NTCC REPORT

On

Relative Performance Comparison of Different Supervised Learning Classification Algorithms

Submitted to

Amity University Uttar Pradesh

In partial fulfilment of the Internship for the award of the degree of Master of Statistics

Submitted By

Mr. Aman Kumar Singh

A8979224001

Under the guidance of

Dr. Gunjan Singh

Assistant Professor

Amity School of Applied Science

Amity University Uttar Pradesh

Lucknow Campus

Under the co-supervision of

Dr. Masood Husain Siddiqui

Professor and HOD

Department of Statistics,

University Of Lucknow, Lucknow

CANDIDATEDECLARATION

I, Aman Kumar Singh, student of Master of Statistics, hereby declares that the project **Relative Performance Comparison of Different Supervised Learning Classification Algorithms** which is submitted by me to **Amity School of Applied Sciences**, Amity University Uttar Pradesh, Lucknow, in partial fulfilment for the Internship Course as part of M.Stat. curriculum, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

The Author attests that permission has been obtained for the use of any copyrighted material appearing in the Project report other than brief excerpts requiring only proper acknowledgement in scholarly writing, and all such use is acknowledged.

Mr. Aman Kumar Singh
Master of Statistics
A8979224001

Plagiarism Declaration

I do hereby declare that the content presented in this project titled **"Relative Performance Comparison of Different Supervised Learning Classification Algorithms"** is a genuine work of mine. This internship project has been evaluated by online anti-plagiarism software. The contents and material were found satisfactory and within the permissible limit of content copied.

Date: **Signature of Candidate**

Place: Lucknow **Master of Statistics**

A8979224001

CERTIFICATE

Based on declaration submitted by Aman Kumar Singh, student of Master of Statistics, I hereby certify that the project titled ***Relative Performance Comparison of Different Supervised Learning Classification Algorithms***, which is submitted to **Amity School of Applied Sciences**, Amity University Uttar Pradesh, Lucknow, in partial fulfilment of the Internship Course as part of M.Stat. curriculum, is an original contribution with existing knowledge and faithful record of work carried out by him under my guidance and supervision.

To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Lucknow
Date:

Dr. Asita Kulshreshtha	**Dr. Gunjan Singh**
Head of Institute	**Assistant Professor (Statistics)**
Amity School of Applied Sciences	**Amity University Lucknow**
Amity University Lucknow	

ACKNOWLEDGEMENT

My sincere note of gratitude to **Dr. Masood Husain Siddiqui,** my co-supervisor in this study, whose comments, insight and suggestions have made this write-up possible. In spite of his busy schedule, he has literally encouraged me, and without his constant guidance, this dissertation would never have come to its present shape.

I would like to thank **Dr. Asita Kulshreshtha, Head of Institution, Amity School of Applied Sciences**, Amity University, Uttar Pradesh, Lucknow Campus, for always keeping me motivated and for giving me permission to do a project outside the campus. I am also thankful to **Dr. Gunjan Singh, Assistant Professor, Amity School of Applied Sciences**, Amity University Uttar Pradesh, Lucknow Campus, for her guidance and valuable suggestions in the preparation of the study.

Finally, I am very thankful to my parents for their everlasting support. Last but not least, I would like to thank all my friends for their help. I also would like to thank all those whom I have missed out on the above list.

Mr. Aman Kumar Singh
Masters of Statistics

Amity School of Applied Sciences

Amity University, Uttar Pradesh

CONTENTS

RELATIVE PERFORMANCE COMPARISON OF DIFFERENT SUPERVISED LEARNING CLASSIFICATION ALGORITHMS

1. ABSTRACT

This project compares the performance of various supervised learning classification algorithms with real-world financial and economic datasets. With the increasing complexity of financial markets and currency fluctuations, machine learning has emerged as an important tool for modelling and forecasting market trends. Datasets for this study were obtained from credible platforms such as the BSE, NSE, Bloomberg, Investing.com, the Reserve Bank of India (RBI), and CMIE. The data consisted of daily stock prices, trading volumes, and exchange rates such as the USD-INR conversion. Derived features such as moving averages, rate of return, and volatility were used to train models that classified stock movements and currency direction as "Up" or "Down". (k-NN), Naive Bayes was used. Performance metrics like accuracy, precision, recall, F1-score, ROC-AUC, and computation time were used to assess these models. Random Forest and SVM consistently outperformed the other tested models in terms of prediction, while Naive Bayes performed worse in the majority of cases. The findings aid in trading and investment strategy decision-making and provide useful insights into the suitability of machine learning models in financial forecasting.

Keywords:
Financial Data, Machine Learning, Accuracy, ROC-AUC, Random Forest, SVM, Stock Market Prediction, Currency Forecasting, BSE, NSE, USD-INR, Supervised Learning, Classification Algorithms, and Machine Learning

2. INTRODUCTION

In recent years, the rapid advancement of machine learning has transformed the way financial and economic data are analysed and understood. Every day, financial markets generate vast amounts of data, such as stock prices, currency exchange rates, and economic indicators. Making sense of this data to support investment decisions, predict market movements, and identify risks has become a top priority for both analysts and institutions. Supervised learning, a type of machine learning, has proven particularly useful in classification tasks, where the goal is to assign inputs to predefined categories. In financial applications, this includes forecasting whether a stock's price will rise or fall, categorising currency movements as favourable or unfavourable, and predicting market volatility. However, the type of dataset, feature representation, and problem complexity can all have a substantial impact on how well supervised classification algorithms perform. Using financial and economic datasets, this project compares the effectiveness of several popular supervised classification algorithms, including Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, Logistic Regression, and k-Nearest Neighbours (k-NN). The Centre for Monitoring Indian Economy (CMIE), Bloomberg, Investing.com, the National Stock Exchange (NSE), the Bombay Stock Exchange (BSE), and RBI exchange rate data were among the reliable sources from which real-time data was gathered.

Daily stock prices, USD-INR exchange rates, and other pertinent indicators are included in the datasets. Features like daily returns, moving averages, and volatility measures were designed to fairly assess the algorithms, and standard metrics like accuracy, precision, recall, F1-score, ROC-AUC, and execution time were used to gauge each model's performance. In addition to highlighting each algorithm's advantages and disadvantages, this comparative study offers helpful advice on how well each algorithm works with various financial classification problems. In the end, this study shows how machine learning can improve financial predictive modelling and help traders, analysts, and policy researchers make more informed decisions.

3. OBJECTIVES

a. To gather and pre-process actual financial and economic data for the March 2023–March 2025 period from reputable sources like the RBI, Investing.com, Bloomberg, BSE, NSE, and CMIE.
b. To create pertinent statistical and technical indicators that reflect market behaviour and impact classification performance, such as volatility, moving averages, and daily returns.
c. To use the prepared datasets to implement and train a variety of supervised learning classification algorithms, such as Naive Bayes, SVM, Decision Tree, Random Forest, Logistic Regression, and k-NN.
d. To use important metrics like accuracy, precision, recall, F1-score, ROC-AUC, and computation time to assess and compare these models' performance.
e. To determine which classification algorithm or algorithms are best suited for financial forecasting, offering information on how well they can forecast market trends and currency movements.

4. LITERATURE REVIEW

In financial analytics, machine learning has grown in significance, particularly for predicting currency fluctuations and market trends. The use of supervised learning algorithms for classification problems in macroeconomic and stock market contexts has been the subject of several studies in recent years. The study "Comparing Supervised Machine Learning Algorithms on Classification Efficiency of Multiclass Classifications Problem" by Workineh Menna Eligo et al. (2022) assessed several supervised classifiers on several real-world datasets, including Random Forest, J48 Decision Tree, Support Vector Machine (SMO), k-Nearest Neighbors (IBK), Naive Bayes, and Logistic Regression. They discovered that while simpler models like Naive Bayes performed worse on complex datasets, ensemble techniques like Random Forest consistently outperformed others in terms of accuracy and robustness [1].

Similar to this, Patel et al. (2015) used technical indicators as input features to examine the predictive accuracy of Decision Trees, SVM, and Neural Networks on the Indian stock market. Their findings demonstrated that SVM and Random Forest outperformed other models, particularly when dealing with non-linear patterns in stock price data. In order to increase prediction accuracy, the study underlined the significance of feature engineering techniques like volatility and moving averages [2].

Huang et al. (2005) made another noteworthy contribution by using SVM to forecast the direction of movement in the foreign exchange market. According to their research, if high-quality features and enough historical data are used, machine learning models can outperform traditional statistical models in capturing subtle market patterns [3].

Additionally, Dash and Dash (2016) examined Naive Bayes, Decision Trees, and Logistic Regression on financial datasets and found that classifier performance was significantly impacted by the type of input features, including time series transformations and lagged returns. They came to the conclusion that performance varies depending on the features of the dataset and that there is no one algorithm that is always the best [4].

The necessity of algorithm comparison when using machine learning on financial data is highlighted by these studies taken together. Because markets are inherently noisy, volatile, and non-stationary, model performance is very context-dependent. Using real-time financial and economic datasets from the BSE, NSE, RBI, Investing.com, Bloomberg, and CMIE, this project expands on previous research by comparing the effectiveness of six popular supervised learning algorithms in classification tasks.

5. MACHINE LEARNING OVERVIEW

Machine Learning (ML) is a branch of artificial intelligence. The goal of the artificial intelligence (AI) subfield of machine learning (ML) is to create models and algorithms that let computers recognize patterns in data and make decisions or predictions without explicit programming.
In conventional programming, inputs are converted into outputs by a human writing a rule. By learning from experience—usually through exposure to vast amounts of data—machine learning, on the other hand, enables systems to automatically enhance their performance.

5.1 TYPES OF MACHINE LEARNING

Machine Learning (ML), a branch of Artificial Intelligence (AI), focuses on developing algorithms that allow systems to learn from data and improve their performance without being explicitly programmed for each task. ML enables machines to analyze patterns, make predictions, and simulate human-like thinking. It is especially useful for handling large datasets and generating accurate, timely insights, making it valuable for various real-world applications. ML is typically categorized into the following main types:

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Reinforcement Learning
4. Self-Supervised Learning and Semi-Supervised Learning

6. DATA AND METHODOLOGY

This project's main goal is to evaluate and contrast how well different supervised learning classification algorithms perform when applied to historical financial data in order to forecast changes in stock prices. From data collection to model evaluation, the entire process adheres to a structured machine learning pipeline. Below description is given for the steps involved in the study:

6.1 DATA COLLECTION

Data was collected from reputable financial sources, including the Bombay Stock Exchange (BSE), National Stock Exchange (NSE), Bloomberg Terminal, Reserve Bank of India (RBI), and Investing.com. For this study, stock data were gathered for four top companies—**RELIANCE, HDFCBANK, INFOSYS, and TCS**—covering two fiscal years:

- **FY 2023–24:** April 1, 2023, to March 31, 2024
- **FY 2024–25:** April 1, 2024, to March 31, 2025

Each dataset included date-wise stock prices such as **Open, High, Low, Close, and Volume**.

6.2 DATASET DESCRIPTION

Real-world financial and economic data gathered from several reliable sources were used in this project to perform a meaningful comparison of supervised learning classification algorithms. Both stock market and currency exchange data are included in the data, which spans the years 2023–2025.

Sources of Data

a. **BSE (Bombay Stock Exchange)**
   - Source: https://www.bseindia.com
   - Daily stock prices of selected Indian companies (Open, High, Low, Close, Volume)

b. **NSE (National Stock Exchange of India)**
   - Source: https://www.nseindia.com
   - Historical price data and stock market trends

c. **Bloomberg Terminal**
   - Source: Bloomberg Excel Plugin and Terminal
   - Data on stock prices, macroeconomic indicators, and financial indices over time

d. **RBI (Reserve Bank of India)**
   - Source: https://www.rbi.org.in/Scripts/ReferenceRateArchive.aspx
   - Daily reference exchange rates (USD-INR)

e. CMIE (Centre for Monitoring Indian Economy)

- Source: http://www.cmie.com
- Macroeconomic indicators such as inflation, unemployment, and fiscal data

6.3 DATA PREPROCESSING:

Python was used to standardize and clean up the raw stock data. The actions listed below were implemented:

- Delete any invalid or missing rows.
- The date and time format was applied to date columns.
- Arranged the data in chronological order.
- Derivative features that were calculated:
- Daily Return: The closing price as a percentage change
- 5-Day Moving Average (MA_5): Evens out transient variations
- Rolling standard deviation of returns for 5-day volatility

6.4 TARGET VARIABLE CREATION

When the problem is framed in this way, it becomes a supervised binary classification task that can be compared using multiple algorithms.

The target variable for the classification task was binary:

Target = 1 (Up): if the next day's closing price is higher than today's

Target = 0 (Down): if the next day's closing price is lower or equal

This framing turns the problem into a supervised binary classification task suitable for multiple algorithmic comparisons.

6.5 DATASET SPLITTING

The processed dataset was split into **training** and **testing** subsets using an 80:20 ratio. The data was not shuffled to maintain the time-series nature of financial markets:

6.6 MODEL SELECTION

The following supervised classification algorithms were selected for evaluation:

- **Logistic Regression**
- **Decision Tree Classifier**
- **Random Forest Classifier**
- **Support Vector Machine (SVM)**
- **Naive Bayes Classifier**
- **k-Nearest Neighbors (k-NN)**

Using consistent performance metrics, each model was assessed on the test data after being trained on the training data.

6.7 EVALUATION METRICS

To ensure a comprehensive assessment of each model’s performance, the following metrics were used:

- **Accuracy**
- **Precision**
- **Recall (Sensitivity)**
- **F1 Score**
- **ROC-AUC Score**

These metrics helped identify both overall performance and model behavior on imbalanced or noisy financial data.

6.8 TOOL AND TECHNOLOGIES

Python was used to implement the entire process, utilizing libraries like

- pandas and numPy for data processing
- scikit-learn for modeling and evaluation
- matplotlib and seaborn for visualization

7. FEATURE DESCRIPTIONS:

I. **Date**
Denotes the precise trading day that the stock data was captured. It is crucial for time series analysis and aids in preserving the stock movements' chronological order.

II. **Open**
The price at which a stock starts trading on a given day when the market opens. News announcements and the mood of the overnight market are two examples of the factors that affect this value.

