the trades.

- **Low** - In stock data, a "low value" typically refers to the lowest price that a particular stock has reached within a specified time period, such as a day, week, month, or year. This information is often provided along with other data points such as the stock's opening price, closing price, and volume of shares traded. Low values are important for investors and analysts as they can provide insight into the performance of a company's stock over time and can help identify trends and patterns in the market.

3.2 Data Modeling and Analysis for Experiment 2

3.2.1 Data Exploration and Analysis
Dataset Description

The dataset contains 1258 rows with 7 columns

Date	Open	High	Low	Close	Volume	Turnover
2016-09-15	8743.85	8751.95	8704.35	8742.55	149889703	7.347340e+10
2016-09-16	8780.85	8847.65	8750.50	8779.85	274039556	1.443063e+11
2016-09-19	8788.45	8824.30	8774.20	8808.40	135814177	7.364640e+10
2016-09-20	8816.10	8816.45	8759.30	8775.90	140718486	7.035070e+10
2016-09-21	8790.30	8826.85	8757.30	8777.15	153441330	7.572410e+10

Figure 20 Nifty 50 Dataset

Data visualization:

Figure 21 Close of Nifty 50

This figure depicts the graph of the closing values of the Nifty 50, during 2017 to 2021

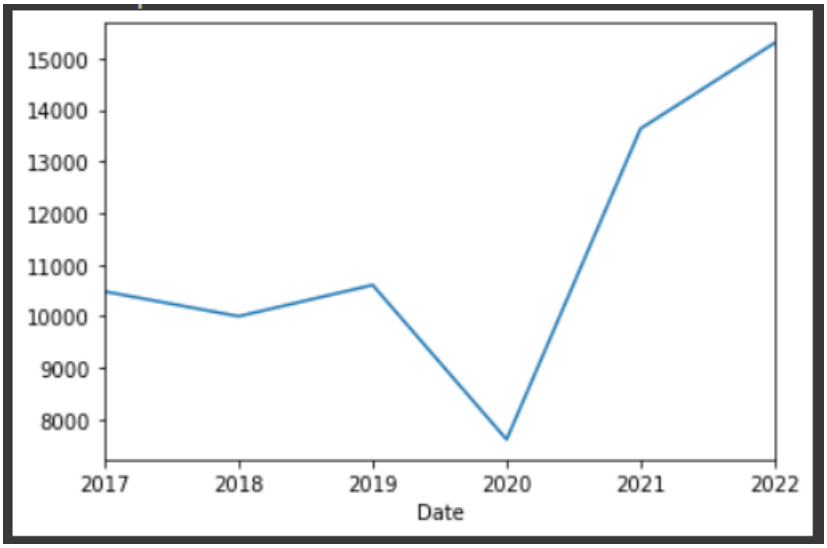


Figure 22 Minimum close value of each year

3.2.2 Modelling

LSTM stands for Long Short-Term Memory, and it is a type of recurrent neural network (RNN) that is commonly used for processing and analyzing sequential data.

LSTM models are designed to overcome the vanishing gradient problem that occurs in traditional RNNs, which can make it difficult for the model to retain information over long periods of time. The LSTM model achieves this by using a series of gates to selectively control the flow of information, allowing it to learn and remember long-term dependencies within the data.

In practical applications, LSTM models have been successfully used for a variety of tasks such as speech recognition, language modeling, and time series prediction, including in the field of finance for stock price forecasting.

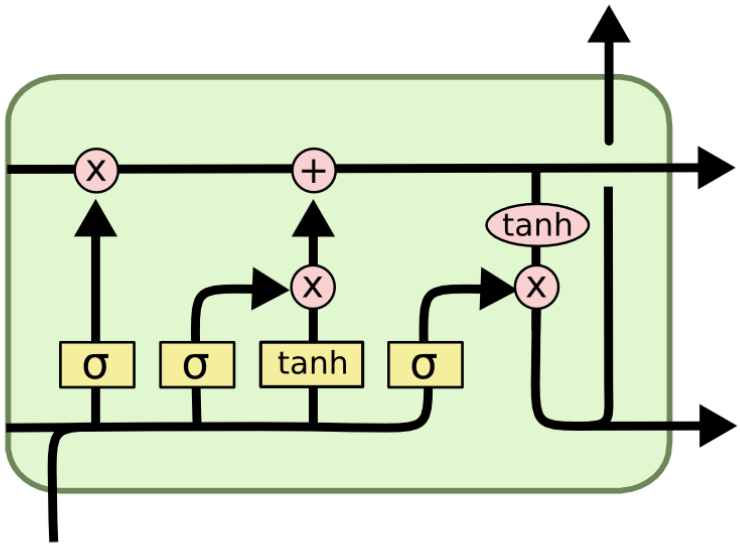


Figure 23 LSTM Model

In a LSTM (Long Short-Term Memory) model, the neurons work together in a complex manner to process sequential data.At each time step, the input to the LSTM model is passed

through a series of gates that determine which information to keep and which information to discard. The three primary gates in a standard LSTM model are:

The forget gate: This gate determines which information from the previous time step should be forgotten, based on the current input.

The input gate: This gate determines which new information should be added to the current memory state.

The output gate: This gate determines which information from the current memory state should be used to make a prediction or generate an output.

Each of these gates has its own set of trainable weights, which are learned during the training process based on the specific task the LSTM model is being used for.

The memory state in an LSTM model is like a conveyor belt, where information flows through and can be added or removed at each time step. This allows the model to capture long-term dependencies in the input data.

Overall, the complex interactions between the gates and the memory state in an LSTM model allow it to effectively process and analyze sequential data, making it a popular choice for many applications in machine learning and artificial intelligence.

The LSTM cell has three gates that help to control the flow of information:

Forget gate: Determines which information from the previous cell state to forget and which information to keep. The forget gate takes as input the previous hidden state h_{t-1} and the current input x_t, and outputs a vector f_t of values between 0 and 1 for each element of the previous cell state c_{t-1}. A value of 1 means to keep the information and a value of 0 means to forget it.

Input gate: Determines which information from the current input to store in the current cell state. The input gate takes as input the previous hidden state h_{t-1} and the current input x_t, and outputs a vector i_t of values between 0 and 1 for each element of the candidate memory vector g_t. A value of 1 means to keep the information and a value of 0 means to discard it.

Output gate: Determines which information from the current cell state to output as the hidden state. The output gate takes as input the previous hidden state h_{t-1} and the current input x_t, and outputs a vector o_t of values between 0 and 1 for each element of the current cell state c_t. The output gate acts as a filter, allowing only relevant information to be passed to the output.

The LSTM cell also has a candidate memory vector g_t and a current cell state c_t. The candidate memory vector is computed using the input gate and the current input x_t, and the current cell state is computed using the forget gate, the input gate, and the candidate memory vector.

Finally, the output of the LSTM cell is the hidden state vector h_t, which is computed by applying the output gate to the current cell state c_t and passing the result through the tanh function to squash the values between -1 and 1.

The equations for an LSTM cell can be written as follows:

Forget Gate: $f_t = \sigma(W_f[x_t, h_{t-1}] + b_f)$
Input Gate: $i_t = \sigma(W_i[x_t, h_{t-1}] + b_i)$
Candidate Memory: $g_t = \tanh(W_g[x_t, h_{t-1}] + b_g)$
Current Memory: $c_t = f_t * c_{t-1} + i_t * g_t$
Output Gate: $o_t = \sigma(W_o[x_t, h_{t-1}] + b_o)$
Hidden State: $h_t = o_t * \tanh(c_t)$
where:

x_t is the input at time step t

h_{t-1} is the hidden state of the previous time step (initially set to 0)

f_t, i_t, and o_t are the forget, input, and output gates respectively

W_f, W_i, W_g, and W_o are the weight matrices for the forget, input, candidate memory, and output gates respectively

b_f, b_i, b_g, and b_o are the bias vectors for the forget, input, candidate memory, and output gates respectively

g_t is the candidate memory vector

c_t is the current memory vector

h_t is the hidden state vector

In summary, the LSTM model is able to selectively remember and forget information over time by using the forget and input gates, while also allowing the network to handle long-term dependencies through the current cell state and candidate memory vector

3.2.3 Model Evaluation

A proposed model for predicting the stock market using LSTM achieved an accuracy score of 96.54%, which was evaluated against the results published in a research paper by the American Journal of Applied Sciences. The paper, titled "Random Forest Based Feature Selection of Macroeconomic Variables for Stock Market Prediction", which used an LSTMRNN model to predict the stock market, which resulted in an accuracy score of 89.57%.

This suggests that the proposed model has achieved a higher level of accuracy than the LSTMRNN model used in the research paper. However, it is important to note that the two models may have been evaluated on different datasets or using different performance metrics, so a direct comparison between the two results may not be entirely accurate. Nonetheless, the high accuracy score achieved by the proposed model using LSTM is promising and indicates that it may be a useful tool for stock market prediction.

