Stock Market Prediction and Portfolio Optimization Using Machine Learning: Evidence from India

A THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE dEGREE OF

DOCTOR OF PHILOSOPHY

in Commerce

IN THE Goa Business School
GOA UNIVERSITY

By

Nazareth Noella Avilia
Goa Business School
Goa University
Goa

APRIL *2025*

DECLARATION

I, Ms. Nazareth Noella Avilia, hereby declare that this thesis represents work which has been carried out by me and that it has not been submitted, either in part or in full, to any other University or Institution for the award of any research degree.

Place: Taleigao Plateau.

Date: 30-04-2025

Ms. Nazareth Noella Avilia

CERTIFICATE

I hereby certify that the above Declaration of the candidate, Ms. Nazareth Noella Avilia is true and the work was carried out under my supervision.

Prof. Y.V. Reddy

Goa Business School, Goa University

Acknowledgment

Every accomplishment within these pages speaks not of my abilities but of God's limitless grace and wisdom that has gently carried me through. I thank God for the gift of a curious mind to wonder, a heart eager to learn, the wisdom to discern, the health to endure, and the clarity to walk this journey to its completion. I thank the Almighty for the way He guided my transition from Junior Research Fellow to Senior Research Fellow in His perfect timing, reminding me that He has held every step of this journey in His hands. Every page in here, every insight, every breakthrough, every obstacle overcome, every lesson learned, stands as a reminder that without Him, none of this would have been possible.

I am extremely grateful to my guide, Senior Prof. Dr. Y.V. Reddy, a true gentleman in every aspect of life. His dedication to this journey was remarkable – never leaving any stone unturned and always willing to go to great lengths to ensure I had every support I needed. His patience with a questioning mind like mine and his quiet belief in my potential has left an indelible mark on this work. His unwavering commitment to quality inspired me to undertake meaningful research and contribute through impactful publications. His humility, wisdom, and constant reassurance created an environment where I could grow, explore, and learn without hesitation. It has been a privilege to learn under a mentor who so effortlessly combined brilliance with kindness. For his steadfast guidance and the many unseen ways he shaped this journey, I remain profoundly thankful.

I sincerely thank my Departmental Research Committee (DRC) expert members, Dr. Harip Khanapuri and Dr. Pournima Dhume, for their constructive suggestions. Their guidance and thoughtful critiques helped shape this research into what it is today.

I deeply appreciate the faculty and administrative staff of the Goa Business School, the Library, and the Academic and Examination Division of Goa University, whose cooperation and assistance have facilitated various aspects of my academic journey. I am equally thankful to the Sports Department for providing the opportunity to represent Goa University in cricket, which remains among the lighter joys of this journey.

I am thankful to my research colleagues, whose discussions and shared experiences have enriched this journey and contributed meaningfully to my growth as a researcher. A special note of thanks is due to my friends, whose encouragement, understanding, and support have been a source of strength and comfort during the more challenging phases of this endeavor.

No words could ever express the depth of gratitude I hold towards my beloved parents, who have given me the time and space to chase my academic dreams. Their constant prayers, quiet sacrifices, unconditional love, and unshakable belief in me have been the foundation upon which this achievement rests. They have celebrated every small victory and have been my greatest cheerleaders. This achievement carries their hopes, and every silent prayer whispered along the way. And therefore, this degree is as much theirs as it is mine.

I am grateful to my brothers, who, without many words, gave me the gift of silent courage. Even from a distance, their presence was a constant reminder that I was never walking this path alone.

For every guiding hand, every encouraging word, and every quiet prayer that carried me here, I thank you from the depths of my heart.

Noella Avilia Nazareth

Dedicated with deep gratitude to

My Parents,
Mr. Joseph Vincent Nazareth
Mrs. Irene Margaret Nazareth
Whose love and sacrifices shaped my journey,

&

My Guide,
Prof. Dr. Y.V. Reddy
Who recognized my potential well ahead of time,
And with whose guidance I reached beyond what I thought possible.

With your support, I have seen my dreams come true –
For all I have become is thanks to you.

Contents

List of Tables

List of Figures

Abbreviations

ANN: Artificial Neural Network

CEPS: Cash Earnings per Share

EBIT: Earnings before Interest and Tax

EV: Enterprise Value

EWP: Equally Weighted Portfolio

LSTM: Long-Short Term Memory

MDS: Multi-Dimensional Scaling

MI: Mutual Information

MLP: Multi-Layer Perceptron

MSR: Maximum Sharpe Ratio

MVP: Minimum Volatility Portfolio

ROI: Returns on Investment

SVM: Support Vector Machine

Chapter 1

Introduction

Machine learning, a specialized branch of artificial intelligence, has emerged as a widely researched area for its diverse applications across various fields. It can process vast volumes of data with unparalleled speed and precision. Moreover, it enables systems to automatically learn and improve from past experiences without being explicitly programmed. The numerous advantages of machine learning techniques have carved their way into all spheres, including finance. The most common application of these tools is process automation, which reduces repetitive tasks and replaces manual work, providing higher customer satisfaction, reducing time, and optimizing costs. Banks and Financial Institutions are also absorbing the benefits of machine learning tools by using them for financial monitoring (Kong, 2024), risk detection and management (Murugan & Kala T, 2023), determining credit scores (Gu, Wu, Hu, & Yu, 2024), recommending financial products (Sharaf, El-Din Hemdan, El-Sayed, & El-Bahnasawy, 2022), detecting fraudulent transactions, laundering money (Domashova & Zabelina, 2021), underwriting and Robo-advisory (Shanmuganathan, 2020) amongst others.

Financial robo-advisors and chatbots provide banking assistance to clients, asset allocation systems provide risk-return assessments to investors, whilst automated insurance services are made available to policyholders; the financial applications of machine learning are interminable. To highlight a few real-world examples of machine learning applications in the financial sector, the ING Group adopted an AI-based credit risk system to alert lenders whether there are likely to be any issues in debt repayment. Investment companies like Betterment and Wealthfront provide robo-advisory services to customers when handling their investments. ZestFinance helps build credit risk models, while Lemonade has automated its insurance services (Cheung, 2020).

Algorithms form the base for running any machine learning tool, which assists a system in automatically learning from past experiences without human intervention. The main aim of using such an expert system is to continuously improve the algorithm's performance. Machine learning models can identify trends or patterns in data with high accuracy.

Moreover, these models are competent in handling complex and voluminous data, of which the financial sector has an excess. Due to its ability to analyze nonlinear data, which is dominant in financial time series, machine learning techniques have gained an edge over traditional tools and techniques.

Predicting the stock market continues to be a fascinating yet challenging area of research due to the financial time series being noisy, chaotic, and non-stationary. Despite the Efficient Market Hypothesis theory proposed by Fama (1965) suggesting the impossibility of market outperformance, analysts contend that its prediction based on technological developments can help traders gain monetary benefits. Understanding the stock market and its innumerable interconnected factors has gained the attention of investors and researchers. A wide range of factors, including global interactions in foreign investments, political turmoil, natural calamities, company and management-related changes, financial news, changes in exchange rates, inflation, market sentiments, and social moods of the people, are some acts that can affect the stock market. Nevertheless, knowing these factors and guesstimating their impact on the stock market is insufficient. There is a need to incorporate these factors into a predictive model by converting the available data into valuable features, which are significant attributes of machine learning and deep learning models. The stock market time series being noisy and volatile has proved to be a test of robustness to determine the performance of these machine learning models.

Traditionally, the trend of future stock prices was determined by analysts or investors using fundamental analysis (by studying the financial statements or accounting information of companies) or technical analysis (by studying past stock prices and trading volume). Research shows that with technological advancements, peoples' sentiments, news updates, blogs, and web searches also shape stock market prices alongside other economic, political, and social factors. Analysts deploy machine learning techniques in stock market prediction due to their higher degree of accuracy.

While managing portfolios, investors allocate their wealth across various assets to generate initial profits and periodically rebalance their portfolios (Li, 2019). This requires making swift investment decisions while navigating vast amounts of financial data. However, the human brain does not easily and efficiently process such vast data records. Machine learning tools, particularly neural networks, offer a solution to this challenge. The neural network model mimics the human brain's ability to recognize complex patterns and relations in data, uncovering insights that may go unnoticed by humans. As a result, neural networks have become indispensable in finance and investment.

1.1 Need for the study

Despite the abundance of literature on stock market prediction, researchers have given limited attention to evaluating market behavior from the perspective of event-induced volatility in emerging economies. Markets exhibit varying responses to different types of events, particularly unanticipated ones. Anticipated events often lead to calculated responses, while unforeseen events trigger varied reactions depending on the timing and context. Understanding these dynamics is crucial, as they significantly impact investor behavior, market stability, and decision-making. Thus, this study attempts to conduct an empirical analysis, which would have significant implications for all market stakeholders.

1.2 Scope of the study

The present study focuses on applying machine learning techniques in predicting the Nifty broad market and sectoral indices alongside the closing prices of stocks from 2012 to 2024. The predictions are made daily, over various periods covering both unforeseen and scheduled events. The scheduled events include the implementation of the GST regime in 2017 and the national general elections of 2019 and 2024. Unforeseen events include demonetization in 2016, the global COVID-19 pandemic in 2020, the invasion of Ukraine by Russia in 2022, and the Hamas attack on Israel in 2023. The stock market data includes the daily open, high, and low prices of the selected indices/stocks. Meanwhile, the macroeconomic variables include the U.S. Dollar – Rupee exchange rate, broad money (M3), gold prices, and GDP at market price. The feature selection results will help stakeholders identify the most relevant features, enhancing market predictions' accuracy and efficiency.