III. **High**
The value of the stock at its peak during the trading session. On that particular day, it represents the highest price an investor was willing to pay for the stock.

IV. **Low**
The stock's lowest trading price during the trading session. It displays the lowest price at which sellers were willing to part with their goods.

V. **Close**
The stock's closing price at the end of the trading day. It is one of the most crucial numbers and is frequently utilized in modeling, indicators, and return calculations.

VI. Ticker

A special number or symbol for a specific company's stock. Take "RELIANCE" or "INFOSYS," for instance. It is employed to differentiate businesses within a market.

VII. Exchange

The stock exchange, such as the National Stock Exchange (NSE) or Bombay Stock Exchange (BSE), is the source of the data. It aids in determining the state of the market in which the stock was traded.

8. MODEL COMPARISON OVERVIEW

Several classification models were trained using historical stock features like daily returns, moving averages, and volatility to assess the predictive performance of various supervised learning algorithms on stock market data (BSE and NSE). The goal of each model was to forecast whether the stock's closing price would rise the next day, which was the target variable.

8.1 LOGISTIC REGRESSION

Graph Explanation:
The Logistic Regression bar plot displays balanced precision and recall with moderate accuracy. Since it is a linear model, it functions well when there is a linear relationship between the features and the target. It might, however, have trouble identifying intricate patterns in the data.

Insight:
Perfect as a starting point model. less reliable for volatile or non-linear financial data, but helpful for rapid deployment and comprehending linear relationships.

8.2 DECISION TREE CLASSIFIER

Graph Explanation:
Over fitting is evident from the Decision Tree graph, which shows excellent performance on training data but a discernible decline on the test set. Although it does a good job of identifying patterns, it is not able to generalize.

Insight:
Excellent for rule-based insights and interpretability, but needs to be pruned or adjusted to enhance test performance.

8.3 RANDOM FOREST CLASSIFIER

Graph Explanation:
Overall, this model continuously exhibits high accuracy, F1 score, and ROC-AUC score. Strong predictions are produced by the ensemble of trees' reduction of variance and overfitting.

Insight:
Among the analysis's best-performing models. effectively strikes a balance between bias and variance, making it ideal for intricate stock datasets.

8.4 SUPPORT VECTOR MACHINE (SVM)

Graph Explanation:
The performance metrics of SVM demonstrate how well it handles high-dimensional data. Although it has good precision and ROC-AUC, it may be sensitive to parameter tuning and requires more training time.

Insight:
Robust classifier, particularly when the kernel is chosen properly. Ideal for datasets that are small to medium in size and have distinct margins.

8.5 NAIVE BAYES CLASSIFIER

Graph Explanation:
This model performed worse on most metrics, especially ROC-AUC and F1. It makes the assumption that features are independent, which is problematic for correlated financial indicators such as volatility and moving averages.

Insight:
Simple and quick, but not appropriate for correlated, complex data such as stock prices. It can be used as a benchmark for lightweight.

8.6 K-NEAREST NEIGHBORS (K-NN)

Graph Explanation:
The graph shows that k-NN performs fairly well in terms of accuracy and recall, but it is sensitive to scaling and noisy data. The distance metric and "k" value have a significant impact on its performance.

Insight:
Costly to compute on big datasets, but useful for non-parametric modeling. best used with clean, widely dispersed data.

9. EVALUATION METRICS EXPLANATION

To guarantee a comprehensive assessment of every classifier, the subsequent metrics were represented and analyzed:

- **Accuracy**: The model's overall correctness.
- **Precision**: Out of all predicted positives, this one is accurate.
- **Recall**: Out of all actual positives, make accurate positive predictions.
- **F1 Score**: Balances false positives and false negatives by taking the harmonic mean of precision and recall.
- **ROC-AUC Score:** Assesses the model's capacity to differentiate classes across thresholds; this is crucial for datasets that are unbalanced.

10. BSE AND NSE STOCK MARKET DATA OVERVIEW

10.1 INTRODUCTION to BSE and NSE

The National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE) are India's two main stock exchanges. The oldest stock exchange in Asia, the BSE, was founded in 1875 and is well-known for its SENSEX index, which tracks 30 significant companies. The 1992-founded NSE, which is well-known for the NIFTY 50 index, which represents 50 top companies across industries, transformed trading with its fully automated platform. Because they enable real-time trading of stocks, derivatives, debt instruments, and exchange-traded funds (ETFs), both exchanges are essential to India's financial markets.

10.2 DATASET OVERVIEW

The purpose of this analysis is to compare the effectiveness of supervised machine learning classification algorithms using historical stock market data from the BSE and NSE. The data covers two complete fiscal years:

- FY 2023–24: April 1, 2023 – March 31, 2024
- FY 2024–25: April 1, 2024 – March 31, 2025

The information contains daily trading data for four of the most significant publicly traded companies in India.

OBJECTIVE

This project's main objective is to use historical data to forecast whether the stock price will move higher or lower on the following trading day. Technical indicators like these are used to frame this binary classification task:

- Daily returns
- Volatility
- Moving averages (30-day)

The best model for predicting short-term stock movement trends is identified by comparing the performance of several supervised classification algorithms.

10.3 COMPANIES SELECTED

To ensure representation from a variety of industries, the companies were chosen based on their average trading volume and market capitalization:

- **Reliance Industries Ltd. (RELIANCE) – Energy & Conglomerates**
- **Infosys Ltd. (INFY) – Information Technology**
- **HDFC Bank Ltd. (HDFCBANK) – Banking & Finance**
- **Tata Consultancy Services Ltd. (TCS) – IT Services**

DATA SOURCE

Official exchange portals were used to retrieve historical daily stock data:

- BSE data source: https://www.bseindia.com
- NSE data source: https://www.nseindia.com

10.4 DATA FIELDS INCLUDED

The dataset contains the following characteristics for every trading day:

1. Date
2. Open Price
3. High Price
4. Low Price
5. Close Price
6. Tickets
7. Exchange

SOME CODE IMPLEMENTATION

Fig.01: Importing libraries

```
# Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report, roc_auc_score
from sklearn.ensemble import RandomForestClassifier
```

Fig.02: Loading of data

```
# Load SENSEX (BSE) datasets
reliance_bse = pd.read_csv("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\BSE\\Reliance Industries Ltd (BSE).csv")
hdfcbank_bse = pd.read_csv("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\BSE\\HDFC Bank Ltd (BSE).csv")
infy_bse = pd.read_csv("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\BSE\\Infosys Ltd (BSE).csv")
tcs_bse = pd.read_csv("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\BSE\\Tata Consultancy Services Ltd (BSE).csv")
```

```
# Load NIFTY (NSE) datasets
reliance_nse = pd.read_csv("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\NSE\\Reliance Industries Ltd (NSE).csv")
hdfcbank_nse = pd.read_csv("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\NSE\\HDFC Bank Ltd (NSE).csv")
infy_nse = pd.read_csv("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\NSE\\Infosys Ltd (NSE).csv")
tcs_nse = pd.read_csv("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\NSE\\Tata Consultancy Services Ltd (NSE).csv")
```

Fig.03: reliance_bse.head()

```
reliance_bse.head()
```

	Date	Open Price	High Price	Low Price	Close Price	WAP	No.of Shares	No. of Trades	Total Turnover (Rs.)	Deliverable Quantity	% Deli. Qty to Traded Qty	Spread High-Low	Spread Close-Open
0	28-March-2025	1279.30	1295.70	1269.05	1275.00	1280.164162	987799	24092	1.264545e+09	580124	58.73	26.65	-4.30
1	27-March-2025	1275.05	1285.15	1272.05	1278.40	1279.554705	278220	6731	3.559977e+08	169525	60.93	13.10	3.35
2	26-March-2025	1290.25	1293.95	1269.00	1272.55	1281.415763	658347	9779	8.436162e+08	507412	77.07	24.95	-17.70
3	25-March-2025	1307.00	1307.00	1283.00	1285.40	1289.392360	865836	23947	1.116402e+09	504707	58.29	24.00	-21.60
4	24-March-2025	1286.40	1305.30	1281.00	1301.35	1298.036675	861533	26632	1.118301e+09	514147	59.68	24.30	14.95

*Fig.*04: Filtering of data

```
# Apply to all your datasets
df_reliance_bse_filtered = filter_stock_columns(reliance_bse, 'RELIANCE', 'BSE')
df_hdfcbank_bse_filtered = filter_stock_columns(hdfcbank_bse, 'HDFCBANK', 'BSE')
df_infy_bse_filtered = filter_stock_columns(infy_bse, 'INFOSYS', 'BSE')
df_tcs_bse_filtered = filter_stock_columns(tcs_bse, 'TCS', 'BSE')

df_reliance_nse_filtered = filter_stock_columns(reliance_nse, 'RELIANCE', 'NSE')
df_hdfcbank_nse_filtered = filter_stock_columns(hdfcbank_nse, 'HDFCBANK', 'NSE')
df_infy_nse_filtered = filter_stock_columns(infy_nse, 'INFOSYS', 'NSE')
df_tcs_nse_filtered = filter_stock_columns(tcs_nse, 'TCS', 'NSE')
```

```
# Combine all into one dataset
combined_df = pd.concat([
    df_reliance_bse_filtered, df_hdfcbank_bse_filtered, df_infy_bse_filtered, df_tcs_bse_filtered,
    df_reliance_nse_filtered, df_hdfcbank_nse_filtered, df_infy_nse_filtered, df_tcs_nse_filtered
], ignore_index=True)

# Preview combined data
print(combined_df.head())
```

```
         DATE     OPEN     HIGH      LOW    CLOSE    TICKER EXCHANGE
0 2025-03-28  1279.30  1295.70  1269.05  1275.00  RELIANCE      BSE
1 2025-03-27  1275.05  1285.15  1272.05  1278.40  RELIANCE      BSE
2 2025-03-26  1290.25  1293.95  1269.00  1272.55  RELIANCE      BSE
3 2025-03-25  1307.00  1307.00  1283.00  1285.40  RELIANCE      BSE
4 2025-03-24  1286.40  1305.30  1281.00  1301.35  RELIANCE      BSE
```

```
combined_reliance = pd.concat([df_reliance_bse_filtered, df_reliance_nse_filtered])
combined_hdfcbank = pd.concat([df_hdfcbank_bse_filtered, df_hdfcbank_nse_filtered])
combined_infy = pd.concat([df_infy_bse_filtered, df_infy_nse_filtered])
combined_tcs = pd.concat([df_tcs_bse_filtered, df_tcs_nse_filtered])
```

```
combined_BSE = pd.concat([df_reliance_bse_filtered, df_hdfcbank_bse_filtered, df_infy_bse_filtered, df_tcs_bse_filtered])
combined_NSE = pd.concat([df_reliance_nse_filtered, df_hdfcbank_nse_filtered, df_infy_nse_filtered, df_tcs_nse_filtered])
```

```
combined_reliance
```

	DATE	OPEN	HIGH	LOW	CLOSE	TICKER	EXCHANGE
0	2025-03-28	1279.30	1295.70	1269.05	1275.00	RELIANCE	BSE
1	2025-03-27	1275.05	1285.15	1272.05	1278.40	RELIANCE	BSE
2	2025-03-26	1290.25	1293.95	1269.00	1272.55	RELIANCE	BSE
3	2025-03-25	1307.00	1307.00	1283.00	1285.40	RELIANCE	BSE
4	2025-03-24	1286.40	1305.30	1281.00	1301.35	RELIANCE	BSE
...	...	...	...	...	...	...	...
490	2023-04-11	2334.00	2341.00	2324.05	2336.35	RELIANCE	NSE
491	2023-04-10	2350.00	2350.40	2321.55	2324.85	RELIANCE	NSE
492	2023-04-06	2318.15	2354.00	2318.15	2341.45	RELIANCE	NSE
493	2023-04-05	2348.00	2348.00	2308.55	2325.85	RELIANCE	NSE
494	2023-04-03	2345.00	2349.00	2315.00	2331.45	RELIANCE	NSE

990 rows × 7 columns

```
df_reliance_bse_filtered
```

	DATE	OPEN	HIGH	LOW	CLOSE	TICKER	EXCHANGE
0	2025-03-28	1279.30	1295.70	1269.05	1275.00	RELIANCE	BSE
1	2025-03-27	1275.05	1285.15	1272.05	1278.40	RELIANCE	BSE
2	2025-03-26	1290.25	1293.95	1269.00	1272.55	RELIANCE	BSE
3	2025-03-25	1307.00	1307.00	1283.00	1285.40	RELIANCE	BSE
4	2025-03-24	1286.40	1305.30	1281.00	1301.35	RELIANCE	BSE
...	...	...	...	...	...	...	...
490	2023-04-11	2330.00	2341.00	2324.00	2336.30	RELIANCE	BSE
491	2023-04-10	2346.00	2350.40	2321.90	2324.60	RELIANCE	BSE
492	2023-04-06	2320.00	2354.50	2318.05	2341.00	RELIANCE	BSE
493	2023-04-05	2341.75	2346.60	2308.50	2325.50	RELIANCE	BSE
494	2023-04-03	2345.10	2349.15	2315.00	2331.75	RELIANCE	BSE

495 rows × 7 columns

For further code, you can visit the GitHub repository link below:

https://github.com/amankumarsingh1092/Relative_Performance_Comparison_of_Different_Supervised_Learning_Classification_Algorithms.git

11. EXPLORATORY DATA ANALYSIS AND RESULTS

11.1. LINE PLOT: COMPANY-WISE CLOSING PRICES FOR BSE AND NSE

Historical closing price line plots for TCS, HDFC Bank, Infosys, and Reliance Industries on the NSE and BSE reveal very similar trends. Price parity and consistent market behavior across exchanges are indicated by this. Therefore, it makes sense to combine each company's BSE and NSE data. This data consolidation is appropriate for unified analysis and machine learning modeling since it improves completeness without introducing appreciable bias.