Table 7 LSTM Prediction

Model Comparison	Proposed model	LSTMRNN Model
Accuracy	96.54	89.57

Experiment 3

3.4 Dataset and Data Preprocessing for Experiment 3

3.4.1 Problem Statement

The objective of the problem statement is to predict the stock market using news headlines using sentimental analysis.

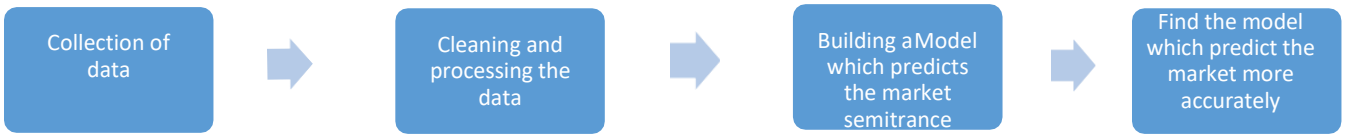
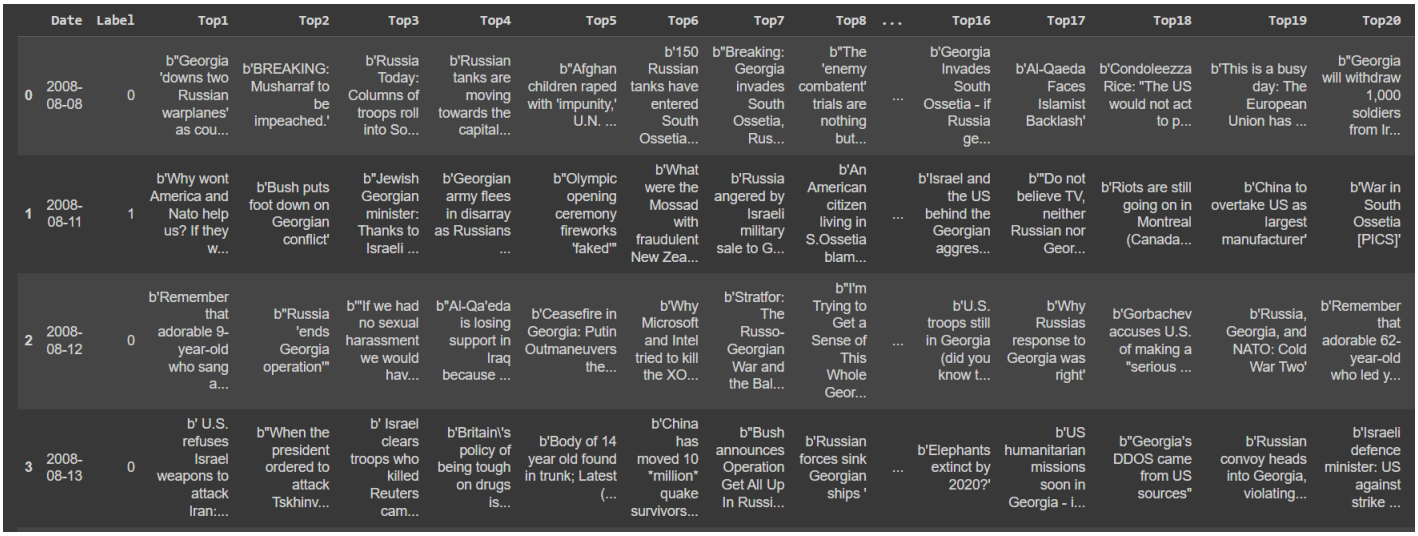


Figure 24 Flowchart of Experiment 3

1. Extraction of data: To extract the Stock market (DJIA) data - utilized the Yahoo Finance website and other resources
2. Cleaning and Preprocessing: The extracted data is then cleaned and processed.
3. Several research papers were consulted to identify the optimal model and method for predicting the market semitrance. Based on this, a high-accuracy model was constructed for prediction.
4. The proposed model's accuracy was compared to that of other models to determine which one was more accurate.