Secondly, the study also evaluates and compares the performance of equity portfolios constructed by an unsupervised machine learning model: K-Means Clustering. We use Monte Carlo simulations to allocate funds and create an Equally Weighted Portfolio (EWP), a Maximum Sharpe Ratio portfolio (MSRP), and a Minimum Volatility Portfolio (MVP). The portfolios are created and evaluated over the same periods (events: unforeseen and scheduled) selected for the first objective. The stocks considered for creating these portfolios belong to the Nifty100 index as of the day of construction of the portfolios. We select a set of representative financial variables, including the stock's returns, standard deviation, and various validation ratios. The findings of this research can guide less-informed investors in constructing more informed and optimized portfolios.

1.3 Objectives of the study

The objectives of the study are:

Objective I: To apply machine learning techniques in developing a model that predicts the Indian stock market during scheduled and unanticipated events

Sub Objectives:

a) To develop a model that predicts the closing price of the broad market indices of the NSE

b) To develop a model that predicts the closing price of the sectoral indices of the NSE

c) To develop a model that predicts the closing price of stocks listed on the NSE

Objective II: To construct portfolios using an unsupervised machine learning technique and evaluate portfolio allocation strategies during scheduled and unanticipated events

Sub Objectives:

a) To apply K-means clustering to construct portfolios

b) To allocate funds amongst stocks in the constructed portfolio

c) To evaluate the performance of each constructed portfolio and allocation strategy

1.4 Significance of the study

This study addresses a critical aspect of stock market analysis by leveraging machine learning models to predict the Nifty broad market and sectoral indices during scheduled and unanticipated events. By focusing on these event-driven scenarios, this research provides a deeper understanding of market behavior under varying conditions, which is especially relevant for the dynamic and volatile nature of emerging markets like India. The study also evaluates the predictive performance of these models in forecasting the closing prices of stocks listed on the NSE, India. This evaluation adds granularity to the analysis, ensuring that the models are effective not only at the broader market level but also at the individual stock level. Moreover, through feature selection, this study will highlight the stock market data and/or macroeconomic indicators that are most relevant in predicting the market. This will enable investors, analysts, and policymakers to make better decisions and formulate strategies.

Further, by applying K-means clustering, the study provides insights into the performance of machine learning-constructed portfolios based on the financial characteristics of the considered stocks. Different fund allocation strategies highlight the performance of these models compared to the performance of the Nifty benchmark indices. The study captures the impact of event-driven volatility on investment strategies by constructing and evaluating portfolios around scheduled and unforeseen market events. This approach is particularly relevant for emerging markets like India, where event-driven fluctuations significantly influence stock performance. The study's findings are critically important to investors, fund managers, and policymakers. Innovative machine learning and simulation techniques can aid in better decision-making and risk management.

1.5 Limitations of the study

This study examines stock market predictions from 2012 to 2024, selected in line with India's base year. India's base years are reference years for various statistical measures, such as GDP, inflation, and industrial production. These base years are periodically updated to reflect the changing structure of the economy. Prior to 2011-12, India updated its base years for GDP and other indicators from 2004-05, and before that in 1999-2000. These base years are critical for making year-on-year comparisons and formulating policies based on reliable economic data. Consequently, the analysis does not include the evaluation of model performance during the significant financial crisis 2008, which profoundly impacted the Indian economy.

Furthermore, three commonly used variables in financial literature - the Consumer Price Index (as a proxy for inflation), the 91-day Treasury Bill rate (representing interest rates), and the unemployment rate—were excluded from the study due to the unavailability of data beginning from 2012. These omissions highlight areas for future research that could enhance the robustness of market prediction models.

This study focuses on constructing portfolios using stocks within the Nifty100 index, excluding the potential exploration of portfolios comprising a broader range of stocks. The clustering performance on a larger dataset remains unexplored and could provide valuable insights on an expanded data pool. Additionally, the study utilizes a specific set of financial indicators for clustering, leaving room to investigate the impact of alternative financial metrics on clustering performance.

Fund allocation strategies are confined to three approaches: the EWP, MSRP, and MVP. While these strategies are robust, future research could delve into advanced allocation methods such as the Risk Parity Portfolio, Factor-Based Portfolio, or Minimum Drawdown Portfolio, among others, to further enhance portfolio performance and diversification.

1.6 Chapterization

Chapter 1 focuses on the need for the study, the scope of the study, the objectives of the study, the significance of the study, the limitations of the study, and the scheme of chapterization.

Chapter 2 provides a theoretical background of the study.

Chapter 3 provides an objective-wise literature review and identifies the research gap.

Chapter 4 elaborates on the detailed research methodology tailored to address the objectives of the study.

Chapter 5 presents the results and discussion on stock market prediction, offering insights into the models' performance.

Chapter 6 focuses on the results and discussion of portfolio optimization.

Chapter 7 summarizes the key findings, the conclusion, policy implications, and the contribution of the study. Additionally, it provides suggestions for future research avenues.

Chapter 2

Theoretical Background

This section provides the theoretical foundation for each objective, offering readers a clearer understanding of the technical concepts and methodologies explored in this study.

2.1 Feature Selection

Feature selection refers to selecting a subset of relevant features from a larger set of available features. The goal of feature selection is to identify the most informative (Zhang, Chen, Bu, & He, 2012) that contribute significantly to the overall predictive power of a machine learning model. Features that provide similar or redundant information can lead to multicollinearity and increase the complexity of the model without adding significant value. Feature selection techniques identify these redundant features for elimination, thereby simplifying the model and reducing overfitting. It is an important step in the process of machine learning as it helps to:

- Improve model performance: Irrelevant or redundant features can introduce noise, increase model complexity, and lead to overfitting. Feature selection helps eliminate such features, enabling the model to focus on the most informative aspects of the data and enhancing its predictive accuracy.
- Reduce dimensionality of the dataset: Feature selection techniques help reduce the dimensionality of the dataset by identifying a subset of relevant features. High-dimensional datasets can pose challenges such as increased computational complexity, longer training time, and the risk of overfitting. By reducing the number of features, the models enhance efficiency, requiring fewer computational resources and reducing the risk of overfitting.
- Enhance model interpretability: Having a smaller set of relevant features makes interpreting and understanding the relationships between the features and the target variable easier. It can provide insights into which features significantly impact the outcome, aiding decision-making and problem-understanding.

Several feature selection techniques are commonly used in machine learning, one of which is the Filter Method. Filter methods assess the relevance of features based on statistical measures or scoring techniques. They evaluate each feature independently of the learning algorithm. Since this study uses numerical input data and target variables, we use the technique of Mutual Information for feature selection.

2.1.1 Mutual Information

The concept of Mutual Information (MI) was introduced by Claude Shannon, an American mathematician and electrical engineer, often called the "father of information theory." Shannon's seminal work in the late 1940s laid the foundation for the field of information theory and introduced fundamental concepts and measures, including mutual information. Shannon's groundbreaking paper, "A Mathematical Theory of Communication," published in 1948, presented mutual information as a measure of the amount of information shared between two random variables. He described MI as a tool to gauge the decrease in uncertainty about one variable when the value of the other variable is known (Shannon, 1948). MI is a metric that quantifies the interdependence between two randomly sampled variables taken concurrently.

Mutual Information MI (X, Y) between 2 (absolutely) continuous random variables X and Y (also called rate of transmission) is defined as:

$$\text{MI}\ (X, Y) = \iint P\ (x, y) \log \frac{P(x,y)}{P(x)P(y)}\, dx\, dy \tag{1}$$

Where $P\ (x, y)$ is the joint probability density function of X and Y and $P(x)$ and $P(y)$ are the marginal density functions associated with continuous random pairs $(X,\ Y)$ (Cover & Thomas, 1991). The MI calculated between two continuous random variables is called differential mutual information (Zeng, 2015), (Pascoal, Oliveira, Pacheco, & Valadas, 2017). Variables with much information have higher entropy values, exhibiting more diverse and less predictable patterns. MI, derived from entropy, measures the extent to which possessing information about one variable decreases uncertainty about another variable. High MI indicates a significant reduction in uncertainty; low MI indicates a slight reduction; and zero MI between two random variables means the variables are independent. Thus, MI helps identify the most informative variables to the target variable.

2.2 Deep Learning Models

Deep learning is a subset of ML that uses artificial neural networks to learn from data. Deep neural networks form the basis of deep learning. A neural network is a series of algorithms that endeavor to recognize underlying relationships in data through a set of processes that mimic the operation of the human brain. Deep learning models consist of multiple hidden layers. Each layer progressively learns by transforming the input data through nonlinear transformations. This hierarchical representation allows deep learning models to discern intricate patterns in the data, making them highly effective. Thus, there are many issues involved in the development of neural networks. These include the choice of input variables, the number of hidden layers and neurons, and their activation functions, network architecture, and parameters. Stock market prediction requires a nonlinear dynamic system. ANNs and LSTMs can adjust and learn these nonlinear relationships between inputs and outputs, which are discussed in the following subsection.

2.2.1 Artificial Neural Networks (ANN)

An ANN is a computational model inspired by the functionality and structure of the human brain. It is a powerful ML technique for solving complex problems by mimicking the behavior of interconnected neurons in the brain.

The primary building block of an ANN is a neuron, also called a perceptron. These perceptrons are organized into layers within the neural network. ANNs have three main layers in a network: the input layer, the hidden layer(s), and the output layer. The input layer receives the features, while the output layer displays the results of the ANN. The hidden layer(s) perform intermediate computations, and it is here where most of the 'learning' occurs. A perceptron takes multiple input values, applies weights, and passes the weighted sum through an activation function to produce an output. The activation function determines whether the neuron should or should not be activated based on the input it receives. The input into the activation function is the weighted sum of the input features from the preceding layer.

Weights represent the connections between neurons in different layers. Each weight represents the strength of the connection between two neurons and determines the influence of one neuron's output on another's input. During the learning process, the neural network adjusts these weights to optimize performance on a specific task.

ANNs can learn complex patterns and make accurate predictions in various disciplines, such as image and speech recognition, natural language processing, or time-series forecasting. They can handle large amounts of data and generalize from examples to generate predictions on unseen data.