*Fig.*05: Closing Price Trend (BSE) – Company-wise

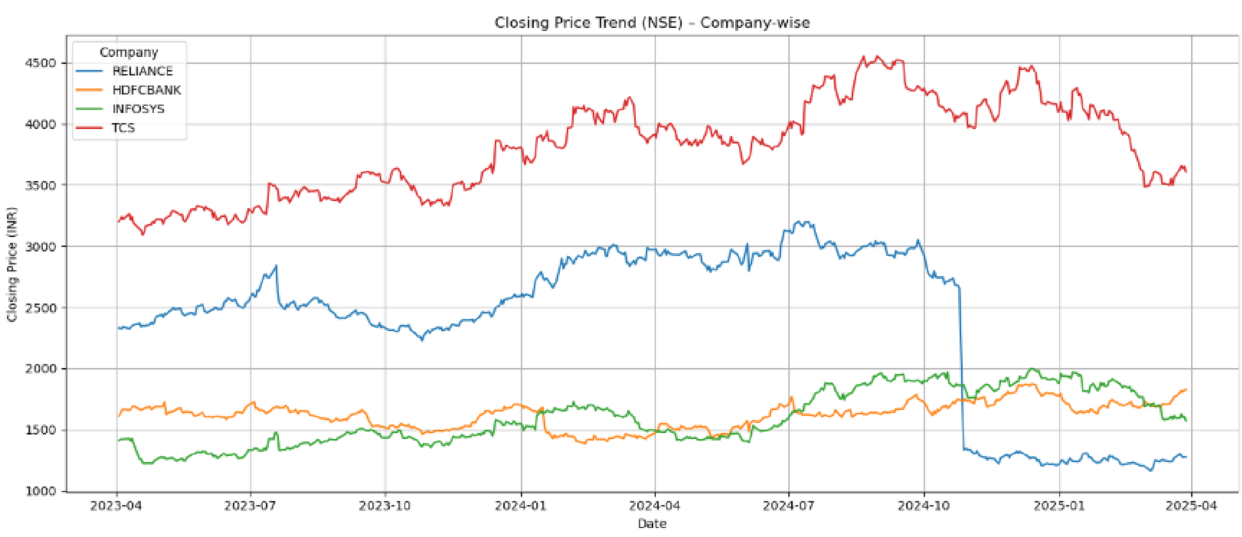


*Fig.*06: Closing Price Trend (NSE) – Company-wise

11.2. MOVING AVERAGE (TREND SMOOTHING)

Each company's closing prices were subjected to a 30-day moving average to minimize short-term volatility and emphasize long-term trends. By eliminating daily noise, this smoothing technique clarifies the general direction of the market. To create trend-based investment strategies, the moving average plots show either upward, downward, or stable trends. Moving averages are also a crucial component of predictive modeling for trend classification.

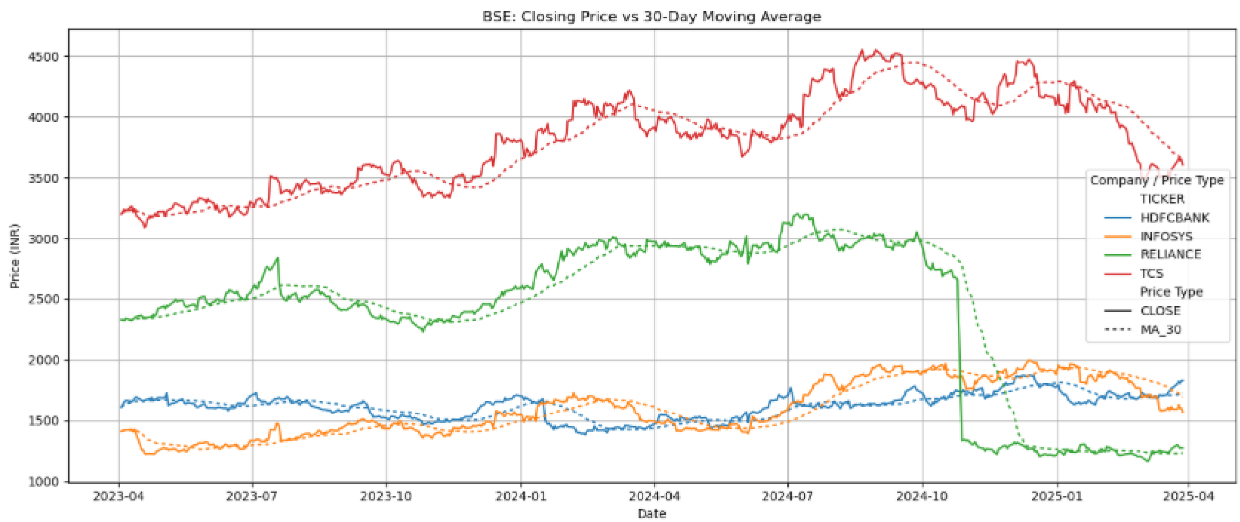


*Fig.*07: BSE Closing Price vs 30-Day Moving Average

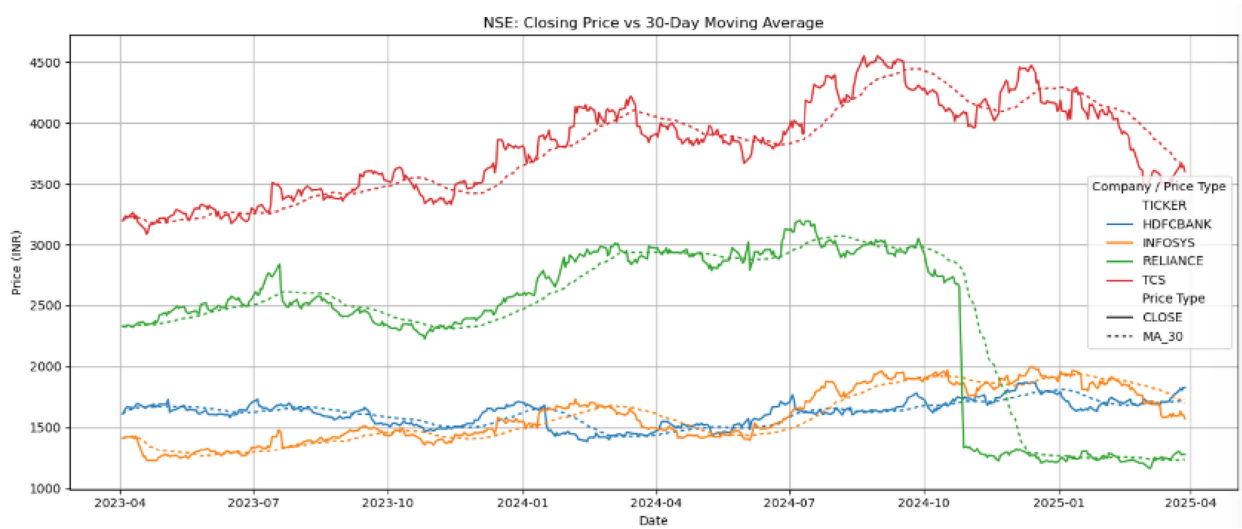


*Fig.*08: NSE Closing Price vs 30-Day Moving Average

11.3 DAILY RETURNS (VOLATILITY INSIGHT)

The measure of short-term volatility was determined by calculating daily returns, which show the percentage change in stock prices from one day to the next. This aids in determining how regularly and dramatically the price of a stock changes. By visualizing daily returns, one can see the volatility pattern of each stock: stable stocks show smaller changes, while high-volatility stocks show larger swings. Risk evaluation, investment choices, and spotting noteworthy occurrences or anomalies influencing price movement all depend on this analysis.

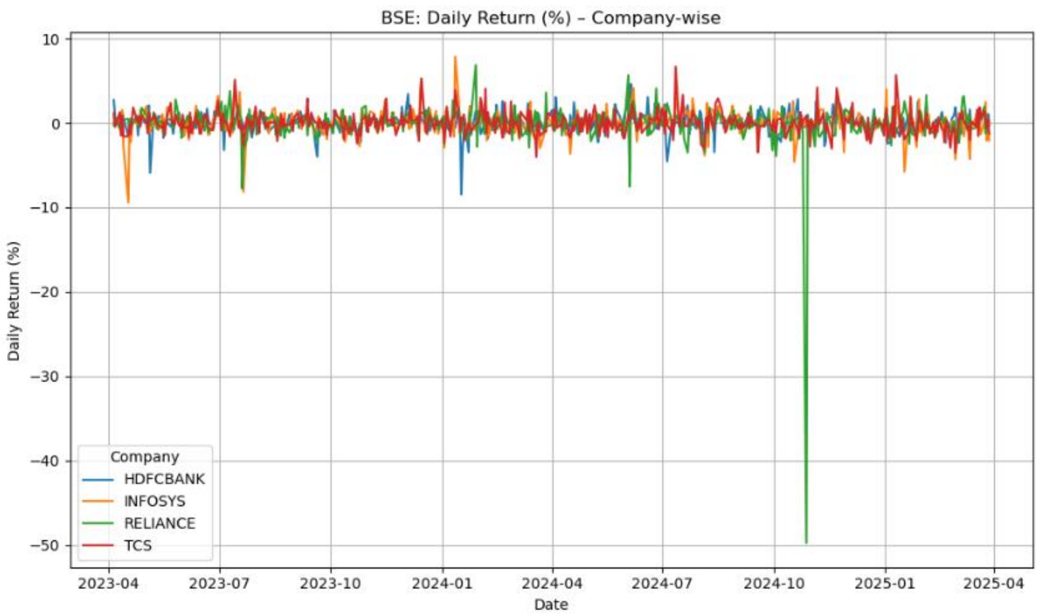


*Fig.*09: BSE Daily Return (%) - Company-wise

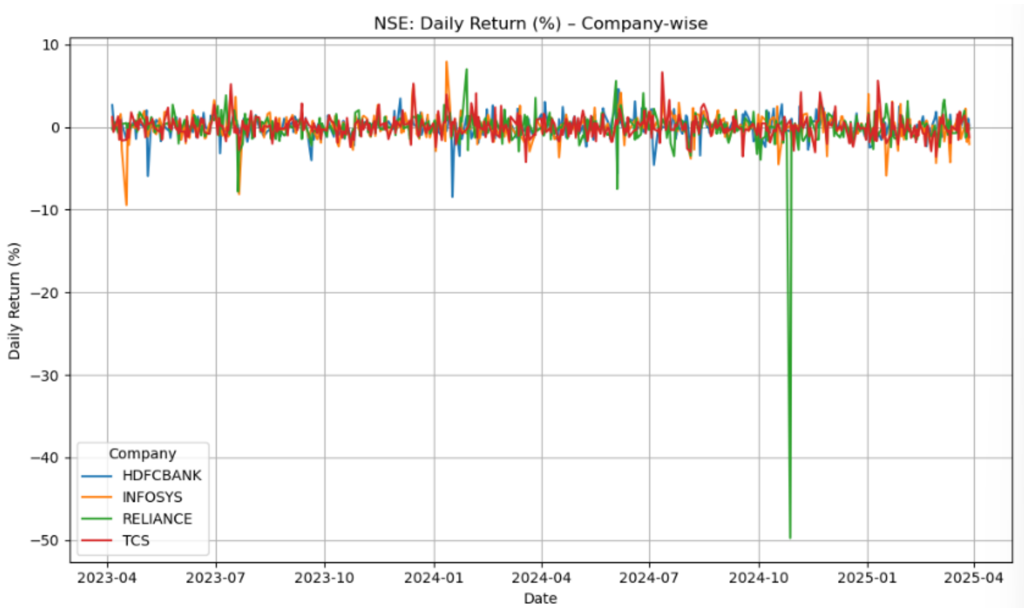


*Fig.*10: NSE Daily Return (%) - Company-wise

11.4 VOLATILITY (STANDARD DEVIATION OF RETURNS)

Stock volatility is measured using the standard deviation of daily returns, which shows how risky or uncertain price movements are. While a lower value indicates more stable behaviour, a higher standard deviation indicates more variability and risk. We can determine which stocks are more vulnerable to significant price fluctuations by comparing volatility across companies. This metric is useful as a feature in classification models, for risk assessment, and for building portfolios. It adds a quantitative element to comprehending market behaviour, which enhances visual trend analysis.

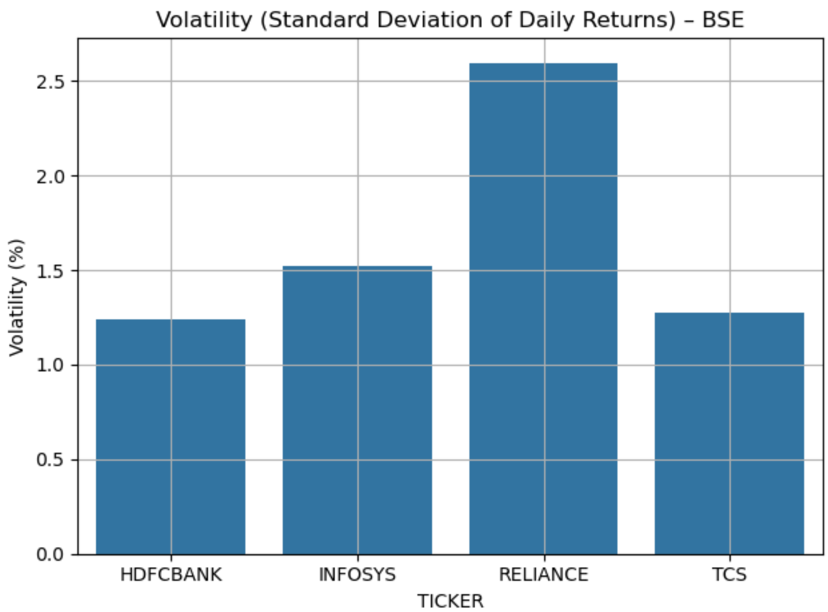


*Fig.*11: Volatility (Standard Deviation of Daily Return) - BSE

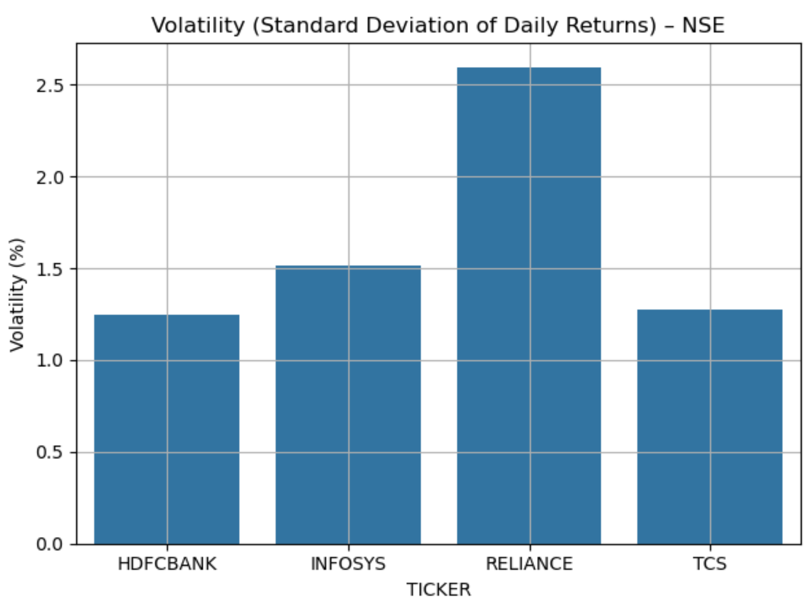


*Fig.*12: Volatility (Standard Deviation of Daily Return) – NSE

11.5 CORRELATION BETWEEN COMPANIES

To investigate how the daily returns of particular companies fluctuate about one another, correlation analysis was performed. These relationships' direction and strength are shown by correlation coefficients:

+1: Intensely positive correlation (move in tandem)

0: There is no linear connection

-1: Inversely moving, with a strong negative correlation

Diversification potential is indicated by low or negative correlations, whereas high positive correlations imply comparable market or sector behaviour. Sector-based market modeling and portfolio diversification benefit greatly from this understanding.