3.4.2 Dataset

The dataset contains the 135 rows with 27 columns



	Date	Label	Top1	Top2	Top3	Top4	Top5	Top6	Top7	Top8	...	Top16	Top17	Top18	Top19	Top20
0	2008-08-08	0	b"Georgia 'downs two Russian warplanes' as cou...	b'BREAKING: Musharraf to be impeached.'	b'Russia Today: Columns of troops roll into So...	b'Russian tanks are moving towards the capital...	b"Afghan children raped with 'impunity,' U.N. ...	b'150 Russian tanks have entered South Ossetia...	b"Breaking: Georgia invades South Ossetia, Rus...	b"The 'enemy combatent' trials are nothing but...	...	b'Georgia Invades South Ossetia - if Russia ge...	b'Al-Qaeda Faces Islamist Backlash'	b'Condoleezza Rice: "The US would not act to p...	b'This is a busy day: The European Union has ...	b"Georgia will withdraw 1,000 soldiers from Ir...
1	2008-08-11	1	b'Why wont America and Nato help us? If they w...	b'Bush puts foot down on Georgian conflict'	b"Jewish Georgian minister: Thanks to Israeli ...	b'Georgian army flees in disarray as Russians ...	b"Olympic opening ceremony fireworks 'faked'"	b'What were the Mossad with fraudulent New Zea...	b'Russia angered by Israeli military sale to G...	b'An American citizen living in S.Ossetia blam...	...	b'Israel and the US behind the Georgian aggres...	b'"Do not believe TV, neither Russian nor Geor...	b'Riots are still going on in Montreal (Canada...	b'China to overtake US as largest manufacturer'	b'War in South Ossetia [PICS]'
2	2008-08-12	0	b'Remember that adorable 9-year-old who sang a...	b"Russia 'ends Georgia operation'"	b"'If we had no sexual harassment we would hav...	b"Al-Qa'eda is losing support in Iraq because ...	b'Ceasefire in Georgia: Putin Outmaneuvers the...	b'Why Microsoft and Intel tried to kill the XO...	b'Stratfor: The Russo-Georgian War and the Bal...	b"I'm Trying to Get a Sense of This Whole Geor...	...	b'U.S. troops still in Georgia (did you know t...	b'Why Russias response to Georgia was right'	b'Gorbachev accuses U.S. of making a "serious ...	b'Russia, Georgia, and NATO: Cold War Two'	b'Remember that adorable 62-year-old who led y...
3	2008-08-13	0	b' U.S. refuses Israel weapons to attack Iran:...	b"When the president ordered to attack Tskhinv...	b' Israel clears troops who killed Reuters cam...	b'Britain\'s policy of being tough on drugs is...	b'Body of 14 year old found in trunk; Latest (...	b'China has moved 10 *million* quake survivors...	b"Bush announces Operation Get All Up In Russi...	b'Russian forces sink Georgian ships '	...	b'Elephants extinct by 2020?'	b'US humanitarian missions soon in Georgia - i...	b"Georgia's DDOS came from US sources"	b'Russian convoy heads into Georgia, violating...	b'Israeli defence minister: US against strike ...

Figure 25 Data set of Objective 3

- **Top** - the Top 1 to top 25 represent the top twenty five news headlines of that day.
- **Label** - The label refers to a categorical variable that represents whether the stock price has increased or decreased compared to the previous day. A value of one typically indicates an increase in stock price, while a value of zero indicates a decrease in stock price.

3.4.3 Modelling

Random Forest Classifier is a supervised machine learning algorithm used for classification tasks. It is an ensemble learning method that builds multiple decision trees and combines their outputs to make a final prediction.

In the Random Forest Classifier algorithm, each decision tree is trained on a randomly selected subset of the training data and a randomly selected subset of the features. This process is known as bagging, and it helps to reduce overfitting by creating diverse trees. The algorithm also uses a technique known as feature importance to determine the most important features for making accurate predictions.To make a prediction using a Random Forest Classifier, the algorithm runs the input data through each of the decision trees and combines their predictions. The final prediction is determined by a majority vote, where the class with the most votes is selected as the final output.

Some advantages of the Random Forest Classifier algorithm include its ability to handle large datasets, its robustness to outliers, and its ability to handle both categorical and continuous data. It is also less prone to overfitting compared to other machine learning algorithms. However, it can be slower to train and predict than other classification algorithms due to the number of decision trees it builds.

Random Forest Classifier is an ensemble learning method that builds multiple decision trees and combines their outputs to make a final prediction. The math behind Random Forest Classifier involves the following steps:Building decision trees: Each decision tree in the Random Forest Classifier algorithm is built using a subset of the training data and a randomly selected subset of the features. The decision tree is built by recursively splitting the data based on the values of the features, with the goal of minimizing the impurity of the resulting subsets. The impurity is measured using metrics such as Gini impurity or entropy.