The following equation shows the relationship between nodes, weights, and biases.

$$y = f(\sum_{i=1}^{n} x_i w_i + b) \tag{2}$$

The weighted sum of inputs for a layer is passed through a nonlinear activation function to another node in the next layer. It can be interpreted as a vector,

where,

$x_1, x_2, \ldots,$ and x_n are inputs,

$w_1, w_2, \ldots,$ and w_n are weights respectively,

n is the number of inputs for the final node,

f is activation function,

b is the bias term added to the weighted sum, and

y is the output of the node after applying the activation function f.

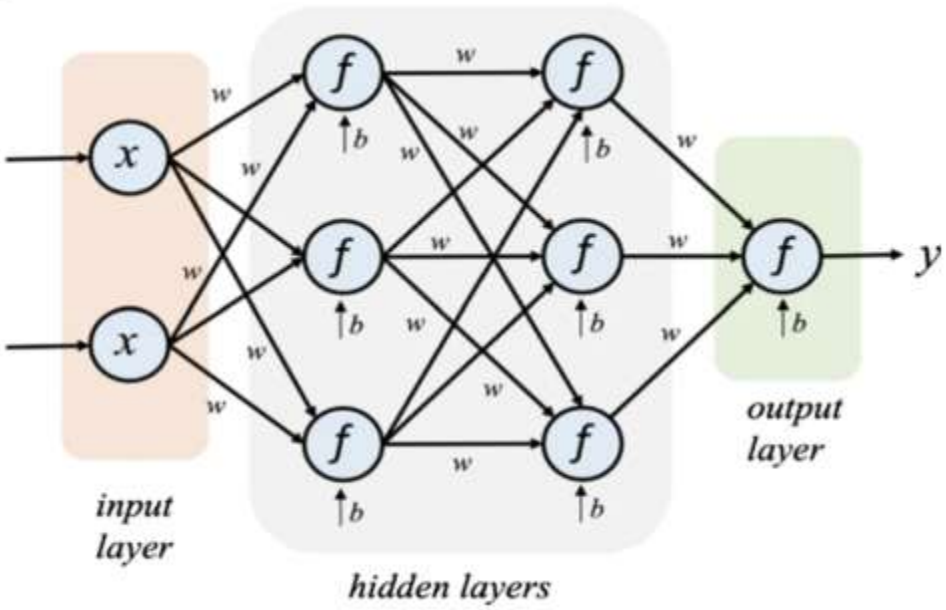


Figure 2.1: An ANN with two hidden layers and a single output.

Source: Ünal & Başçiftçi, (2022)

2.2.2 Long Short-Term Memory (LSTM)

Predicting stock market movements necessitates understanding a dynamic system because of the highly volatile nonlinearity of financial time series (Qiu, et al., 2016). Price return time series exhibit certain regularities associated with market patterns, suggesting the potential to develop trading strategies that could yield risk-free profits (Shternshis, et al., 2022). LSTM can capture and learn these nonlinear connections between input data and desired outputs.

Hochreiter and Schmidhuber (1997) introduced LSTM networks that underwent subsequent refinements in the ensuing years by Gers et al. (2000) and Graves and Schmidhuber (2005). LSTM networks are explicitly designed to grasp long-term dependencies or sequential patterns, overcoming the inherent challenges of RNNs. It can retain information longer than RNNs (Fischer & Krauss, 2018). The structure of an LSTM network comprises an input layer, hidden layers, and an output layer. The incorporation of memory cells distinguishes these networks from others. These memory cells serve as crucial network components, facilitating the retention and utilization of contextual information over extended sequences. LSTM cell takes three different pieces of information: the current input sequence X_t, the short-term memory from the previous cell h_{t-1}, and the long-term memory from the previous cell state C_{t-1} at time t. The forget gate takes the information from X_t and h_{t-1} and produces the output between 0 and 1 through the sigmoid layer, and then it identifies which information to discard from the previous cell state C_{t-1}. When the value is 1, it stores all the information in the cell, while with a value of 0, it forgets all the information from the previous cell state. The calculations for each state and gate are as follows:

$$f_t = sigmoid\ (W_{f,x}x_t + W_{f,h}h_{t-1} + b_f\ ) \tag{3}$$

$$\hat{s}_t = tanh\ (W_{\hat{s},x}\, x_t + W_{\hat{s},h}h_{t-1} + b_{\hat{s}}) \tag{4}$$

$$i_t = sigmoid\ (W_{i,x}x_t + W_{i,h}h_{t-1} + b_i) \tag{5}$$

$$s_t = f_t \circ s_{t-1} + i_t \circ \hat{s}_t \tag{6}$$

$$o_t = sigmoid\ (W_{o,x}x_t + W_{o,h}h_{t-1} + b_o) \tag{7}$$

$$h_t = o_t \circ tanh\ (s_t) \tag{8}$$

Where, $W_{f,x}$, $W_{f,h}$, $W_{\hat{s},x}$, $W_{\hat{s},h}$, $W_{i,x}$, $W_{i,h}$, $W_{o,x}$, *and* $W_{o,h}$ denote weight matrices, b_f, $b_{\hat{s}}$, b_i, and b_o denote bias vectors, f_t, i_t, and o_t signify vectors representing the activation values of their respective gates, s_t and $\hat{s}_t$ represent vectors corresponding to the cell states and candidate values, h_t signifies a vector representing the output of the LSTM layer.

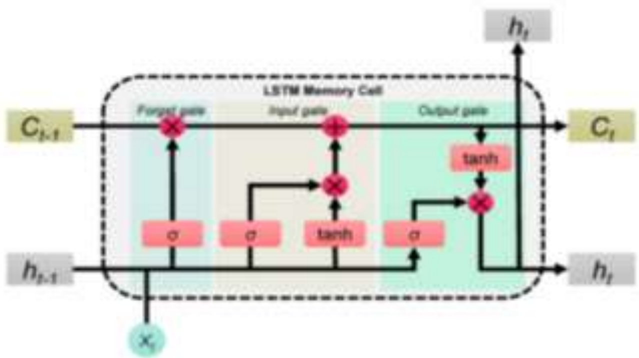


Figure 2.2: The architecture of an LSTM Unit
Source: Yehia (2024)

Chapter 3

Review of Literature

This section comprehensively reviews previous studies on the current research topic. The literature is analyzed systematically for each objective, enabling the identification of specific research gaps. The gaps are highlighted and discussed in this section, laying the foundation for the objectives of the present research.

3.1 Literature Review on Stock Market Prediction

Research in stock markets and machine learning includes forecasting stock prices, stock returns, stock volatility, market index values, and binary classification (direction) of stock movement. While most studies focus on predicting the next day's closing price or return for a stock or index, a few focus on intraday price prediction (Chong, Han, & Park, 2017) (Sun, Xiao, Liu, Zhou, & Xiong, 2019). Beyond stock market forecasting, the state-of-the-art machine learning models also apply to other stock market-related aspects, such as measuring the impact of factors influencing the stock market (Khattak, Ali, & Rizvi, 2021), determining the financial immunity of countries to the COVID-19 pandemic (Zaremba, Kizys, Tzouvanas, Aharon, & Demir, 2021), developing mobile application frameworks to alert potential investors on possible market close prices and develop trading strategies (Chandra & Chand, 2016).

3.1.1 Features for Stock Market Prediction

A taxonomy proposed by Bustos and Quimbaya (2020) classifies the predictive features into structured and unstructured data. While structured data comprises market information, technical indicators, and economic indicators, unstructured data comprises news, social networks, and blogs. The present study adopts the same pattern for analyzing the features for stock market prediction.

In the words of Fama (1965): "To what extent can the past history of a common stock's price be used to make meaningful predictions concerning the future price of the stock?" Historical stock market information in the form of open, close, high, and low prices and volume traded are most commonly used to predict the stock market, individually or in combination with other structured and unstructured data (Nazareth & Reddy, 2023). Long et al. (2020) used a fusion of stock market and trader's information to predict the next day's stock price direction.

The most frequently used technical indicators include Moving Averages, Relative Strength Index (RSI), Williams R, and Stochastic K (Ayala, Garcia-Torres, Noguera, Gomez-Vela, & Divina, 2021). Studies adopted by Chen & Hao (2017) and Basak, Kar, Saha, Khaidem, & Dey (2019) used a combination of market information and technical indicators to predict the Shanghai Stock Exchange (SSE) Composite Index and Shenzhen Stock Exchange (SZSE) Component Index by the former and to predict the direction of prices of selected stocks by the latter.

Khattak, Ali, and Rizvi (2021) consider macroeconomic data such as gold prices, oil prices, Bitcoin prices, EUR/USD exchange rate, and indices of stock markets across countries to find the predictors influencing the European market. Similarly, Kia, Haratizadeh, and Shouraki (2018) observed that a better market prediction model could be achieved by simultaneously using global market data of oil and gold prices along with historical indices of multiple countries to predict the target stock market and commodity prices' direction. Analysts use financial reports to extract liquidity, turnover, and Price/Earnings ratios (Zhang, Chu, & Shen, 2021) (Keyan, Jianan, & Dayong, 2021).

Text mining is gaining popularity in big data due to its ability to extract vast amounts of relevant information from news, blogs, speeches, and other social media websites that can enhance stock prediction. Similarly, analysts design market sentiment as an indicator to represent the opinions of a group of people in specific situations, as expressed on social media platforms concerning a particular stock or the market in general. It involves discovering the polarity of statements, i.e., investors' positive, negative, or neutral attitude towards the stock market, based on a news announcement on social media. For instance, analyzing Trump's tweet's immediate 15-minute impact on DJIA and the S&P500 index results in investors' negative reactions (Kinyua, Mutigwe, Cushing, & Poggi, 2021). Saurabh and Dey (2020) discovered the relationship between the social moods of people on the FTSE 100 index by employing ANN. They observe that the "happy" dimension could significantly improve the stock market's index return prediction. Heston and Sinha (2018) also found that positive news stories increase stock returns quickly, while negative stories receive a long-delayed reaction.