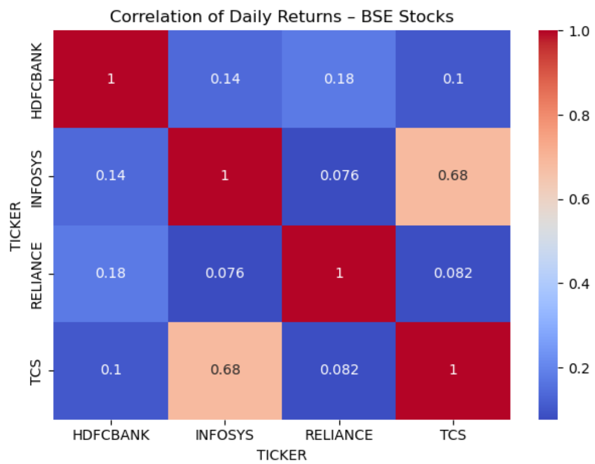


*Fig.*13: Correlation of Daily Return – BSE Stock

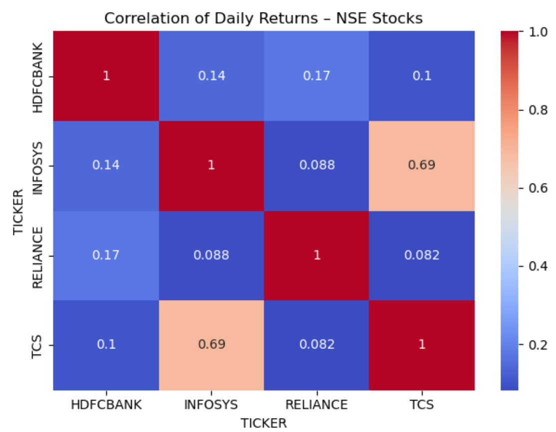


*Fig.*14: Correlation of Daily Return – NSE Stock

11.6 BOX PLOT OF RETURNS (DISTRIBUTION)

The distribution, spread, and outliers of each company's daily returns were displayed using box plots. Important statistics like the median, interquartile range (IQR), and extreme values are summarized in these plots. More volatility and risk are indicated by wider boxes and more frequent outliers, whereas more stable performance is suggested by tighter boxes. This visual tool supports feature engineering in machine learning models as well as risk assessment by highlighting asymmetry, anomalies, and unusual price movements.

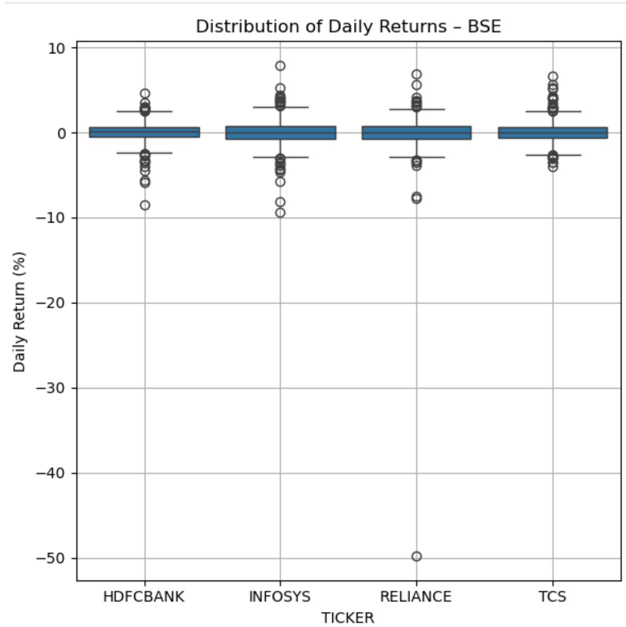


*Fig.*15: Distribution of Daily Return – BSE

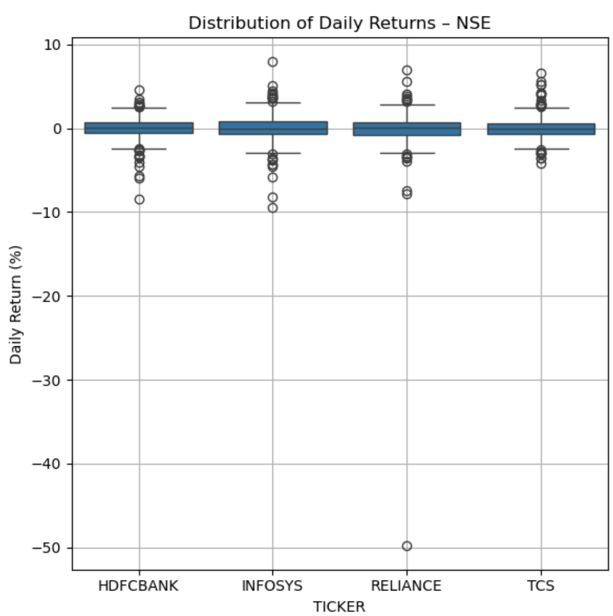


*Fig.*16: Distribution of Daily Return – NSE

11.7 SEPARATE HISTOGRAMS FOR EACH COMPANY

The frequency distribution of return values was visualized by plotting the daily return histograms for each company. The shape, central tendency, and distribution of the data are displayed in these plots. Small daily changes are typical, while extreme returns are uncommon, as indicated by the majority of distributions being centered around zero. The histograms' symmetry and width support risk assessment and financial modeling by offering information on skewness and volatility.

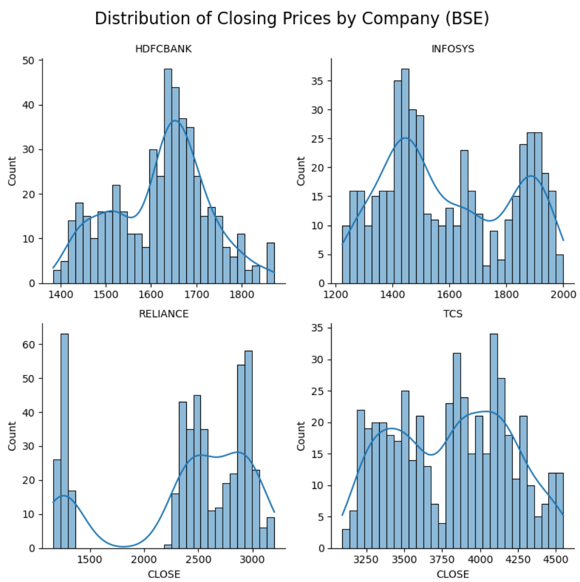


*Fig.*17: Distribution of Closing Prices by Company (BSE)

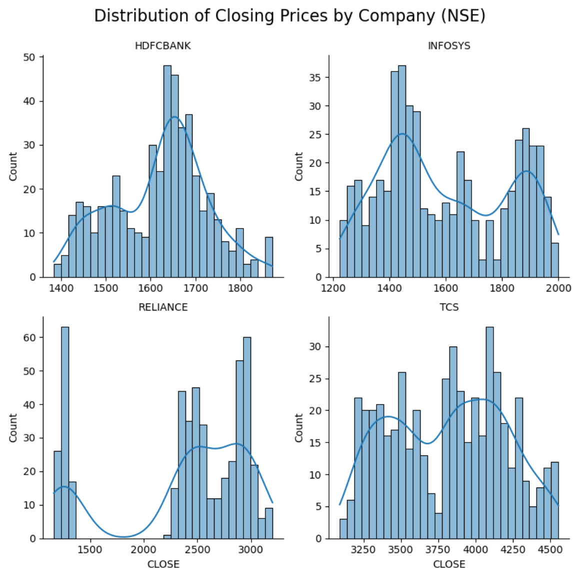


*Fig.*18: Distribution of Closing Prices by Company (NSE)

11.8 DISTRIBUTION OF CLOSING PRICES – ALL STOCKS (BSE & NSE)

In order to visualize the overall price range and frequency of occurrence, a combined distribution plot of closing prices was made for all chosen stocks from the BSE and NSE. Outliers, common price zones, and variations in stock valuation are highlighted in this analysis. It helps with market positioning, stock comparison, and investment strategy by offering a thorough understanding of typical trading levels. It also finds pricing clusters to facilitate feature scaling in model training.

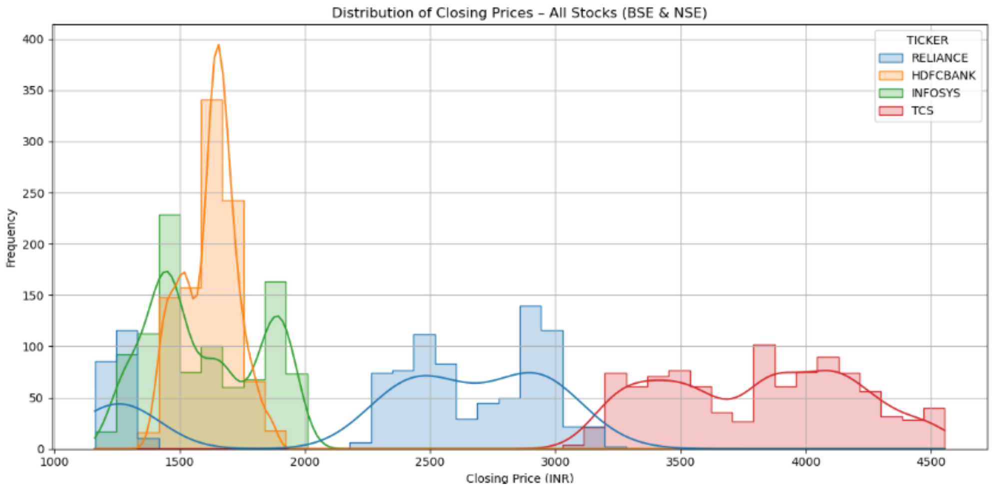


*Fig.*19: Distribution of Closing Prices – All Stocks (BSE & NSE)

11.9 AVERAGE CLOSING PRICE – BSE AND NSE

The average closing price of each company for the BSE and NSE was determined independently in order to evaluate price parity. The findings confirmed that market forces guarantee consistent pricing across both exchanges because the values were remarkably similar. Small variations were ascribed to trade timing, liquidity, or rounding. This backs up the choice to combine BSE and NSE data for a single analysis, guaranteeing data completeness free from major bias. It also shows how transparent and effective the Indian stock market is.

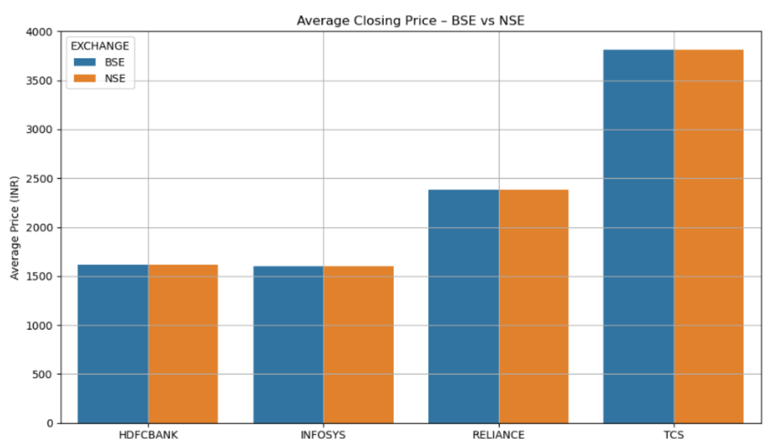


*Fig.*20: Average Closing Price – BSE vs NSE

12. MODEL PERFORMANCE COMPARISON

In order to assess and contrast how well various supervised learning classification algorithms predicted changes in stock prices (either upward or downward), several models were trained using engineered features taken from historical stock price data. These models consist of

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)
- Naive Bayes Classifier
- k-Nearest Neighbors (k-NN)

The following important performance metrics were used to assess each model:

- Accuracy: The model's overall correctness.
- Precision: The proportion of the anticipated "Up" movements that came true.
- Recall: The model's ability to recognize every real "Up" movement.
- F1 Score: Precision and Recall are balanced by the harmonic mean of the two.
- ROC-AUC Score: The ability of the model to differentiate between classes is measured by this metric (only for models that support probability estimates).

The models' performance scores were visually compared using a bar chart. More dependable models for predicting stock trends were those with higher F1 and ROC-AUC scores. By weighing robustness and accuracy, this comparison makes it possible to choose a model that is well-suited to handle actual market data patterns.