Combining decision trees: Once the decision trees are built, the Random Forest Classifier algorithm combines their outputs to make a final prediction. The algorithm runs the input data through each of the decision trees and assigns a class label based on the majority vote of the classes assigned by the individual trees.

Feature importance: The Random Forest Classifier algorithm also calculates the importance of each feature for making accurate predictions. This is done by measuring the decrease in impurity that results from splitting the data on each feature. The features with the largest decrease in impurity are considered the most important for making accurate predictions.Tuning hyperparameters: The Random Forest Classifier algorithm has several hyperparameters that can be tuned to optimize its performance. These include the number of decision trees to build, the size of the subset of features to consider for each split, and the depth of the decision trees.

In summary, the math behind Random Forest Classifier involves building multiple decision trees using a subset of the training data and a randomly selected subset of features, combining their outputs to make a final prediction, and calculating the importance of each feature for making accurate predictions. The algorithm can be optimized by tuning its hyperparameters.

3.4.4 Model Evaluation

A proposed model for predicting the stock market using RandomForestClassifier achieved an accuracy score of 85.71%, which was evaluated against the results published in a research paper by the The school of Electrical Engineering Korea University Seoul, Korea. The paper, titled " Stock Prices Prediction using the Title of Newspaper Articles with Korean Natural Language Processing ", which used an artificial neural network CNN model to predict the stock market, which resulted in an accuracy score of 53%.

This suggests that the proposed model has achieved a higher level of accuracy than the CNN model model used in the research paper. However, it is important to note that the two models may have been evaluated on different datasets or using different performance metrics, so a direct comparison between the two results may not be entirely accurate. Nonetheless, the high accuracy score achieved by the proposed model using RandomForestClassifier is promising and indicates that it may be a useful tool for stock market prediction.

Table 8 Randonforest Classifer

Model Comparison	Proposed model	CNN Model
Accuracy	85.71	53

Chapter 4 Experiments and Results Analysis

4.1 Findings and Results

The initial phase of the project involved conducting a detailed study and holding several meetings with market experts to gain insights into various factors that affect the stock market, and then considered a range of factors, including oil prices, bond prices, repo rates, USD to INR exchange rates, inflation rates, and gold prices, and found that oil prices, USD vs INR exchange rates, and gold prices have a significant influence on the market.

The second objective of the project was to identify the most effective model for predicting the stock market with precision and used various methods and concluded that the LSTM model predicted the market more accurately than other models thus compared the accuracy scores of the proposed model with those of other models and found that the proposed model performed better.

The third objective was to predict the stock market using news headlines through sentimental analysis. The Random Forest Classifier model was used, which achieved an accuracy score of 85.71%. The team compared the accuracy scores of different models and concluded that the proposed model was the most accurate in predicting the market.

Overall, the project involved an in-depth study of various factors affecting the stock market, followed by the identification of the most effective models for predicting the market accurately. The findings of the project have important implications for investors, portfolio managers, traders, and economists who rely on accurate stock market predictions to make informed decisions.

Model Recommendation

- Objective one has allowed us to identify the specific microeconomic factors that have an impact on the stock market. By closely analyzing the movements of these factors, we can gain a better understanding of how the global market is responding, and how it is affecting the Indian Stock Market, specifically the Nifty 50. This information is critical in making informed investment decisions, as it helps us to anticipate and respond to market changes. Furthermore, by monitoring these factors over time, we can gain insight into long-term trends and make more strategic investments. Overall, objective one has provided us with a valuable framework for analyzing the complex and ever-changing dynamics of the stock market.
- Through objective two and three, we can obtain a comprehensive understanding of the upcoming market trends and the sentiment of individual investors. By analyzing both of these factors, we can make accurate predictions about the movements of the domestic market. This information is crucial in making informed investment decisions and maximizing returns. Ultimately, objective two and three provide us with the tools necessary to stay ahead of market trends and capitalize on emerging opportunities.