The stock markets are often volatile and change abruptly due to the country's economic conditions, political situation, and significant events. Therefore, investigating the effect of some major events, specifically global and local events for top stock companies, remains an open research area (Qiu, Song, & Akagi, 2016). Maqsood et al. (2020) analyzed the impact of significant local and global event sentiments through a Twitter dataset on selected U.S., Hong Kong, Turkey, and Pakistan stocks. Their findings illustrate that stock market volatility depends on community sentiments and the economic and political conditions of the country. Local events have had an impact on the performance of prediction algorithms more significantly than global events.

Accurately predicting the stock market calls for selecting relevant features that serve as inputs to machine learning models. While some researchers do not explain the selection of input features to the model but continue to utilize features from previous studies, few researchers have used machine learning models to select the most effective and relevant features that influence the stock market. Feature selection extracts a subset of input variables that efficiently encapsulate the data while minimizing the effects of noise or irrelevant factors and ensuring precise predictive results (Chandrashekar & Sahin, 2014).

Past literature showcases valuable features mined from the original dataset to boost predictor performance. The primary research focus in this field concentrates on using daily stock information (Open, high, low, close, volume), financial ratios, technical indicators, macroeconomic indicators, and financial news (Htun, Biehl, & Petkov, 2023) to predict stock returns. For instance, Tsai and Hsiao (2010) and Tsai et al. (2011) examined a combination of financial ratios and macroeconomic indicators to predict the Taiwan stock market. Cheng, Chen, and Lin (2010) extensively studied macroeconomic indicators and technical features, achieving 76% accuracy with probabilistic backpropagation, C4.5 Tree, and rough set algorithms. Similarly, Petrobras stock movement was predicted by combining technical and fundamental indicators, achieving 87.50% accuracy (de Oliveira, Nobre, & Zárate, 2013). Qiu et al. (2016) applied fuzzy curve analysis for the prediction of the Japanese Stock Market; Baek et al. (2020) utilized Decision Tree, Genetic Algorithm, and Random Forest to study the influence of economic indicators on the U.S. stock market volatility; Lohrmann and Luukka (2019) applied Random Forest for feature selection for predicting intraday stock returns direction. All along these lines, a substantial body of literature has verified and supported the importance of feature selection as an essential step in designing stock prediction models (Hussain, Knowles, Lisboa, & El-Deredy, 2008).

Financial time series often exhibit a nonlinear correlation pattern, which poses difficulties in being captured by Pearson's correlation coefficient, a metric extensively used in existing literature to measure linear correlations (He, Wen, Huang, & Ji, 2022). To capture both linear and nonlinear correlations between variables and to measure the mutual dependence between each input variable and the target variable, we harness the power of Mutual Information (MI) based on entropy for regression. In the theoretical work on the adequacy of MI for feature selection in regression, (Frénay, Doquire, & Verleysen, 2013) demonstrate that features selected with MI criteria are the ones that minimize the MSE and MAE.

3.1.2 Models for Stock Market Prediction

Machine learning models have an edge over traditional prediction models, which do not consider the nonlinearity in data. Classification models are used to obtain a binary output, i.e., to classify whether the stock would move upward or downward. In contrast, regression models forecast a particular value or trend of a stock/index. In addition, deep learning models are widely used in this study area. Some studies focus on comparing models, while a few have used ensemble models to enhance the accuracy of the output.

A study initiated by Atkins et al. (2018) used Latent Dirichlet Allocation for feature reduction and Naïve Bayes for text classification of financial news. Malagrino et al. (2018) also adopted the Bayesian Network to study the dependencies of stock market indices globally on the Ibovespa index. The authors point out this model's simplicity and human-friendly visuals in understanding conditional dependencies and the frequency of co-occurrences. Correspondingly, Khattak et al. (2021) also studied the influence of global market indices on the Eurostox50 index during the COVID-19 crisis using the Least Absolute Shrinkage and Selection Operator (LASSO) regression model. Selection and reporting of only essential variables and learning from the limited information available during a crisis are significant features of the model.

Basak et al. (2019) adopted Random Forest and Gradient Boosted Decision Tree (GBDT) for a trading period ranging from 3 to 90 days for stock price direction classification. The study's findings indicate that the models are apt for trading over more considerable periods, as the accuracy increased with the increase in the trading period. Ayala et al. (2021) compared the performance of Random Forest, linear model, ANN, and Support Vector Regression (SVR) for developing trading signals. ANN models outperformed others, while the Linear model was most suitable for short-period prediction. An experimental framework proposed by Qiu et al. (2016) attempted to improve ANN's accuracy, which was initially trained using the Back Propagation algorithm. Results prove that a hybrid model that combines the genetic algorithm and simulated annealing outperforms back propagation in boosting ANNs' weights and biases.

Chandra and Chand (2016) and Das et al. (2018) adopt a specific category of ANN, which includes RNN. Zhang et al. (2021) point out the vanishing gradient and exploding gradient problem in the RNN model and apply the LSTM model, an advanced version of RNN that could overcome these drawbacks. (Saud & Shakya, 2020) compared the performance of LSTM to RNN and GRU on the datasets of Nepal and Keyan et al. (2021) on the datasets of the Shanghai Stock Exchange. According to the former results, GRU outperforms LSTM and RNN, while the latter favors LSTM over GRU and RNN. However, since these models were trained on different datasets and timeframes, it incapacitates us from weighing up any model. Li, Bu, Li, & Wu (2020) concluded through their study that utilizing investor sentiment as an indicator of the machine learning models proved the superiority of LSTM over logistic regression, Naïve Bayes, and SVM. Further, a study proposed by M.A. Menon (2018) applied four deep learning models to predict the next 10 days' closing prices of certain stocks listed on the NSE and NYSE. The results demonstrated that CNN outperforms LSTM, MLP, and RNN in capturing abrupt changes in the data. However, the experimental findings reveal that the LSTM model with 30 neurons achieves the best fit and highest prediction accuracy, followed by GRU with 50 neurons and CNN with 30 neurons (Pokhrel, et al., 2022). A comprehensive study put forth by Kraus & Feuerriegel (2017) compared the performances of Naïve baseline and traditional machine learning models: Ridge regression, LASSO, Elastic Net, Random Forest, SVM, AdaBoost, Gradient Boosting, Deep learning models: RNN and LSTM, and transfer learning: RNN and LSTM. Transfer learning using RNN and LSTM showed significant results. Nabipour et al. (2020) compare machine learning models: Decision Tree, Random Forest, Adaptive Boosting, XGBoost, Support Vector Classifier (SVC), Naïve Bayes, KNN, Logistic Regression and ANN, and deep learning methods: RNN and LSTM. The results indicate that RNN and LSTM outperform all other prediction models on continuous and binary data. In recent years, the application of LSTM for stock market prediction is due to its better efficiency in learning how public mood impacts financial time series (Malandri, Xing, Orsenigo, Vercellis, & Cambria, 2018). Sentiment analysis has proved crucial in predicting stock markets to determine whether the market is driven by rational decision-making or by investors' emotions and opinions. Xing et al. (2019) developed a Sentiment-Aware Volatility forecasting model, incorporating market sentiments for predicting fluctuations in stock markets. This novel model outperforms GARCH, EGARCH, TARCH, GJR-GARCH, Gaussian-process volatility model, and Variational neural models, including VRNN and LSTM. Ample studies focus on using market sentiment indicators to predict the future price of stock markets, while there is still considerable scope for predicting volatility driven by market sentiments.

In predicting the stock's direction, the average reported accuracy lies between 55% and 65% using various models such as Bayesian networks, random forest, and XGBoost. Weng et al. (2017) conducted a comparison analysis wherein a situation combining market information, technical indicators, Google news counts, and generated features proved to have the highest accuracy of around 85% using decision trees, SVM, and Neural networks. Li et al. (2020) brought out the differences among models and provided evidence of the superiority of the LSTM model to SVM, Naïve Bayes, with an accuracy of 80.20% when incorporating the hourly text-extracted investor sentiment indicator. The ability of the LSTM model to remember both long-term and short-term values has proved its superiority in treating financial time series, thereby becoming the preferred tool for time series analysis. In addition, they can handle very noisy data and work independently from the linearity assumption.

3.2 Literature Review on Portfolio Optimization

Portfolio management includes several tasks, such as selecting assets to form the portfolio, prioritizing assets based on risk tolerance and expected returns, and formulating suitable portfolio strategies. It also includes revising these strategies and rebalancing portfolio composition from time to time to achieve long-term or short-term financial goals. Due to these many related tasks, portfolio selection, portfolio allocation, and portfolio optimization have been used interchangeably in previous literature but with the same underlining aspect of managing a portfolio (Ozbayoglu, Gudelek, & Sezer, 2020).

Selecting a set of potential securities to construct a portfolio from the vast pool of stocks available on the exchange is a formidable challenge. Furthermore, determining the optimal weights to maximize returns while minimizing risks adds to the complexity of this task. Investors must make certain trade-offs when choosing securities due to the limitation of capital. On one hand, investors cannot solely invest in the highest-return securities, as their risk tolerance is insufficient to handle the associated risk. On the other hand, it is also impossible for investors to invest all their capital in risk-free securities because the available returns are too low to satisfy investors' demands (Deng & Yuan, 2021). Hence, the selection of securities and capital allocation are critical problems for investors. It was not until

Markowitz (1952) proposed the Modern Portfolio Theory (MPT).