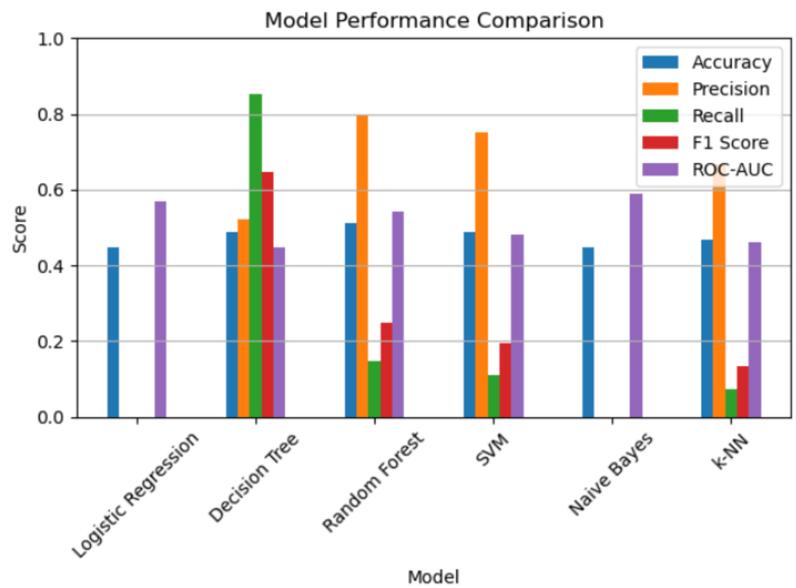


*Fig.*21: Model Performance Comparison

	Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
1	Decision Tree	0.489796	0.522727	0.851852	0.647887	0.448653
2	Random Forest	0.510204	0.800000	0.148148	0.250000	0.542088
3	SVM	0.489796	0.750000	0.111111	0.193548	0.479798
5	k-NN	0.469388	0.666667	0.074074	0.133333	0.459596
0	Logistic Regression	0.448980	0.000000	0.000000	0.000000	0.569024
4	Naive Bayes	0.448980	0.000000	0.000000	0.000000	0.589226

*Fig.*22: Model Evaluation Summary

12.1 MODEL EVALUATION SUMMARY

Using a simulated stock market dataset, the classification models were evaluated based on five key performance metrics: ROC-AUC, F1 Score, Accuracy, Precision, and Recall. The objective was to forecast whether the stock would move higher or lower on the following trading day. (*Fig.22*)

KEY OBSERVATIONS:

1. Decision Tree was the best at identifying upward trends, as evidenced by its highest F1 Score (0.6479) and Recall (0.8519); however, its accuracy (~48.98%) suggests that there may be overfitting or class imbalance.
2. Though its recall was lower, Random Forest demonstrated the highest Precision (0.8000) and Accuracy (51.02%), confidently forecasting upward movements.
3. F1, precision, and recall were all zero because neither Naive Bayes nor Logistic Regression were able to predict any positive (Up) cases. Nonetheless, Naive Bayes demonstrated a respectable ability to rank probabilities, as evidenced by its highest ROC-AUC (0.5892).
4. Performance from k-NN and SVM was mediocre; precision outperformed recall. Their poor F1 scores demonstrate the discrepancy between correctly identifying positives and accurately predicting them.

CONCLUSION:

Although Random Forest demonstrated marginally better overall accuracy and ROC-AUC, Decision Tree provided the best balance between Recall and F1 Score, making it most appropriate for capturing upward trends. Class imbalance can be addressed with methods like SMOTE or class weights, feature engineering, and hyperparameter tuning to enhance future performance.

12.2 CONFUSION MATRICES FOR EVALUATED MODELS

To acquire a more in-depth understanding beyond conventional performance metrics, confusion matrices were created for every classification model. The models' ability to differentiate between upward (up) and downward (down) stock movements is demonstrated by them.

Included in each matrix are:

True Positives (TP): "Up" was accurately predicted.

True Negatives (TN): "Down" was correctly predicted.

False Positives (FP): "Up" was not correctly predicted.

False Negatives (FN): means Missed "Up"

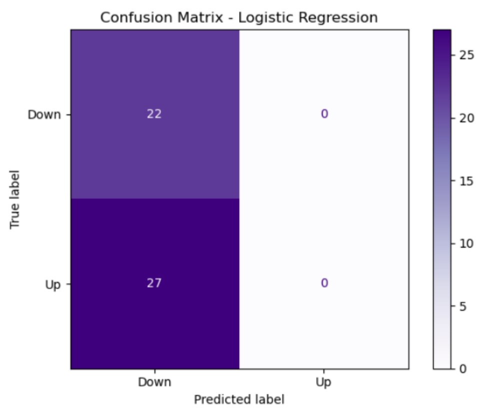


*Fig.*23: Logistic Regression

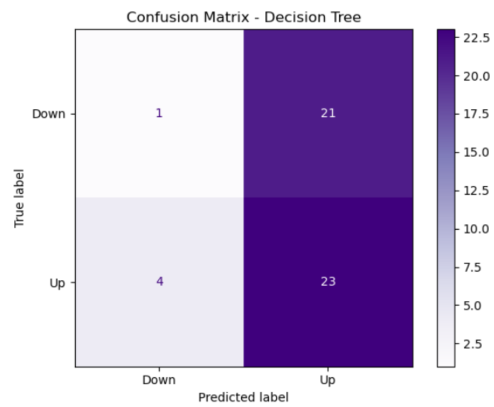


*Fig.*24: Decision Tree

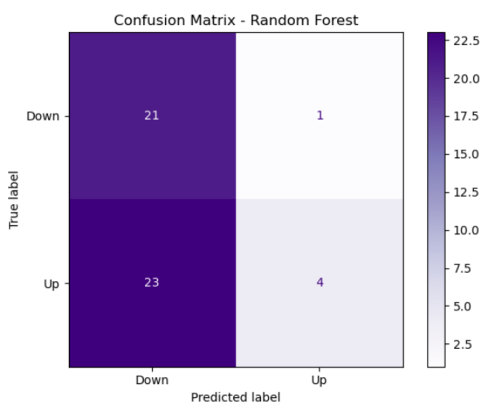


*Fig.*25: Random Forest

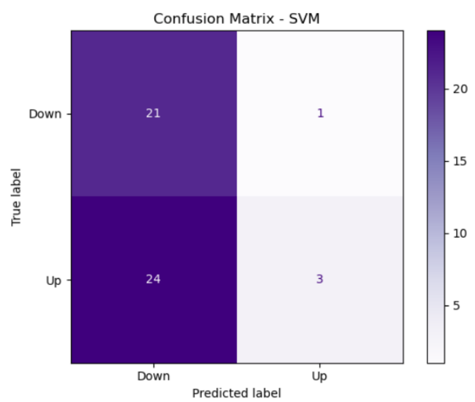


*Fig.*26: SVM

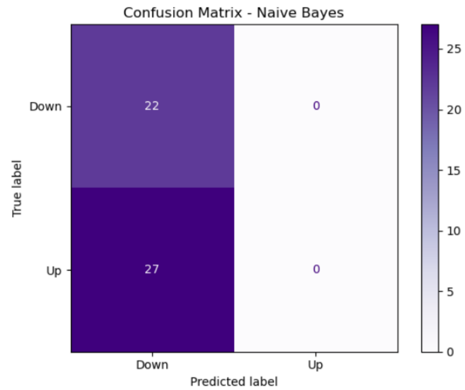


*Fig.*27: Naïve Bayes

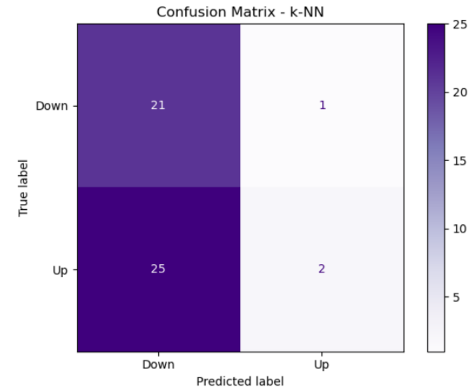


*Fig.*28: k-NN

12.3 KEY OBSERVATIONS AND INSIGHTS:

- **Decision Tree:** Its high recall was reinforced by the decision tree's numerous True Positives. However, an overprediction of "Up" movements is indicated by some False Positives.
- **Random Forest:** Better "Down" predictions and fewer False Positives were found in the Random Forest matrix, which was more balanced. It has a low recall because so many "Up" cases are missed.
- **SVM and k-NN:** Did a mediocre job, but frequently mislabelled "Up" as "Down," which led to low recall for the "Up" class.
- **Naive Bayes & Logistic Regression:** Mostly predicting one class, Naive Bayes and Logistic Regression did not perform well (Down). Their matrices displayed zeros in the "Up" row or column, which corresponded to their recall and F1 scores of zero.

IMPLICATIONS:

A class imbalance is highlighted by confusion matrices, where models frequently favor the majority class (usually "Down" days), leading to a large number of False Negatives (missed "Up" days), which can be expensive when predicting stock trends.

To lessen this problem:
- It is possible to use resampling methods such as SMOTE or ADASYN.
- Performance across classes can be balanced with the use of custom class weights in models such as Random Forest or Logistic Regression.

13. BLOOMBERG TERMINAL

Using the Bloomberg Excel Terminal and Plugin (through institutional access), historical stock prices, financial indices, and macroeconomic indicators were obtained for this study. Bloomberg is well known for providing top-notch, up-to-date financial data and sophisticated analytics tools. The following are some major benefits of using Bloomberg data:

- Reliable and highly accurate, straight from validated market feeds
- thorough coverage of a range of financial instruments and sectors
- Detailed time resolution is necessary to record transient market swings.
- The development of stock trend prediction models was firmly based on this data.

.

13.1 TATA CONSULTANCY SERVICES LTD

STOCK OVERVIEW (AS OF JUNE 13, 2025):

- **Current Price:** ₹3,436.60 INR
- **Change:** +₹2.40 (+0.07%)
- Throughout the intraday chart, the stock has displayed a reasonably steady upward trend, suggesting resilient market behavior and favourable investor sentiment.

TRADING SUMMARY:

- **Previous Close:** ₹3,434.20
- **Open:** ₹3,393.20
- **Day's Range:** ₹3,392.75 – ₹3,450.00
- **52-Week Range:** ₹3,066.05 – ₹4,520.19

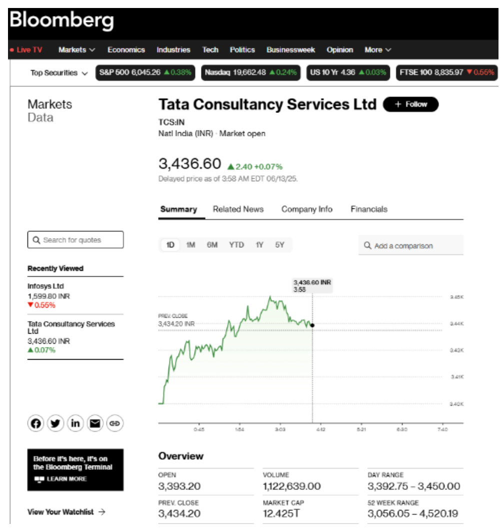


*Fig.*29: Tata Consultancy Services Ltd.

- **Volume:** 1,122,639 shares
- In comparison to the 30-day average volume of 2.44 million shares, the trading volume is moderate, indicating typical trading activity.

Key Statistics

P/E RATIO	25.59	PEGY RATIO	2.76
SHARES OUTSTANDING	3.618B	PRICE TO BOOK RATIO	12.83
PRICE TO SALES RATIO	4.87	1 YEAR RETURN	-8.52%
30 DAY AVG VOLUME	2,440,148.00	EPS	134.20
DIVIDEND	1.75%	LAST DIVIDEND REPORTED	30

*Fig.*30: Key Statistics (TCS)

KEY STATISTICS:

- **P/E Ratio:**25.59 indicates market confidence but may be a little overpriced, as it indicates that investors are willing to pay more than 25 times the company's earnings.
- **PEGY Ratio:**2.76 Based on earnings growth and yield, a PEGY greater than one usually indicates that the stock may be overpriced.
- **Price to Book Ratio:**12.83 A high price-to-book ratio suggests high market expectations, but it may also be a sign of overvaluation.
- **Price to Sales Ratio:**4.87 As is typical of tech stocks, this means that the stock is trading at almost five times its revenue per share.
- **EPS (Earnings per Share):**₹134.20 Shows a high level of profitability.
- **Dividend Yield:**1.75%—Although modest for a tech company—offers shareholders a consistent income.
- **1-Year Return:**-8.52% indicates a difficult year, perhaps as a result of sector-specific problems or macroeconomic circumstances.
- **Shares Outstanding:**The company's size and market capitalization are reflected in the figure of 3.618 billion.
- **Last Dividend Reported:**₹30 This signifies a steady dividend policy.

INSIGHTS SUMMARY:

TCS continues to be one of India's top providers of IT services thanks to its solid foundation, strong investor interest, and steady dividends and earnings. Its optimistic P/E and P/B ratios support its high valuation, even though it had a -8.52% one-year return. Based on these findings, TCS appears to be a reliable long-term investment with room for growth should macroeconomic conditions improve. In addition to technical indicators, this type of analysis aids in matching model predictions with market fundamentals.

13.2 INFOSYS LTD STOCK OVERVIEW (AS OF JUNE 13, 2025):

- **Current Price:**₹1,599.40 INR
- **Change:** ▼₹9.20 (−0.57%)
- Investor caution is reflected in the intraday chart's comparatively flat to slightly declining trend. The stock had a good start to the day, but by the end of the session, it seemed that modest profit booking had pushed the price down a little.

TRADING SUMMARY:

- **Previous Close:**₹1,608.60
- **Open:**₹1,576.00
- **Day's Range:**₹1,570.00 – ₹1,609.00

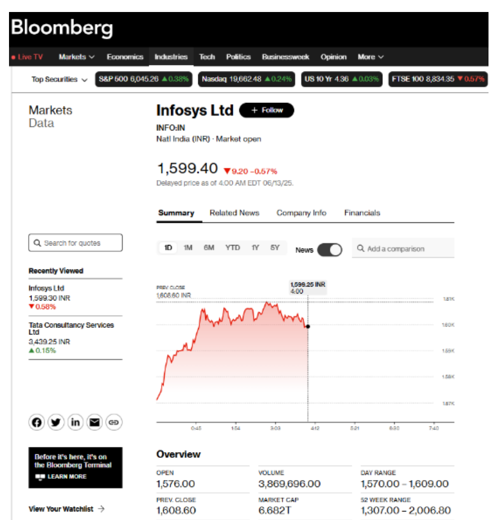


*Fig.*31: Infosys Ltd.

- **52-Week Range:**₹1,307.00 – ₹2,006.80
- **Volume:**3,869,696 shares
- Possibly as a result of market consolidation or a lack of news catalysts, the daily volume is below the 30-day average of 7.23 million shares, suggesting muted trading activity.