Prediction of Nifty 50 :

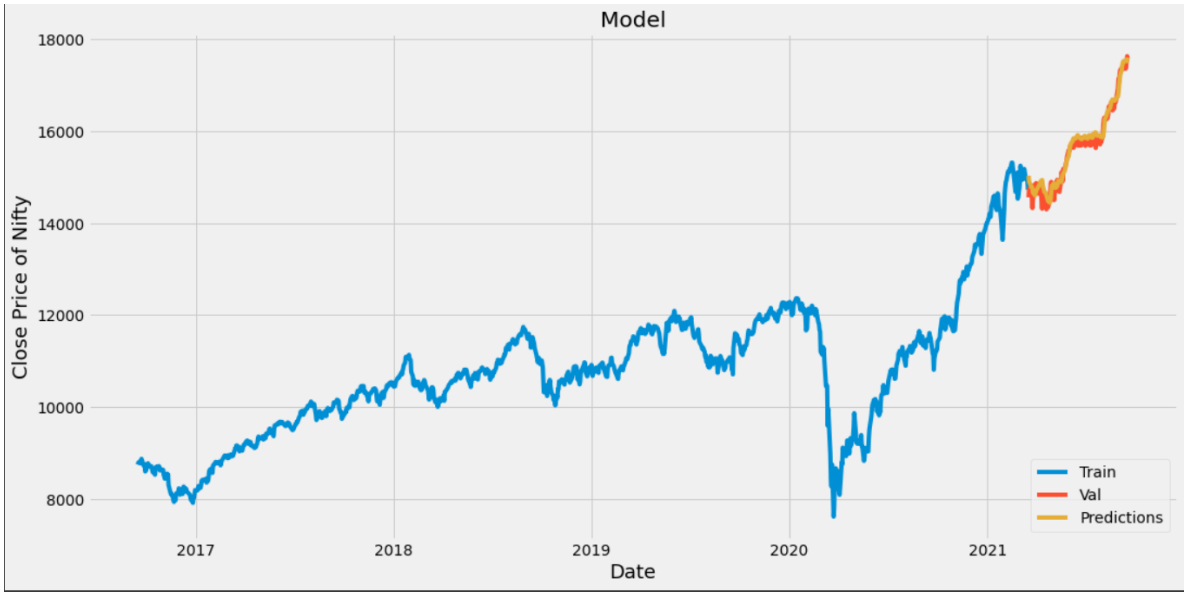


Figure 26 Prediction Of Nifty 50

4.2 Objectives Achieved

Following objectives were achieved in the course of this project:

- The initial experiment successfully identified the crucial economic indicators that have a significant impact on changing market trends.
- The second objective achieved a remarkable stock price prediction accuracy of 96.54%, surpassing the accuracy and performance of alternative methods.
- In the final phase of the first experiment, a model was developed to assess market sentiment using newspaper reports, resulting in an impressive accuracy score of 85.71%.

4.3 Limitations of the study and future enhancements

Following are the limitations of the project:

1. The objectives did not include real-time data analysis, which would be a significant milestone for intraday and day traders if implemented.
2. The unavailability of certain datasets, such as the total volume of stocks bought and sold within a day, prevented accurate prediction of both upper and lower bounds of shares. If this dataset were available, it could enhance the precision of our predictions.

Following are the future enhancements planned:

1. Dynamic analysis and prediction of stocks would be feasible if real-time data is made available through a stockbroker or a stock exchange, provided that the dataset is also made dynamic.
2. Further exploration can done into constructing a stock indicator using LSTM models due to their superior accuracy compared to other models.

Chapter 5 Summary

The project involved a detailed study and meetings with market experts to identify factors that affect the stock market, concluding that oil prices, USD vs INR exchange rates, and gold prices have a significant influence. The second objective then identified the LSTM model as the most effective for predicting the market, and the Random Forest Classifier model for sentimental analysis achieved an accuracy score of 85.71%. These findings have important implications for investors, portfolio managers, traders, and economists in making informed decisions.

References

Determinants of Equity Share Prices in India: A Panel Data Approach

By P. Srinivasan

Factors Affecting Stock Market in India.

By Dr. Perways Alam

Relevance of Macro Economic factors for the Indian stock market

By Urvin Shah

Factors Affecting Stock Prices in India: A Time Series Analysis

By Arpit Bhargava, Ankush Bhargava, and Surbhi Jain

STOCK TREND PREDICTION USING NEWS SENTIMENT ANALYSIS.

By Kalyani Joshi, Prof. Bharathi H.N, and Prof. Jyothi Rao

Stock market prediction using machine learning classifiers and social media, news

By Wasiat Khanq, Mustansar Ali Ghazanfar, Muhammad Awais Azam, Amin Karami, Khaled H. Alyoubi, Ahmed S. Alfakeeh

ANNEXURE