Way back in 1952, Markowitz (1952) designed the Mean-Variance (MV) portfolio, wherein the average of the historical data of a stock's return is the expected return, and the variance of these returns acts as the risk (Milhomem & Dantas, 2020). The majority of the studies to date utilize the MV portfolio strategy. Recent advancements in technology have made online portfolio selection possible. Here, decisions are made online as and when financial data is updated. Reinforcement learning is gaining popularity in investment portfolios nowadays due to its ability to make decisions by observing the state of the environment.
Portfolio selection theory focuses on searching for the most appropriate categories of assets to hold under uncertainty, according to the characteristics of assets. By analyzing these characteristics, investors can determine the most suitable asset categories to include in their portfolios (Monsour, Cherif, & Abdelfattah, 2019). Portfolios must be constructed so that their assets can outperform the market's cross-sectional median or benchmark index.

3.2.1 Models for Portfolio Optimization

This section reviews various models used for portfolio management and its closely related functions, such as portfolio construction, asset selection, fund allocation, and portfolio optimization.

Several strategies have been implemented in literature for creating portfolios. The most common one followed includes predicting stock prices/returns, and based on the predictive performance, a decision is made to include the stock in a portfolio or not. For example, Chen, Zhang, Mehlawat & Jia (2021) employ XGBoost for stock price prediction combining it with the Mean-Variance (MV) model to construct a portfolio; Ma, Han, & Wang (2021) applies Random Forest, SVR, CNN, MLP, LSTM and Linear Autoregressive Integrated Moving Average (ARIMA) incorporated with MV and Omega Portfolio. Wang, Li, Zhang & Liu (2020) apply LSTM and SVM for prediction and optimize portfolios through MV and 1/N. According to their comparative study, LSTM+MV outperformed LSTM+1/N, SVM+MV, and SVM+1/N in portfolio optimization by achieving higher annual cumulative returns, Sharpe ratio per triennium, and average returns to risk. Paiva et al. (2019) use SVM as a binary classifier for day trading operations intending to gain a specific target daily return of 1%, 1.5%, and 2%. The proposed model SVM + MV outperforms SVM + 1/N and Random Forest + 1/N.

Vo et al. (2019) propose a reinforcement learning model incorporating MV and

Environmental Social Governance ratings (MV-ESG) to create a socially responsible investment portfolio. These studies adopt a two-step approach: forecasting future returns and taking appropriate trading decision rules. However, Koratamaddi et al. (2021) apply deep reinforcement learning techniques to create an intelligent automated trader that would merge the two steps and act as an investor in the real world. The model adopted was called Adaptive Sentiment-Aware Deep Deterministic Policy Gradients (DDPG). Thus, reinforcement learning in portfolio management has eliminated tedious activities that investors would otherwise have to perform.

Tan et al. (2019) use Random Forest to directly select stocks from the Chinese stock market in the short and long run. Hargreaves et al. (2013) use neural networks and Logistic Regression to determine the optimum stock portfolio selection in the Healthcare Sector and Financial Sector on the Australian Stock Exchange. Gunasekaran et al. (2015) developed a system of portfolio optimization using the Neuro-Fuzzy System on the BSE Sensex stock index.

While making decisions regarding stock selection, investors usually aim to select stocks when both the risks and the returns are consistent, such that the risk is low and the return is high (Chen & Huang, 2009). In the literature of modern portfolio theory, the initial stock selection approach developed for investment focuses on clustering stocks based on risk tolerance and return expectations (Markowitz, 1952), (Kılıc¸man & Sivalingam, 2010) (Mirnoor & Shariati, 2012). Risk is measured using variance when employing this approach, while asset return data are assumed to be normally distributed. However, this is misleading, as the risk measures obtained are inconsistent with investors' preferences, and the assumption made on the asset return data is invalid for a real market scenario, which is actually left or right-skewed (Rom & Ferguson, 1994). Thus, methods based on cluster analysis are presented as alternative stock selection approaches that aim at grouping multiple stocks simultaneously, such that stocks belonging to one cluster are similar, and those belonging to different clusters are dissimilar (Nanda, Mahanty, & Tiwari, 2010). Mirkin (1996) defines clustering as "a mathematical technique designed for revealing classification structures in the data collected in the real-world phenomena." It aids in visually positioning assets by revealing underlying similarities. Researchers have used clustering techniques for different purposes. For example, Jung & Chang (2016) clustered stocks using partial correlation coefficients to study the relationship between firms over different periods; Chen et al. (2014) analyzed the spatial structure of stock interactions; and D'Urso, Cappelli, Lallo, & Massari (2013) classified financial time series of Euro exchange rates against international currencies using cluster analysis. This study aims to cluster stocks of similar characteristics, creating portfolios for better risk management, diversification, and optimized investment strategies. Clustering can simplify markets by reducing dimensionality and complexity, facilitating portfolio optimization.

Chen and Huang (2009) evaluated the performance of 122 Equity Mutual Funds from 47 mutual trust companies in Taiwan using k-means clustering. They classified these funds into four clusters: Inferior performing funds, stable funds, aggressive funds, and performance funds, based on four evaluation indices: rates of return, standard deviation, turnover rate, and Treynor index. They eliminated the two clusters that yielded lesser profits and selected the portfolio from the remaining two. The authors optimized the portfolio in two ways: maximizing the expected return subject to future risk and minimizing the future risk subject to expected return. Similarly, Kılıc¸man & Sivalingam (2010) analyzed 38 equity mutual funds of three banks in Malaysia using the same method without considering the turnover ratio. Similarly, Mirnoor and Shariat (2012) also applied K-means clustering to optimize the portfolio selection of equity mutual funds in Iran. However, only three clusters were considered by excluding the cluster of stable funds. Moreover, the Sortino ratio was used instead of the turnover ratio.

Apart from k-means clustering, researchers compared the performance of various clustering techniques. For example, B.S. & Mathew (2016) evaluated the performance of the Partitioning-based technique (k-means), Hierarchical technique (Agglomerative algorithm), Density-based technique (DBSCAN), and Model-based technique (EM) with the help of validation indices such as C-index, Jaccard index, Rand index, and Silhouette index. The K-means algorithm in the partitioning-based technique and the EM algorithm in the model-based technique showed better performance than hierarchical and density-based techniques. In the context of the Indian stock market, Nanda et al. (2010) compared the clustering performance of k-means, Self-Organizing Maps, and Fuzzy C-means for 106 stocks from BSE. The main aim of clustering these stocks was to find the least diversity within the group and the most difference among the clusters or groups formed. Various validity indices such as the Silhouette index, Davies-Bouldin index, Xie and Beni's index, Partition index, and Separation index were used to find the optimal number of clusters. Stocks were then selected from each cluster to build a portfolio. The portfolio's average weekly returns were graphically analyzed for 6 months. The findings reveal that the three constructed portfolios beat the benchmark index (BSE Sensex), with K-means clustering outperforming SOM and Fuzzy C-means.

Generally, high-quality stocks are selected based on their predicted return to form a portfolio. Researchers predict these returns using fundamental as well as technical indicators. Most studies use fundamental data, such as historical open, high, low, close price, and volume traded data, to form their portfolios (Nazareth & Reddy, 2023). Researchers use K-means clustering as an unsupervised machine learning model that does not rely on labeled data for training but instead identifies patterns or groupings within the data autonomously. However, several factors are fed into the model to identify the underlying structures. Almost all studies that use clustering for portfolio management have repeatedly made use of returns and standard deviation for identifying stock classification patterns ((Chen & Huang, 2009), (Kılıc¸man & Sivalingam, 2010), (Mirnoori & Shariati, 2012), (Alqaryouti, Farouk, & Siyam, 2018)). Other factors used along with the returns and standard deviation are Treynor's Index, Sortino ratio, and asset turnover. Nanda et al. (2010) used a more structured set of factors for clustering BSE Sensex stocks. The users divide these factors into returns and validation ratios. The returns fed to the model include 1 day, 1 week, and 30 days covering short-term stock performance and 3 months, 6 months, and 1 year covering the long-term performance of the stocks. The validation ratios include: Price Earning (P/E), Price to Book Value (P/BV), Price/Cash EPS, Enterprise Value/EBIDTA, and Market Capitalization/Sales. Fallahpour et al. (2014) applied the same factors for portfolio management on TSE50, Iran, by eliminating Price/Cash EPS, Enterprise Value/EBIDTA, and 1-year returns. While reviewing the studies that apply clustering for portfolio management, almost all studies used data from less than a year to identify classification structures. At the same time, the 2-step model that initially focuses on stock price forecasting requires a large quantity of data to train the forecasting model.

The Sharpe ratio is the most commonly used metric to measure investments. It represents an amount of return (in units) obtained by taking a unit of risk. The greater a portfolio's Sharpe ratio, the better its risk-adjusted performance. Juan (2022) reports a Sharpe Ratio of 9.31, using an Attention-based LSTM network to form a portfolio of 20 stocks from CSI 300. Meanwhile, using the same model for portfolio formation on the S&P500, a Sharpe Ratio of 2.77 is reported. This model, combined with statistical arbitrage, can generate positive trading performance based on the daily and weekly returns of S&P 500 stocks. However, in the real world, investors are concerned with the returns they achieve and do not wholly rely on the Sharpe ratio.

In the context of fund allocation, previous studies have emphasized the creation of an EWP, MSRP, and/or an MVP for a set of stocks (Meher & Mishra, 2024), (Li & Zhong, 2022), (Prasad, Jaiswal, Shakya, & Singh, 2020). In an EWP, each stock is allocated an equal percentage of the total investment. In an MSRP, the aim is to optimize the risk-adjusted return by maximizing the Sharpe ratio. Thus, weights are assigned such that the portfolio with the highest Sharpe ratio is selected. The goal of MVP is to minimize the portfolio's overall volatility (risk) while still considering all the stocks. Thus, weights that result in the lowest overall volatility amongst all portfolios are selected. The allocation of funds to each stock is determined using Monte Carlo simulations. Monte Carlo Simulation is a mathematical technique that estimates the possible outcomes of an uncertain event. This method involves generating numerous random portfolio weight combinations, followed by selecting the portfolio weights that produce the highest Sharpe ratio (in the case of MSRP) or selecting the portfolio weights that produce the lowest overall volatility (in the case of MVP).