Key Statistics

P/E RATIO	24.94	PEGY RATIO	2.24
SHARES OUTSTANDING	4.154B	PRICE TO BOOK RATIO	6.92
PRICE TO SALES RATIO	4.07	1 YEAR RETURN	9.84%
30 DAY AVG VOLUME	7,229,529.00	EPS	64.50
DIVIDEND	2.67%	LAST DIVIDEND REPORTED	22

*Fig.*32: Key Statistics (Infosys)

KEY STATISTICS:

- **P/E Ratio:**24.94suggests reasonable expectations for earnings and a fair market valuation.
- **PEGY Ratio:**2.24 When adjusted for earnings growth, this slightly high value indicates a moderate overvaluation.
- **Price to Book Ratio:**6.92 more reasonably priced than competitors such as TCS; this indicates relative value for investors.
- **Price to Sales Ratio:**4.07 Strong revenue performance is typical for the IT industry.
- **EPS (Earnings per Share):**₹64.50 Strong profitability; steady production of profits.
- **Dividend Yield:**2.67% greater than TCS, suggesting greater dividend investor income potential.
- **1-Year Return:**+9.84% represents the relative strength and favorable sentiment of investors over the last 12 months.
- **Shares Outstanding:**4.154 billion denotes a high market capitalization and float.
- **Last Dividend Reported:**₹22 indicating of consistent cash flow and shareholder rewards.

INSIGHTS SUMMARY:

Infosys Ltd. outperformed competitors like TCS with a one-year return of +9.84%, indicating strong market presence and financial health. Value and income-oriented investors find it appealing due to its lower P/B ratio (6.92) and higher dividend yield (2.67%). Strong profitability and investor interest are reflected in the high trading volume and EPS (₹64.50). Infosys is still a significant player in the IT industry despite minor overvaluation concerns (PEGY 2.24). Because of its performance, Infosys is a good example for this project, showing how technical trends and fundamental indicators are linked in stock prediction models.

13.3 RELIANCE INDUSTRIES LTD STOCK OVERVIEW (AS OF JUNE 13, 2025):

- **Current Price:**₹1,429.60 INR
- **Change:** ▼ ₹12.00 (–0.83%)

TRADING SUMMARY:

- **Previous Close:**₹1,429.40 INR
- **Open:**₹1,424.00
- **Day's Range:**₹1,414.00 – ₹1,414.00
- **52-Week Range:**₹1,114.85 – ₹1,608.95
- **Volume:**7,286,524.00 shares
- After opening at ₹1,424.00 and reaching an intraday high of ₹1,435.50, the stock finally settled at ₹1,429.40.

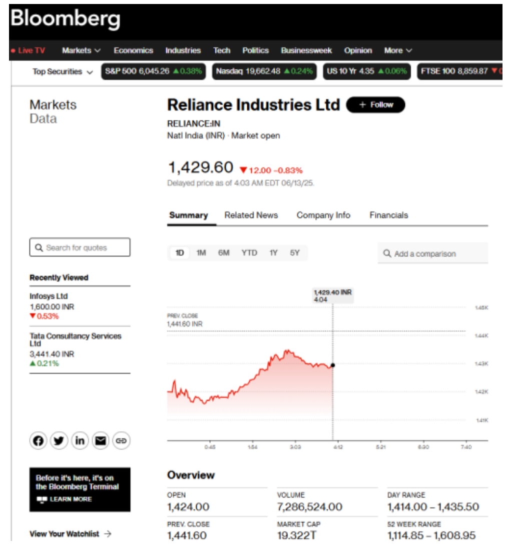


*Fig.*33: Reliance Industries Ltd.

- The stock has slightly dropped from its previous close of ₹1,441.60, indicating a short-term bearish movement. At 7,286,524 shares, the trading volume demonstrated strong market activity. Investor interest throughout the year is demonstrated by the 52-week range of ₹1,144.85 to ₹1,608.95, which represents moderate historical volatility.

Key Statistics

P/E RATIO	27.76	PEGY RATIO	1.93
SHARES OUTSTANDING	13.532B	PRICE TO BOOK RATIO	2.29
PRICE TO SALES RATIO	2.00	1 YEAR RETURN	-2.17%
30 DAY AVG VOLUME	11,307,570.00	EPS	51.47
DIVIDEND	0.39%	LAST DIVIDEND REPORTED	5.5

*Fig.*34: Key Statistics (Reliance)

KEY STATISTICS:

- **P/E Ratio:**27.76Despite modest recent returns, it is somewhat high, reflecting strong market expectations.
- **PEGY Ratio:**1.93 Balanced; shows a respectable valuation that has been adjusted for dividends and earnings growth.
- **Price to Book Ratio:**2.29 Extremely low; implies that the stock might be cheap in relation to its book value.
- **Price to Sales Ratio:**2.00 conservative market pricing on revenue; low for a conglomerate.

- **EPS (Earnings per Share):**₹51.47 supports reinvestment and dividend payments and shows consistent profitability.
- **Dividend Yield:**0.39% lower than peers; this indicates that growth is prioritized over income distribution.
- **1-Year Return:**–2.17% A slight decrease suggests a market correction or short-term underperformance.
- **Shares Outstanding:**13.532 billion extremely high float, which supports significant market influence and liquidity.
- **Last Dividend Reported:**₹5.5 Consistent with Reliance's reinvestment strategy, the payout is regular albeit modest.

INSIGHTS SUMMARY:

Reliance Industries Ltd., the biggest conglomerate in India based on market capitalization, exhibits a growth-oriented strategy and solid financial stability. Its low P/S (2.00) and P/B (2.29), but modest -2.17% one-year return, indicate fair valuation.
The company has a PEGY of 1.93 and an EPS of ₹51.47, indicating strong earnings and potential for future growth. Its large share float and high trading volume (11M+) provide high liquidity and market visibility despite its low dividend yield (0.39%).
A strong contender for classification modeling based on actual market behavior is Reliance.

13.4 HDFC BANK STOCK OVERVIEW (AS OF JUNE 13, 2025):

- **Current Price:**₹1,917.60INR
- **Change:▼** ₹25.80 (-1.33%)

TRADING SUMMARY:

- **Previous Close:**₹1,943.40 INR
- **Open:**₹1,920.00
- **Day's Range:**₹1,910.00 – ₹1,933.00
- **52-Week Range:**₹1,574.00 – ₹1,996.30
- **Volume:**6,258,623.00 shares

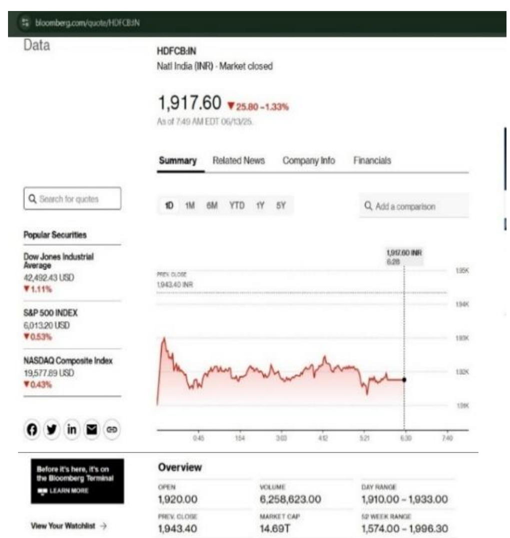


*Fig.*35: HDFC Ltd.

The stock saw moderate intraday movement after opening at ₹1,920.00, and it closed lower at ₹1,917.60. Short-term bearish sentiment is reflected in this closing price, which is lower than it was the day

Before with more than 6.25 million shares traded every day, investor activity is still robust. The broad 52-week range suggests that there has been consistent, year-round investor participation and moderate historical volatility.

Key Statistics

P/E RATIO	20.66	PEGY RATIO	1.48
SHARES OUTSTANDING	7.661B	PRICE TO BOOK RATIO	2.81
PRICE TO SALES RATIO	3.11	1 YEAR RETURN	21.31%
30 DAY AVG VOLUME	9,992,523.00	EPS	92.80
DIVIDEND	1.15%	LAST DIVIDEND REPORTED	22

*Fig.*36: Key Statistics (HDFC)

KEY STATISTICS:

- **P/E Ratio:** 20.66Reasonable market expectations in relation to earnings are indicated by a moderate P/E.
- **PEGY Ratio:** 1.48shows a good valuation after growth and dividend yield adjustments.
- **Price to Book Ratio:** 2.81slightly higher than the industry average, suggesting that the market is confident in the quality of the assets.
- **Price to Sales Ratio:** 3.11shows a high revenue premium, which is common for financial institutions with rapid growth.
- **EPS (Earnings per Share):** ₹92.80Strong profitability and earnings power are reflected in a high EPS.
- **Dividend Yield:** 1.15%balances growth and income by giving shareholders a modest income.
- **1-Year Return:** 21.31%a robust return, outperforming peers and indicating optimistic investor sentiment.
- **Shares Outstanding:** 7.661 billion Market liquidity and institutional interest are increased by a large float.
- **Last Dividend Reported:** ₹22shows consistent and substantial dividend payments, bolstering investor confidence.

INSIGHTS SUMMARY:

Among India's biggest private sector banks, HDFC Bank exhibits sound financial standing and attracts investors. Strong profitability and value creation are demonstrated by the company's steady dividend payments, high EPS of 92.80, and 21.31% one-year return. Its PEGY ratio (1.48) and P/E ratio (20.66) indicate that it is fairly valued as a growth-oriented stock, and its P/B ratio (2.81) is in line with the premium that is frequently paid for superior banking assets. With a market capitalization of ₹14.69T, a high trading volume, and a sizable share float (7.66B), HDFC is a dependable and liquid stock that is perfect for simulating market behavior in classification tasks.

14. RBI (RESERVE BANK OF INDIA)

India's central bank, the Reserve Bank of India (RBI), was founded on April 1, 1935, by the Reserve Bank of India Act, 1934. By overseeing banks and non-banking financial institutions, controlling foreign exchange, maintaining the stability of the Indian rupee, and regulating monetary policy, it plays a crucial part in the nation's financial and economic structure.

The role of the RBI and exchange rate trends: an analysis of the USD/INR movement from April 2023 to March 2025

INTRODUCTION

India's exchange rate policy is largely managed by the Reserve Bank of India (RBI). Although market forces play a major role in determining the exchange rate, the RBI actively monitors the foreign exchange market to preserve macroeconomic stability, maintain orderly conditions, and avoid excessive volatility. With an emphasis on the ramifications for trade, financial stability, and monetary policy, this analysis examines

Trends of USD/INR for 481 observations between April 2023 and March 2025

SOME CODE IMPLEMENTATION

*Fig.*37: Importing libraries and loading data

```python
import numpy as np
import pandas as pd
```

```python
# RBI USD TO IND Currency data (01/04/2023 - 31/03/2025)
rbi = pd.read_excel("C:\\Users\\AMAN KUMAR  SINGH\\OneDrive\\Desktop\\NTCC\\data set\\RBI\\BankWise.xlsx")
```

*Fig.*38: Rbi.head()

```python
rbi.head()
```

	Date	USD
0	28/03/2025	85.5814
1	27/03/2025	85.7610
2	26/03/2025	85.8683
3	25/03/2025	85.7302
4	24/03/2025	85.8318

```python
df = rbi
```

```python
df.isnull().sum()
```

```
USD    0
dtype: int64
```

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 481 entries, 0 to 480
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Date    481 non-null    object
 1   USD     481 non-null    float64
dtypes: float64(1), object(1)
memory usage: 7.6+ KB
```

For further code, you can visit the GitHub repository link below:

https://github.com/amankumarsingh1092/Relative_Performance_Comparison_of_Different_Supervised_Learning_Classification_Algorithms.git

RESULTS AND OUTCOMES

14.1. EXCHANGE RATE TREND ANALYSIS (USD/INR)

Overview and Summary Statistics

Period: **April 2023–March 2025**

Entries: 481 observations per day

USD (US dollar to Indian rupee) is the currency being tracked.

Statistic	Value
Mean Rate	₹83.68
Median Rate	₹83.36
Minimum Rate	₹81.65
Maximum Rate	₹87.59
Standard Dev.	₹1.32

Interpretation: The range between the lowest and highest values of the exchange rate is about ₹6, indicating moderate fluctuation. Relative stability with sporadic spikes is indicated by the standard deviation.

14.2 LINE CHART: USD/INR EXCHANGE RATE

The USD to INR exchange rate's daily fluctuations over a two-year period, from April 2023 to March 2025, are depicted in the line chart below. The general trend and variations of the Indian Rupee (INR) in relation to the US Dollar (USD) are clearly shown by this graphic.

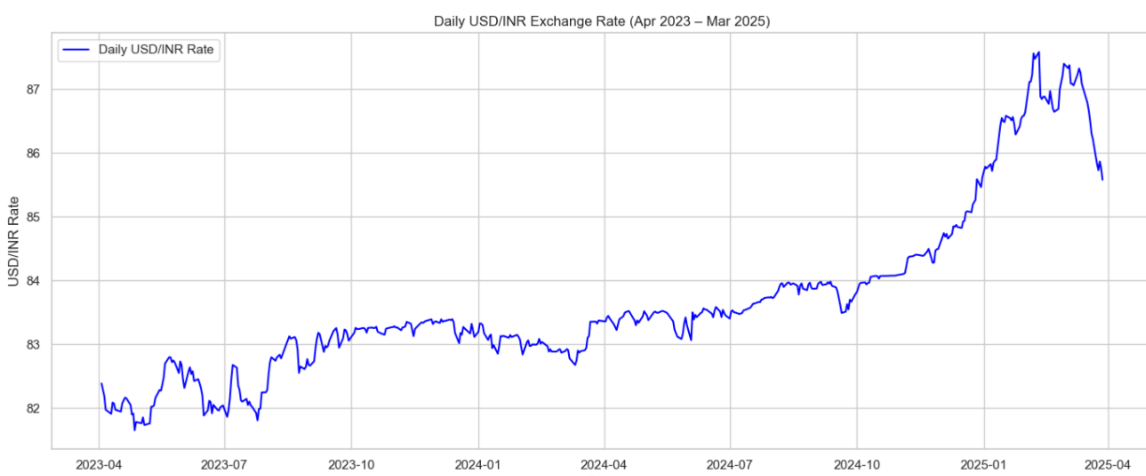


*Fig.*39: Daily USD/INR Exchange Rate (Apr 2023 – Mar 2025)

KEY OBSERVATIONS:

- The Indian Rupee's gradual depreciation over time is indicated by the exchange rate's upward trajectory.
- Early 2023 saw the lowest value (~₹81.65), and early 2025 saw the highest value (~₹87.59).
- The line shows a generally smooth trend, indicating controlled market dynamics with few sharp spikes, despite sporadic short-term volatility**.**

INTERPRETATION: Important macroeconomic variables like trade imbalances, inflation, geopolitical risks, and changes in interest rates globally are all reflected in the exchange rate line chart. It is consistent with the RBI's plan to reduce excessive volatility and increase exports by permitting gradual depreciation.
Long-term financial forecasting, trade competitiveness, and monetary policy can all be better understood with the help of this analysis.

14.3 30-DAY ROLLING AVERAGE PLOT: USD/INR EXCHANGE RATE

The USD/INR exchange rate from April 2023 to March 2025 is shown in a smoothed visual form using the 30-Day Rolling Average Plot. This method is frequently applied in time series analysis to highlight the underlying trend in financial data and eliminate short-term noise.

WHAT THE PLOT DISPLAYS:

- Because of market volatility, economic news, and global events, the daily exchange rate line (shown in lighter color) exhibits frequent fluctuations.
- By taking the average of the previous 30 days at each point, the 30-day rolling average line (bold/red) smoothes out these swings.
- Without being skewed by daily fluctuations, this rolling line shows the long-term trajectory of the Rupee's value against the US dollar.

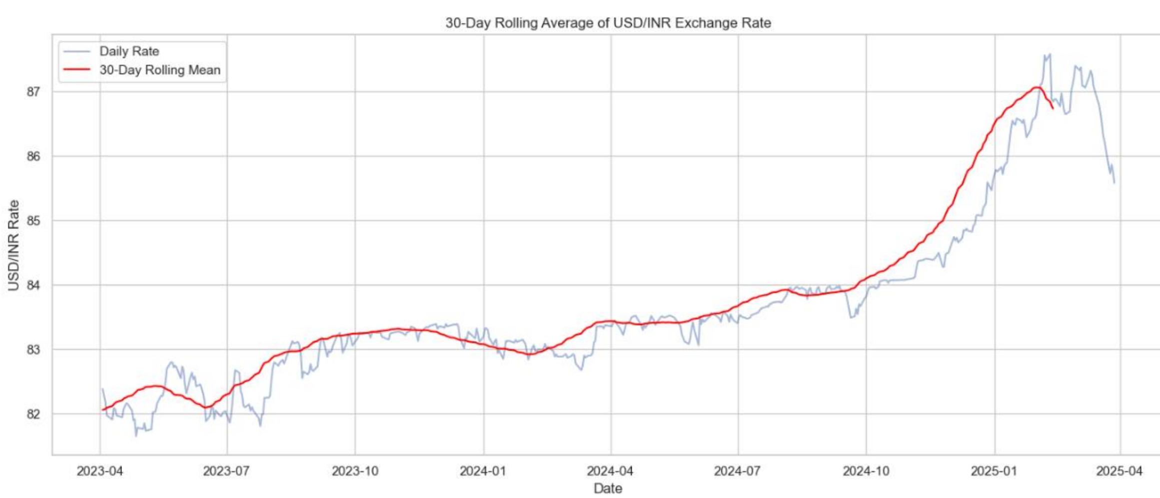


*Fig.*40: 30-Day Rolling Average of USD/INR Exchange Rate

KEY FINDINGS:

- Throughout the examined period, the INR consistently weakened, as evidenced by the rolling average's steady upward slope.
- Small plateaus in the rolling line indicate brief intervals of stability, which may have been brought on by RBI intervention or better economic fundamentals at the time.
- Late 2024 and early 2025 saw the rolling average's steepest upward slope, suggesting a more rapid depreciation phase during that time.

INTERPRETATION:

The 30-day rolling average emphasizes the INR's steady decline against the USD while highlighting brief intervals of stability. The influence of long-term economic circumstances and policy choices on currency value is reflected in this trend. The figure provides important information for economic analysis and forecasting since it shows macro-financial stress, which may be caused by inflation, trade shocks, or fiscal deficits.

14.4 MONTHLY AVERAGE AND PERCENTAGE CHANGE

The Monthly Average and Percentage Change chart offers a macro-level view of the monthly performance of the Indian Rupee (INR) in relation to the US Dollar (USD) during the two-year period between April 2023 and March 2025. This dual-axis visualization provides a clear picture of trends and changes in currency valuation by combining the average monthly exchange rate with the month-over-month percentage change.

WHAT THE CHART SHOWS:

- The bar graph illustrates how the value of the Rupee changed over time by displaying the average USD/INR rate for each month.
- The line graph shows the direction and magnitude of monthly movements by superimposing the percentage change from the prior month.

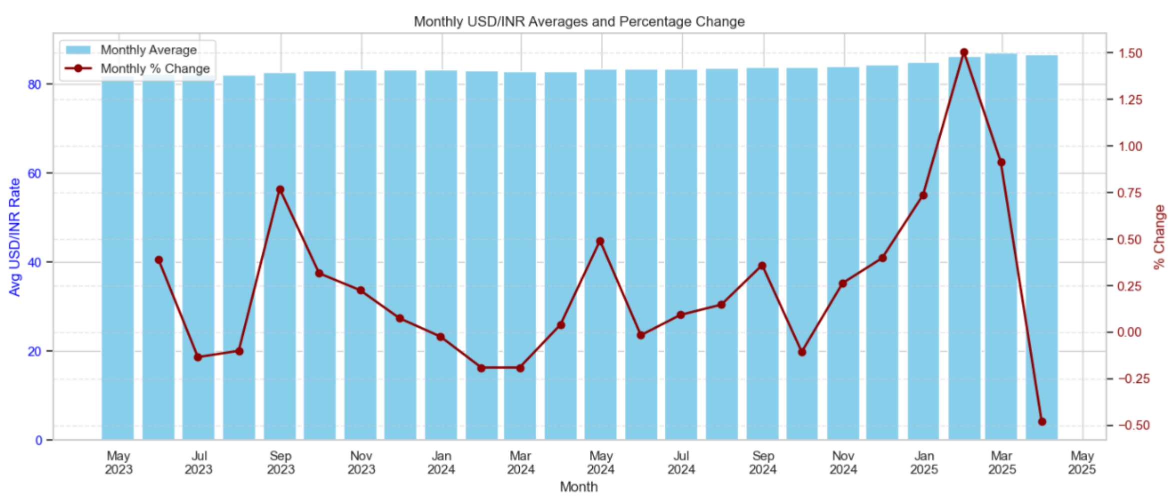


*Fig.*41: Monthly USD/INR Average and Percentage Change

KEY OBSERVATIONS:

- The monthly average indicates a steady rise, indicating the Rupee's long-term decline.
- From October 2024 to February 2025, there are notable periods of steeper depreciation that are characterized by steady increases in percentage, with some months seeing a peak of about +1.5%.
- After months of decline, a small negative percentage change in March 2025 indicates a mild correction or stabilization of the exchange rate.

INTERPRETATION:

The information shows that the Indian rupee has been steadily declining, most likely as a result of trade deficits, capital flight, and changes in interest rates around the world. Identifying periods of policy impact, such as RBI interventions, and volatility spikes is made easier by the percentage change line.

- This chart is useful because:
- Monitoring changes in volatility over time
- Assessment of the regularity of depreciation and appreciation
- Making important features for time series classification and forecasting models

14.5 CONCLUSION:

[A] Data Summary
Daily USD to INR exchange rates from financial institutions or the RBI's published rates are included in the uploaded dataset. Four hundred and eighty-one observations were made between April 2023 and March 2025.

a. Average Exchange Rate: ₹83.68

b. Maximum: ₹87.59

c. Minimum: ₹81.65

d. Overall Trend: The USD/INR moved upward, indicating that the Indian Rupee was depreciating.

[B] Potential Impact of the RBI and Policy Consequences

a. Exchange Rate Control
The current account deficit and external shocks like global interest rate hikes and geopolitical tensions were probably the main causes of the Indian Rupee's steady depreciation in 2024 and the first part of 2025. The steady pattern with few peaks indicates that the RBI permitted a slow depreciation to boost exports while potentially stepping in to control excessive volatility.

b. Controlling inflation

The cost of imports, particularly oil and necessities, rises when the rupee depreciates, which puts pressure on inflation. The RBI may have tightened monetary policy in response by:

Increasing the repo rates To absorb excess liquidity, open market operations are carried out. The goals of these policies are to preserve economic equilibrium and stabilize prices.

c. Reserves of Foreign Exchange

According to the data, the RBI did not aim for a fixed exchange rate; rather, it used its foreign exchange reserves to control short-term volatility. Persistent depreciation may indicate a balanced intervention strategy or a strategic shift toward market-driven rates or pressure on reserves.

d. Capital Movements and Investor Attitude

Exchange rate fluctuations have an impact on foreign institutional investors (FIIs). The RBI frequently steps in to preserve the stability of the domestic market when a falling rupee leads to capital flight.

[C] Observations of Monthly Volatility

The exchange rate increased gradually between November 2024 and February 2025, reaching its highest point in February 2025. March's minor correction (-0.48%) might be a sign of RBI stabilization or intervention as a result of better macroeconomic conditions.

[D]Implications for Strategy

Examining trends in exchange rates is essential for:

- Importers and exporters: To evaluate the effects on expenses and earnings
- Investors: To determine the direction of the RBI policy
- Policymakers: To connect inflation, growth, and trade with currency trends
- Forecasting Models: To provide precise data for predictive tools such as LSTM, ARIMA, and others

The Rupee's consistent depreciation between 2023 and 2025 is indicative of the RBI's emphasis on controlling volatility while striking a balance between growth, inflation, and external stability.

15. CMIE (CENTRE FOR MONITORING INDIAN ECONOMY PVT.LTD.)

Founded in 1976, the Centre for Monitoring Indian Economy (CMIE) is one of India's top independent economic think tanks. It helps researchers, analysts, companies, and policymakers make well-informed decisions by offering top-notch, up-to-date data and analytical tools on financial markets, industry trends, and macroeconomic conditions. Here are some ways that CMIE is a useful instrument for economic analysis:

15.1 IMPORTANT FEATURES OF CMIE PLATFORM:

15.1.1. ECONOMIC INDICATORS

Indicator	Value	Change	Base Period	Interpretation
Unemployment Rate (30-day avg.)	6.7%	-0.1%	–	slightly higher employment rates.
Consumer Sentiments Index	114.5	0.0	Sep–Dec 2015	Consumer confidence is steady overall.
Consumer Expectations Index	113.1	0.0	Sep–Dec 2015	The future outlook of consumers remains unchanged.
Current Economic Conditions Index	116.7	0.0	Sep–Dec 2015	The way that consumers perceive the state of the economy is unaltered.

15.1.2 QUARTERLY FINANCIALS OF LISTED COMPANIES:

Updated on: 16 June 2025, 05:30 AM

A. ALL LISTED COMPANIES

Indicator	Jun '24	Sep '24	Dec '24	Mar '25
Income	9.4%	7.3%	6.9%	6.4%
Expenses	10.6%	8.3%	6.1%	6.9%
Net Profit	4.6%	1.3%	8.9%	9.5%
PAT Margin (%)	9.9%	10.0%	9.9%	10.1%
Count of Companies	4,649	4,588	4,600	4,367

B. NON-FINANCIAL COMPANIES

Indicator	Jun '24	Sep '24	Dec '24	Mar '25
Income	6.5%	4.7%	5.1%	5.5%
Expenses	8.2%	6.2%	4.3%	5.8%
Net Profit	-2.7%	-8.7%	5.4%	11.6%
PAT Margin (%)	7.8%	7.6%	7.7%	8.3%
Net Fixed Assets	–	4.7%	–	7.2%