Metropolis and Ulam proposed the Monte Carlo simulation while developing atomic bombs to solve the random neuron diffusion problem. With the continuous advancement of computers, Monte Carlo simulations have been widely used in economic research for option pricing and financial decision-making (Tobisova, Senova, & Rozenberg, 2022). They are typically used for sampling, estimation, and optimization. Specifically, they can be used to price financial instruments and analyze risk (Kroese, Brereton, Taimre, & Botev, 2014). Platon & Constantinescu (2014) selected an investment project based on the risks estimated by Monte Carlo methods.

It is difficult for investors to hold too many stocks and measure their performance regularly. Hence, Paiva et al. (2019) recommend that an average of 7 assets in a portfolio be held per day. Based on this precept, Chen et al. (2021) created a portfolio consisting of stocks with a cardinality ranging from 6 to 10, while Wang et al. (2020) used 4 to 10 stocks. Most studies under review consider using ten or fewer stocks, as an individual investor can easily manage them. Tan et al. (2019) selected 20 stocks to form the portfolio to be held for 20 days, and Vo et al. (2019) empirically tested that the ESG portfolio best performed when it consisted of 7, 18, and 12 particular stocks in 2016, 2017, and 2018 respectively. However, there is no hard and fast rule on the number of stocks an investor must hold.

3.3 Research Gap

While abundant literature focuses on stock market predictions, most studies emphasize developed economies, with limited research addressing the dynamics of developing markets. Furthermore, developed and developing markets often have distinct characteristics that can alter the model's predictive performance. The former tends to be more mature, efficient, and stable, while the latter can be more volatile and vulnerable to unique factors, as evidenced in the stock market time series. Gao, Zhang, and Yang (2020) provide evidence that predictive models exhibit superior performance in developed financial markets compared to developing ones. Given that India is recognized as one of the fastest-growing economies on the global stage, numerous factors differentiate its financial time series from those of already developed economies. This study bridges the gap by applying deep learning models to forecast India's prominent benchmark indices. These indices reflect the broader market sentiments of the Indian economy. Moreover, there is a notable gap in studies on predicting sectoral indices using machine learning models. This study attempts to predict 13 sectors of the Indian stock market, followed by the prediction of closing prices of stocks listed on the NSE.

Furthermore, previous studies predominantly focus on exploring regular market conditions, leaving a gap in examining model performance during periods of significant volatility, mainly triggered by unscheduled events. This empirical study focuses on assessing the predictive abilities of deep learning models during critical economic and geopolitical shifts that have the potential to cause volatility in the Indian market. These include scheduled national events such as the national general elections of 2019 and 2024 and the implementation of the GST regime. The unforeseen events under study include a domestic event: demonetization and the global COVID-19 pandemic, the invasion of Ukraine by Russia, and the Hamas attack on Israel.

While researchers have studied numerous portfolio optimization methods, most research primarily emphasizes stock price prediction as the basis for portfolio construction. In this approach, authors use the predicted stock prices as the basis for stock selection. However, clustering algorithms, which use historical data to identify inherent stock classification structures, offer a more reliable alternative than predictive models.

In the context of clustering, previous studies have demonstrated the superior performance of K-means clustering compared to other techniques in identifying underlying stock classification structures. Diversification of risk, often measured through the correlation of returns, is a critical aspect of portfolio management. Existing studies typically select stocks from different clusters to achieve diversification, operating on the premise that selecting stocks from separate clusters is likely to exhibit a lower correlation, thereby enhancing portfolio diversification. As an unsupervised machine learning model, the K-means clustering algorithm organizes stocks based on its internal logic and patterns, which are not always transparent. Given its black-box nature, we analyze the correlation patterns of stocks within each cluster.

Although researchers have explored this approach to some extent, there remains a significant geographical gap in the literature regarding the application of clustering techniques to the Indian stock market. The limited studies available, such as those by Nanda et al. (2010) for 2007–2008 and B.S. and Mathew (2016) for the first half of 2015, provide insufficient coverage. Furthermore, there is a lack of studies evaluating the performance of portfolios constructed through clustering during periods of market crisis or events that induce significant volatility. To address this gap, our research evaluates the performance of portfolios created using clustering techniques across various market events, including scheduled and unforeseen national and global events. By focusing on the Indian stock market and incorporating periods of heightened market turbulence, this study aims to provide novel insights into the robustness and effectiveness of clustering-based portfolio optimization.

In the context of fund allocation, several studies have employed Monte Carlo simulations on either randomly selected stocks or stocks within a specific sector. However, only a limited number of studies explicitly test the performance of the allocated weights. To bridge this gap, Monte Carlo simulations are applied to determine the optimal allocation of funds after stock clustering. Additionally, we construct an Equally Weighted Portfolio, Maximum Sharpe Ratio portfolio, and Minimum Volatility portfolio. We evaluate the performance of all these portfolios by creating a dummy investment strategy.

Chapter 4

Research Methodology

This section presents a detailed explanation of the research methodology employed to achieve each proposed objective.

4.1 Research Methodology for Objective I

4.1.1 Period of the Study

The study covers over twelve years, from 01^{st} January 2012 to 31^{st} October 2024. During this period, several events impacted the Indian economy from a geopolitical, financial, economic, technological, and healthcare viewpoint. For instance, tax reforms, governance changes, terror attacks, and law enactments are notable examples. There have also been disruptions in global trade due to incidents such as the pandemic and military conflicts, which caused supply chain breakdowns worldwide. The selected study period covers events that have caused volatility in the Indian stock market. This study investigates the predictive performance of machine learning models during seven such events. The scheduled events under study include the implementation of the GST regime in 2017 and the national general elections of 2019 and 2024. Unforeseen events include demonetization in 2016, the global COVID-19 pandemic in 2020, the invasion of Ukraine by Russia in 2022, and the Hamas attack on Israel in 2023.

4.1.2 Sample Design

A model was developed to predict the Nifty50, Nifty100, Nifty200, and Nifty500 broad market indices as they represent 54%, 66.4%, 79.6%, and 92% of the free float market capitalization of the stocks listed on NSE as of 30^{th} September 2024, respectively.

A model was developed to predict the sectoral indices, including Nifty Auto Index, Nifty Bank Index, Nifty Energy Index, Nifty Financial Services Index, Nifty FMCG Index, Nifty IT Index, Nifty Infrastructure Index, Nifty Media Index, Nifty Metal Index, Nifty Pharma Index, Nifty PSU Bank Index, Nifty Realty Index, and Nifty Services Index.

The stocks under study include HCL Technologies Ltd., DLF Ltd., ITC Ltd, Mahindra & Mahindra Ltd, State Bank of India, Zydus Lifesciences Ltd., PVR Inox Ltd., Hindalco Industries Ltd., Reliance Industries Ltd., and Siemens Ltd.

4.1.3 Data Variables and Sources

We categorized the input features into stock price data and macroeconomic variables. The stock market data encompasses the daily open, high, and low prices of the selected indices/stocks. On the other hand, the macroeconomic variables consist of the daily U.S. Dollar – Rupee exchange rate representing foreign exchange, monthly broad money (M3) representing money supply, monthly gold price representing commodities, and quarterly GDP at market price, which indicates growth and economic activity. However, three variables often used in the literature, the Consumer Price Index representing inflation, the 91-day treasury bill representing interest rates, and the unemployment rate, were not part of this study due to the unavailability of data since 2012. The closing price of the index/stock is the target output.

Table 4.1: Overview of features, frequency, and sources

Features		**Frequency**	**Source**
Stock market:	Open Price	Daily	NSE
	High Price	Daily	NSE
	Low Price	Daily	NSE
Macroeconomic Variables:	U.S. Dollar – Rupee Exchange rate	Daily	RBI
	Broad Money	Monthly	RBI
	Gold price	Monthly	RBI
	GDP	Quarterly	RBI

We source the daily stock price data for the four broad market indices, sectoral indices, and selected stocks from the official NSE website. The U.S. Dollar-Rupee exchange rate, M3, gold price, and GDP are sourced from the official website of the RBI. Table 4.1 provides the list of potential features, their frequencies, and sources.

4.1.4 Methodology

The research design framework for stock market prediction is depicted in Figure 4.1. Data preparation and handling were executed in Python 3.8, supported by numpy, pandas, and scikit-learn packages. The models were built and trained using Keras to initialize a sequential model object. One of the reasons for preprocessing the data before training the model is to identify missing data. A forward-fill technique was applied to bridge the gap between low-frequency (monthly and quarterly) data and high-frequency (daily) data and to preserve the sequential relationship.