Current Assets	–	8.8%	–	10.5%
Current Liabilities	–	3.0%	–	7.8%
Borrowings	–	8.3%	–	6.0%
Reserves & Surplus	–	13.3%	–	12.8%
Count of Companies	3,472	3,448	3,466	3,313

Key Insights:

- Revenue, profit, and margin improvements were all steadily increasing every quarter.
- June and September 2024 saw non-financial companies' profit growth turn negative, but March 2025 saw a robust recovery (+11.6%).
- Across all sectors, PAT margins increased, suggesting increased profitability.
- Increases in current assets, borrowings, and reserves show that non-financial firms have become more financially stable.
- A drop in listed companies from year-to-year points to delisting or industry consolidation.

15.1.3 ANNUAL FINANCIALS OF ALL COMPANIES:

A. ALL LISTED COMPANIES

Indicator	FY23	FY24	FY25
Income	18.4%	6.8%	9.6%
Expenses	18.4%	5.0%	8.9%
Net Profit	9.3%	36.1%	15.7%
PAT Margin (%)	5.1%	6.7%	12.2%
Assets	10.6%	10.8%	10.8%
Net Worth	11.7%	13.8%	17.5%
RONW (%)	9.9%	12.1%	18.6%
Count of Companies	35,177	29,786	258

B. NON-FINANCIAL COMPANIES

Indicator	FY23	FY24	FY25
Income	18.4%	4.7%	7.3%
Expenses	18.8%	3.2%	7.6%
Net Profit	-3.1%	35.7%	5.3%
PAT Margin (%)	3.9%	5.3%	12.4%
Net Fixed Assets	5.8%	6.0%	7.4%
Net Worth	9.9%	11.8%	15.6%
RONW (%)	9.6%	12.1%	23.1%
Debt / Equity (times)	0.9	0.8	0.3
Interest Cover (times)	3.0	3.5	11.0
Net Working Capital Cycle (days)	65	66	26
Count of Companies	27,800	23,798	193

KEY TAKEAWAYS:

- All businesses and non-financial firms saw a significant improvement in profitability in FY25, as evidenced by higher PAT margins and Return on Net Worth (RONW).
- A recovery in the base effect was a major factor in FY24's robust net profit growth.
- Non-financial businesses demonstrated greater interest coverage and lower debt-to-equity ratios, indicating greater financial efficiency.
- Stronger liquidity and operational efficiency are indicated by a significant decrease in the net working capital cycle (from 66 to 26 days).
- Data lag or continuous database updates are probably to blame for the dramatic drop in the number of companies reported for FY25.

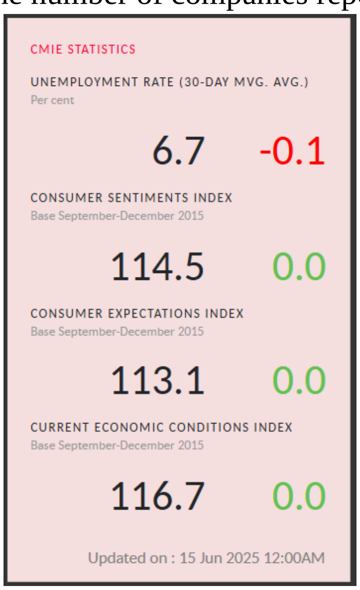
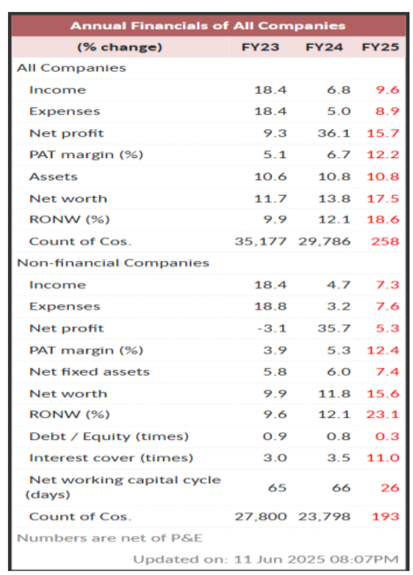


*Fig.*42: CMIE Statistics



Annual Financials of All Companies

(% change)	FY23	FY24	FY25
All Companies			
Income	18.4	6.8	9.6
Expenses	18.4	5.0	8.9
Net profit	9.3	36.1	15.7
PAT margin (%)	5.1	6.7	12.2
Assets	10.6	10.8	10.8
Net worth	11.7	13.8	17.5
RONW (%)	9.9	12.1	18.6
Count of Cos.	35,177	29,786	258
Non-financial Companies			
Income	18.4	4.7	7.3
Expenses	18.8	3.2	7.6
Net profit	-3.1	35.7	5.3
PAT margin (%)	3.9	5.3	12.4
Net fixed assets	5.8	6.0	7.4
Net worth	9.9	11.8	15.6
RONW (%)	9.6	12.1	23.1
Debt / Equity (times)	0.9	0.8	0.3
Interest cover (times)	3.0	3.5	11.0
Net working capital cycle (days)	65	66	26
Count of Cos.	27,800	23,798	193

Numbers are net of P&E

Updated on: 11 Jun 2025 08:07PM

*Fig.*43: Annual Financial Statements of All Companies

A. High-frequency, Real-time data

Near real-time information on important economic metrics, including employment, inflation, investment, consumption, and production, is provided by CMIE. This facilitates prompt policy responses and early economic shift detection. Studies such as "Quantitative and Qualitative Improvement in Jobs" and "Employment Bounces Back," for instance, track the post-COVID employment recovery using high-frequency data, highlighting trends in both the quantity and quality of jobs.

B. Sector-Specific Insights

CMIE offers comprehensive insights into important industries, including manufacturing, services, FMCG, and power, emphasizing the cost, profitability, and growth trends necessary for sectoral analysis. Reports like "Non-Fin Services Profits Soar" and "FMCG Cos. Margins Improve" show shifting profit dynamics, assisting analysts in monitoring economic drivers and investment trends.

C. Tracking Financial and Corporate Performance

Business strategy, credit research, stock analysis, and comparative financial evaluations are made possible by CMIE's platforms, such as Prowess, which provide access to the financial statements of thousands of Indian companies. For instance, information from "POL Cos. Drag Mfg. Profit Growth in FY25" aids in assessing the effects of shifting commodity prices or policy changes on the profitability of the manufacturing sector.

D. Analysis of Macroeconomic and Policy Impacts

By providing information on important national indicators like GDP growth, inflation, fiscal deficits, monetary trends, and developments in the external sector, CMIE makes it possible to analyze how policy choices affect economic results. As an illustration of how commodity availability affects infrastructure demand and economic activity, consider the report "Coal to Pull Back Rail Freight Growth in Q1."

E. The Rural Economy and Consumer Attitude

CMIE's Consumer Pyramids Household Survey (CPHS) provides a thorough understanding of household income, spending, employment, and sentiment trends, which makes it useful for researching rural dynamics and consumption-driven growth. For instance, "Good Kharif Odds Lift Rural Sentiments" illustrates how monsoon expectations increase rural confidence, which affects demand and inflation, using CPHS data.

F. Behavioural Analysis and Political Economy

In order to demonstrate how social and political factors affect economic outcomes, CMIE also looks at voter preferences and political behaviour. Election trends that can influence or reflect local economic priorities are highlighted in reports such as "Voter Preferences Shift in Big States" and "Bandwagon Effect."

G. Relationships between Climate and the Environment

CMIE finds risks and vulnerabilities, particularly in infrastructure, agriculture, and rural livelihoods, by combining economic trends with climate data. One example is "Southwest

Monsoon Stalls," which emphasizes how delayed monsoons can upset economic equilibrium, especially in rural areas.

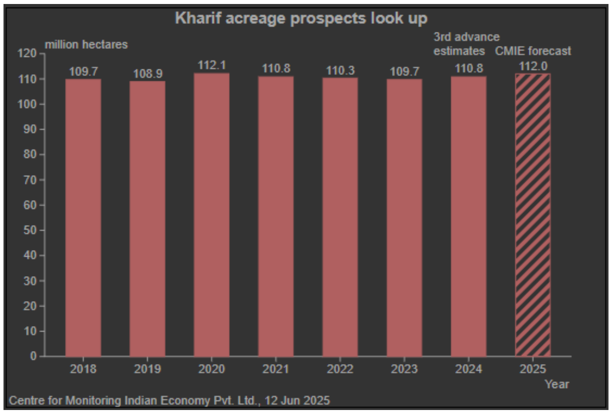


*Fig.*44: Kharif acreage prospects look up.

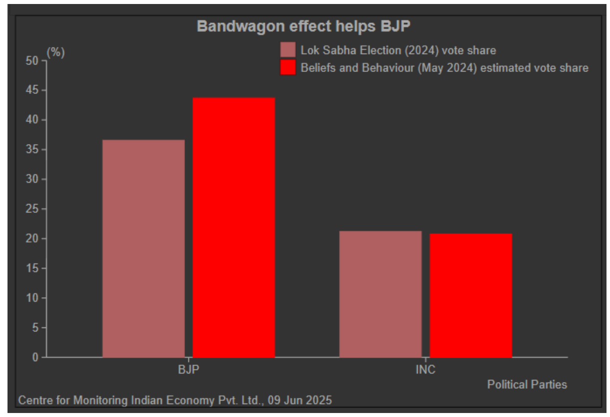


*Fig.*45: Bandwagon effect helps the BJP

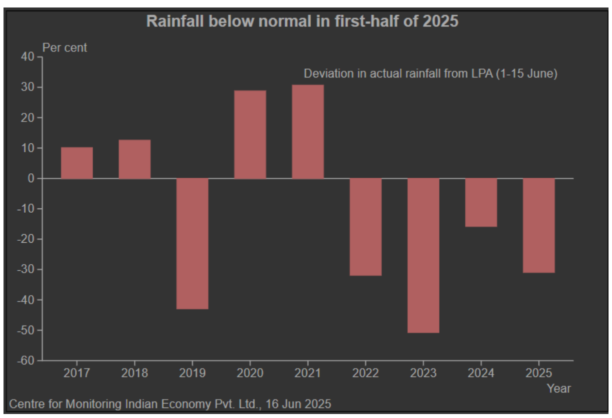


*Fig.*46: Rainfall below normal in the first half of 2025.

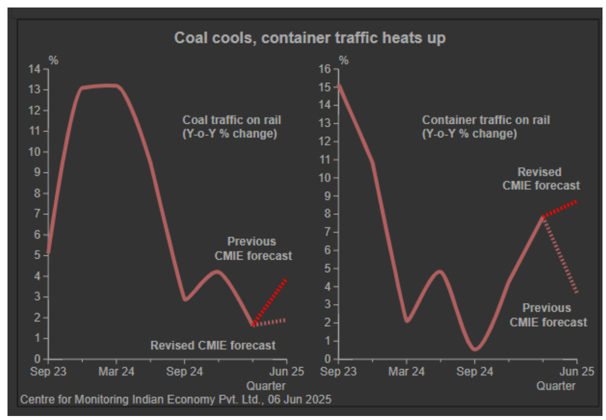


*Fig.*47: Coal cools, container traffic heats up

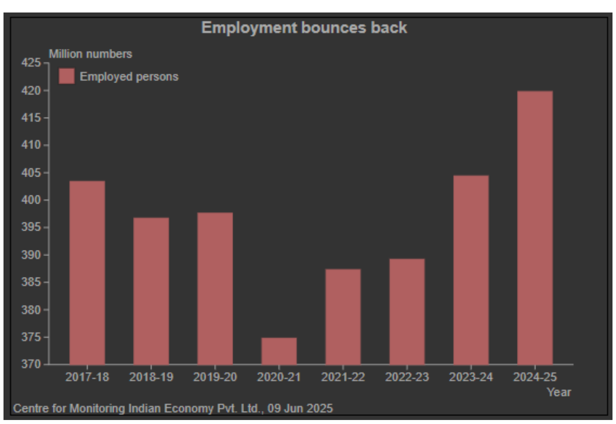


*Fig.*48: Employment bounces back.

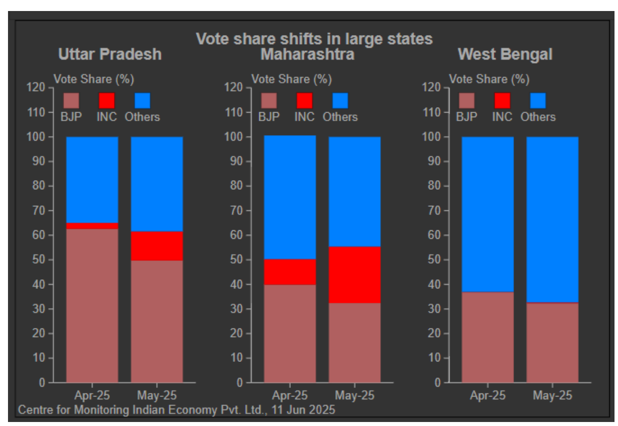


*Fig.*49: Vote share shifts in the large state of Maharashtra

16. CONCLUSION

Using real-world financial and economic data from sources such as the BSE, NSE, Bloomberg, RBI, and CMIE, this project showed how supervised learning algorithms can be applied. Models like Random Forest, Decision Tree, SVM, k-NN, Logistic Regression, and Naive Bayes were tested for forecasting currency trends and stock price movements using data from FY 2023–24 and FY 2024–25.

- Decision Tree was excellent at identifying upward trends and capturing true positives.
- Random Forest provided more accuracy and generalization.
- Naive Bayes and logistic regression performed poorly because of their strict presumptions.

The RBI exchange rate data showed controlled Rupee depreciation and policy intervention, while technical indicators (volatility, moving averages, and daily returns) enhanced prediction. Case studies on TCS, Infosys, Reliance, and HDFC Bank improved sector-specific analysis, and CMIE and Bloomberg contributed useful macroeconomic and corporate data. All things considered, the study demonstrates that machine learning supports risk management, investment choices, and market strategy by assisting with financial forecasting. Deep learning, sophisticated feature engineering, and class imbalance handling (like SMOTE) are possible future advancements.

17. REFERENCES

1. Workineh Menna Eligo, et al. (2022). *Comparing Supervised Machine Learning Algorithms on Classification Efficiency of Multiclass Classification Problems*.

2. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). *Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques*. Expert Systems with Applications.

3. Huang, W., Nakamori, Y., & Wang, S. Y. (2005). *Forecasting stock market movement direction with support vector machine*. Computers & Operations Research.

4. Dash, M., & Dash, P. K. (2016). *A comparative study of supervised machine learning algorithms for stock market trend prediction*. Financial Innovation.

DATA SOURCES

- **BSE (Bombay Stock Exchange)**
  Source: https://www.bseindia.com
- **NSE (National Stock Exchange)**
  Source: https://www.nseindia.com
- **Bloomberg Terminal**
- TCS: *Tata Consultancy Services Ltd Stock Price*
  TCS: Tata Consultancy Services Ltd Stock Price Quote - Natl India - Bloomberg

- Infosys: *Infosys Ltd Stock Price*
  INFO: Infosys Ltd Stock Price Quote - Natl India - Bloomberg
- Reliance: *Reliance Industries Ltd Stock Price*
  RIL: Reliance Industries Ltd Stock Price Quote - - Bloomberg
- HDFC Bank: *HDFC Bank Ltd Stock Price*
  HDFCB: HDFC Bank Ltd Stock Price Quote - Natl India - Bloomberg
- **RBI (Reserve Bank of India)**
  USD-INR Exchange Rates Archive:
  https://www.rbi.org.in/Scripts/ReferenceRateArchive.aspx
- **CMIE (Centre for Monitoring Indian Economy)**
  Source: https://www.cmie.com
- **Image & Code References**
- **Jupyter Notebook Visualizations**: Using Python in Jupyter Notebook and libraries like matplotlib, seaborn, pandas, and scikit-learn, all analytical graphs and model evaluation plots (such as line charts, box plots, ROC curves, and confusion matrices) were produced *Source: Created by the author as a component of the project.*
- **Geeks for Geeks Diagrams**: Geeks for Geeks provided a few conceptual illustrations that were modified for educational purposes, such as machine learning types and confusion matrices. *Source: GeeksforGeeks. (n.d.).https://www.geeksforgeeks.org/supervised-machine-learning/*
- **GitHub Repository**: You can find the full Python code and visualizations at: *https://github.com/amankumarsingh1092/Relative_Performance_Comparison_of_Different_Supervised_Learning_Classification_Algorithms*