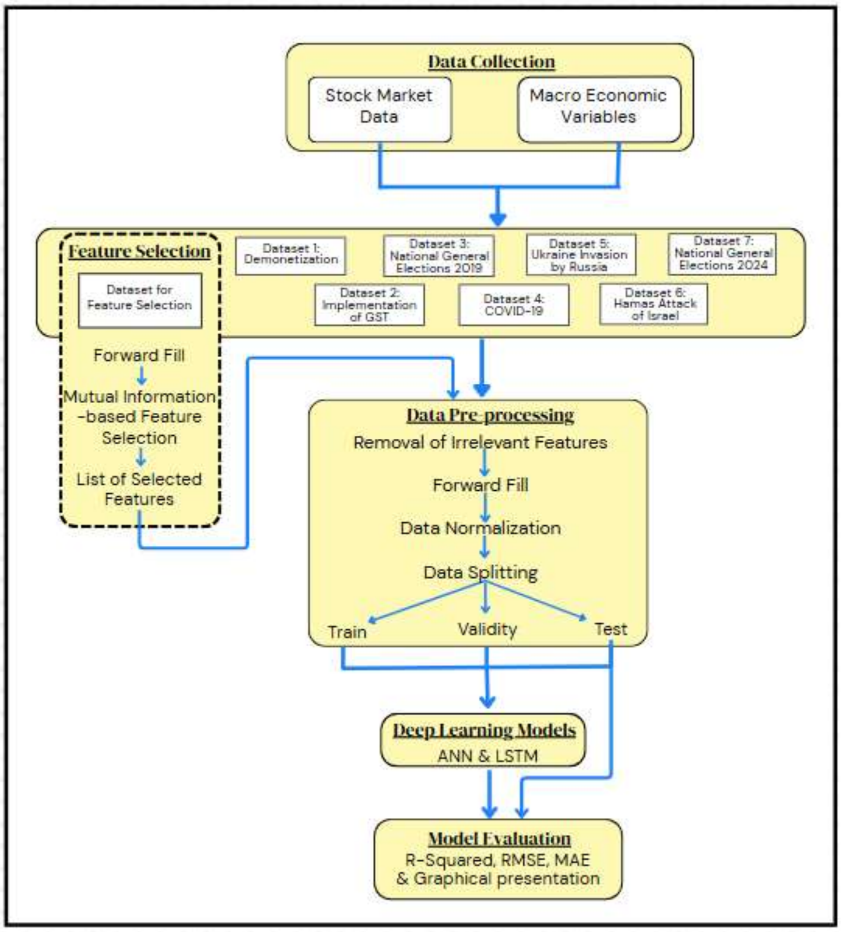


Figure 4.1: Research design framework for stock market prediction

To identify the most significant variables contributing to the prediction of the target variable, the mutual dependence between each input variable and the target variable was measured from 01st January 2012 to 30th September 2012. The 75th percentile of input variables with the highest MI scores was selected as input to the deep learning models.

After selecting the most informative features, each dataset was split into 80% training and 20% testing. From the training dataset, 20% was used for validation. The training data was normalized using MinMaxScaler from sklearn.preprocessing. This technique scales each feature in the range of [0, 1] by using:

$$xnorm = (x - xmin) / (xmax - xmin) \tag{9}$$

As the order of the data is critical in time-series prediction, it was divided into consecutive windows of validation. For illustration, we employed a sliding window approach in which the initial five observations (from the first to the fifth) were utilized to predict the subsequent observation. In the subsequent sliding window, we employed observations from the second to the sixth to forecast the following consecutive observations, and this pattern continues iteratively.

For each event, the prediction began from the first trading day of the month prior to the month in which the event occurred. The test sets were meticulously formed by considering at least a month's (average 21 trading days) time stamps preceding the event day to ensure our forecasting analysis accounted for the dynamics leading to these critical events.

The training set was selected 4 years prior to the start of the test set, proportionally being in the ratio of 4 years:1 year (train: test). The event dates, the train, and the test datasets are detailed in Table 4.2.

Table 4.2: Train and test period for each event

Dataset	Event Date	Training Period	Testing Period
Development Set (Feature Selection)	XX	01st January 2012 – 30th September 2016	
Demonetization	08th November 2016	1st October 2012 – 30st September 2016	1st October 2016 – 30st September 2017
GST Implementation	01st July 2017	1st June 2013 – 31st May 2017	1st June 2017 – 31st May 2018
General Elections 2019	23rd May 2019	1st April 2015 – 31st March 2019	1st April 2019 – 31st March 2020
Covid-19 Pandemic	11th March 2020	1st February 2016 – 31st January 2020	1st February 2020 – 31st January 2021
Invasion of Ukraine by Russia	24th February 2022	1st January 2018 – 31st December 2021	1st January 2022 – 31st December 2022
Hamas attack on Israel	7th October 2023	1st September 2019 – 31st August 2023	1st September 2023 – 31st August 2024
General Elections 2024	4th June 2024	1st May 2022 –30th April 2024 (2 years)	1st May 2024 – 31st October 2024 (6 months)

The number of layers, dropout ratio, epochs, time-stamps, and neurons are tuned for model configuration. These elements significantly influence the model's predictive performance, and their optimization is integral to our research. Both ANN and LSTM contain an input layer, single/multiple hidden layers, a dropout layer, and a dense output layer.

Neurons: The number of neurons in the input layer equals the number of input parameters in the regression problem. This experiment addresses the count of neurons that provide a better prediction accuracy and a superior fit to each model. Another fully connected layer is added with nodes ranging from 10 to 200 and a Tangent Hyperbolic activation function.

Dropout ratio: A dropout regularization technique recommended by Gal and Ghahramani (2016) is applied within the recurrent layer to avoid overfitting. The dropout ratio ranges from 0.05 to 0.50. This approach randomly drops a portion of the input units to 0 during each training update, thereby reducing the risk of overfitting and facilitating better generalization. Then, the final fully connected layer is added with a single neuron and a linear activation function. This layer produces the prediction output for the regression task.

The time-stamp or look-back values are crucial when working with sequential data, where the order and timing of observations matter. The choice of the look-back value is essential as it affects the model's ability to capture patterns and dependencies in the data. A more considerable look-back value allows the model to consider more historical information, potentially capturing long-term trends. However, it may also introduce noise and increase computational complexity. On the other hand, a smaller look-back value might need to capture more historical information, which could lead to oversimplified predictions. The model is tuned with several look-back values: 5, 10, 15, 21, 42, and 63-day data. 5, 10, and 15 days represent the average trading days in 1-week, 2-week, and 3-week, respectively. 21, 42, and 63 days represent the average trading days in 1-month, 2-months, and 3-months, respectively. This experimental framework addresses the look-back period required to predict the next day's close price more accurately.

(It is imperative to note that within each LSTM layer, the neurons return sequences instead of a single output. The input sequences have time steps, each with the number of input features. As the input sequence is processed, its features are presented to the LSTM network. The input sequence is processed one timestep at a time, producing an output after the final timestep).

Layers: In forecasting applications, it is common to apply feedforward Neural Networks with a single hidden layer and a solitary output node (Sharma, Hota, Brown, & Handa, 2022). We investigate this assumption by accommodating multi-hidden layers (ranging from 1 to 8) in the network, keeping the same number of nodes in each hidden layer.
Epochs and batch size: The number of epochs determines how often the model will see the entire dataset. Too few epochs may lead to underfitting, as the model has not seen enough data to learn complex patterns. On the other hand, too many epochs can lead to overfitting, where the model starts memorizing the training data instead of learning the underlying patterns. The model is thus trained on different number of epochs: 10, 20, 30, 40, 50, 60, 70, and 80 epochs, with a batch size of 16, such that the weights are updated after processing each batch consisting of 16 input sequences.

During training, the model learns to minimize the Mean-Squared Error (MSE) loss between the predicted outputs and the actual target values. The goal is to optimize the model's parameters (weights and biases) to make accurate predictions on unseen data. During training, the validation set is utilized to evaluate the value of the loss function. The selected loss function, MSE, is calculated at the end of each epoch.

Both models, ANN and LSTM, are compiled using the Adam optimizer, an efficient optimization algorithm commonly used for deep learning.

Finally, a dense layer with a single output unit performs a linear transformation to generate the regression prediction.

We have experimented with different architectures and parameters to obtain optimal configurations. The performance of this optimal network is then used for comparison. We highlight the convention that we do not stop at the modeling phase; instead, it is a continuous procedure of building and testing different parameters that produce the model with the highest accuracy. The complete set of fine-tuned alternatives is detailed in Table 4.3.

The performance of the models is evaluated and compared using R-squared (R^2), the Root Mean Squared Error (RMSE), and the Mean Absolute Error (MAE) metrics.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{11}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{12}$$

Where;

N is the number of observations,

y_i is the actual value of the i^{th} observation and

$\hat{y}_i$ is the predicted value of the i^{th} observation

$\bar{y}_i$ is the mean of y

Table 4.3: Hyper-parameters and their candidate values

Hyper-Parameter	**Candidate Values**
Layers	1, 2, 3, 4, 5, 6, 7, and 8
Neurons	10, 20, 40, 60, 80,100, 120, 140, 160, 180 and 200
Dropout Ratio	0.05, 0.1, 0.2, 0.3, 0.4, and 0.5
Epochs	10, 20, 30, 40, 50, 60, 70, and 80
Time Stamp	5, 10, 15, 21, 42 and 63 days
Learning Rate	0.01
Batch Size	16
Optimization Algorithm	Adam Optimizer
Loss Function	Mean Squared Error

4.2 Research Methodology for Objective II

4.2.1 Period of the Study

We construct portfolios, allocate funds, and evaluate the framework's performance by simulating a dummy investment held for one year. This one-year period serves as the test period, consistent with the test period defined in Objective 1. The variables used for portfolio construction are derived from data collected during the year preceding the test period. Table 4.4 presents a detailed breakdown of the variable extraction period for each event and the corresponding investment holding period or test period.

4.2.2 Sample Design

Portfolios are created based on stocks constituted in the Nifty 100 index. The list of stocks in the Nifty100 index for each of the years when the events occurred is provided in the Annexure.

Table 4.4: Data sourcing and testing windows

Event	Event date	Variables sourced for the period	Investment holding Period
Demonetization	08th November 2016	1st October 2015 – 30th September 2016	1st October 2016 – 30st September 2017
GST implementation	01st July 2017	1st June 2016 – 31st May 2017	1st June 2017 – 31st May 2018
General Elections 2019	23rd May 2019	1st April 2018 – 31st March 2019	1st April 2019 – 31st March 2020
Covid-19 pandemic	11th March 2020	1st February 2019 – 31st January 2020	1st February 2020 – 31st January 2021
Invasion of Ukraine by Russia	24th February 2022	1st January 2021 – 31st December 2021	1st January 2022 – 31st December 2022
Hamas attack on Israel	7th October 2023	1st September 2022 – 31st August 2023	1st September 2023 – 31st August 2024
General Elections 2024	4th June 2024	1st May 2023 – 30th April 2024	1st May 2024 – 31st October 2024 (6 months)

4.2.3 Data Variables and Sources

A set of representative financial indicators, including the stock's returns, standard deviation, and validation ratios, are selected to examine the stock performance. We calculated the daily, weekly, monthly, quarterly, and yearly returns, as well as the standard deviation of all stocks. The validation ratios include Price-to-Earnings (P/E), Price-to-Book Value (P/BV), Price-to-Cash Earnings per share (CEPS), Enterprise Value-to-Earnings before Interest and Tax (EV/EBIT), and Market Capitalization-to-Sales. A high P/E ratio implies that the market is more willing to pay for the company's earnings as the market has high hopes for the future, and as a result, it bids up the price. The P/BV ratio compares the market value of a company's share to its book value. A lower P/BV ratio could be a sign that the stock is undervalued. The Price/Cash Earnings per share (CEPS) indicates how much operating cash a company generates relative to its stock price. EV/EBIT gives investors an idea of how much they pay for the company's operating profit, regardless of its capital structure or tax situation. The Market Capitalization-to-Sales ratio indicates how the market values every rupee of the company's sales. Simply put, it shows a company's value in relation to its sales. The variables for stock classification are listed in Table 4.5.

The stock closing prices and validation ratios were extracted from the CMIE database. Data that was unavailable for specific companies were excluded from this study.

Table 4.5: Stock classification variables

	Variables
Returns	Daily
	Weekly
	Monthly
	Quarterly
	Yearly
Standard Deviation	Daily
Validation Ratios	Price-to-Earnings (P/E)
	Price-to-Book Value (P/BV)
	Price-to-Cash Earnings Per Share (P/CEPS)
	Enterprise Value-to-Earnings Before Interest and Taxes (EV/EBIT)
	Market Capitalization-to-Sales

4.2.4 Methodology

The research design framework adopted for portfolio optimization is depicted in Figure 4.2. The entire process is summed up in three main steps: Clustering, allocation of funds, and investment and evaluation.

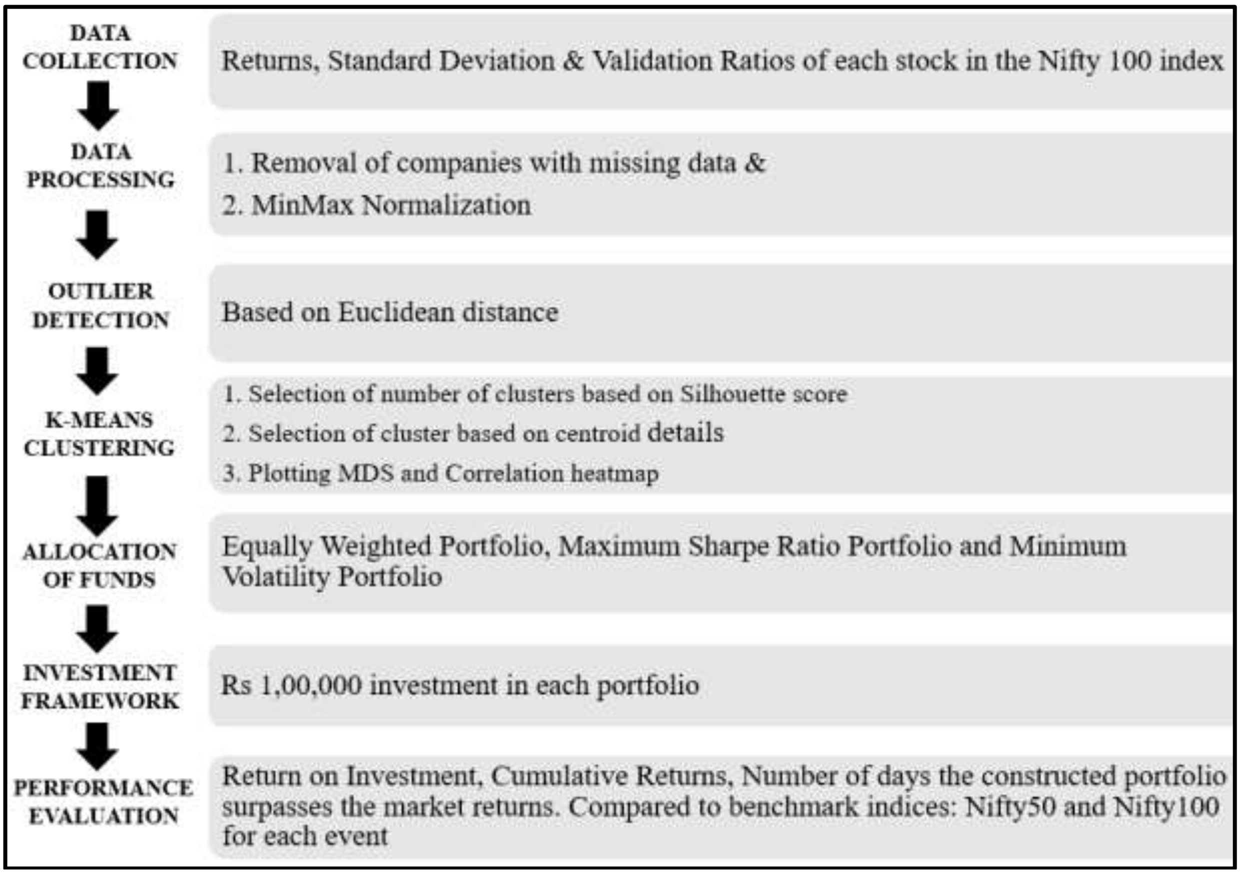


Figure 4.2: Research design framework for portfolio optimization

4.2.4.1 Clustering

Since the measurement units of each variable are different from one another, the data was first normalized. This ensures that the clustering results are on the same basis and, thus, more accurate. We use the Min-Max normalization technique to scale the features to the range of [0, 1]. Before clustering the data, the outliers are detected and eliminated. The remaining rows of each company's return, standard deviation, and validation ratios are sent to the clustering model.

The k-means clustering model is optimized using K-Means++ with 20 re-runs limited to 100 steps. Here, the number of clusters is decided based on the highest Silhouette score. The stocks are plotted on a Multi-Dimensional Scale (MDS). MDS is a dimensionality reduction technique that projects high-dimensional data onto a lower-dimensional space while preserving the pairwise distances between the data points as much as possible.

Next, to select the appropriate cluster, we determine the centroid characteristics of each cluster formed. We create a correlation heatmap to understand the nature of correlation amongst the stocks in the selected cluster.

4.2.4.2 Fund Allocation

To evaluate the effectiveness of the selected clusters formed by the K-Means clustering model, we implement three distinct fund allocation strategies: EWP, MSRP, and MVP. In an EWP, we allocate an equal percentage of the total investment to each selected stock. The remaining two portfolios are constructed using the Monte Carlo, which simulates 1,00,000 random portfolios with different weight combinations. The returns, volatility, and Sharpe ratio are calculated for each simulated portfolio. The combination of weights that produce the highest Sharpe ratio is selected to create the MSRP. The combination of weights that result in the portfolio's lowest overall volatility is selected to create the MVP.

4.2.4.3 Investment Framework and Evaluation

We invest an amount of Rs.1,00,000 in each portfolio as per the weights simulated by Monte Carlo and an EWP. The number of shares to be purchased and the investment amount in each stock is then calculated. Since this is an analytical study, the shares are purchased in fractions.

The performance of each portfolio is evaluated by calculating the Returns on Investment (ROI) and visualizing the daily cumulative returns during the investment holding period for each event. The ROI and cumulative returns are then compared to the benchmark indices (Nifty50 and Nifty100).

$$\textbf{Return on Investment (ROI)} = \frac{\text{(Current Sale value – Purchase Value)}}{\text{Purchase Value}} \times 100 \qquad (13)$$

Cumulative Returns (graphically depicted)

Chapter 5

Stock Market Prediction

5.1 Introduction

This chapter is divided into two parts. First, we present and discuss the results stemming from applying Mutual Information based on entropy to select features for predicting the Nifty broad market indices, sectoral indices, and selected stocks. Then, we delve into the performance of LSTM and ANN models in predicting the NSE indices and stocks across the seven events.

5.2 Results and Discussion

5.2.1 Feature Selection Using Mutual Information

Figures 5.1, 5.2, and 5.3 illustrate the Mutual Information scores for the broad market indices, sectoral indices, and stocks, respectively. Features with higher MI scores share more information with the target, while those with lower MI scores contribute little to reducing uncertainty. Training the machine learning models with all available features may lead to overfitting. Therefore, based on the MI scores, the 75th percentile of input variables that contribute to predicting the target close price are selected to build the model.

Intriguingly, the findings reveal that the 75th percentile subset primarily comprised the daily high and low prices, highlighting the significance of price volatility as a key determinant in predicting all the Nifty indices and stock prices.

Stock market data are more endogenous to the indices' behavior. In contrast, the macroeconomic indicators capture exogenous factors, such as global economic trends or national economic policies, which might not directly correlate with short-term stock market movements. Thus, focusing on short-term predictions might overlook the predictive power of macroeconomic indicators over extended periods. It is essential to consider that economic data reflects past trends rather than immediate market sentiment; therefore, it becomes challenging to predict short-term market movements accurately. Macroeconomic indicators might be more relevant for predicting longer-term trends in the stock market rather than short-term fluctuations. Moreover, it is possible that the stock market data already captures the information carried by macroeconomic variables. Thus, adding macroeconomic variables to predict Nifty indices or stock prices does not provide significant additional insight. Also, the frequency of data collection for certain macroeconomic variables (broad money, gold price, and GDP) does not align with the high-frequency stock market data. This discrepancy in time scales may have contributed to the diminishing strength of the relationship between the two sets of variables.

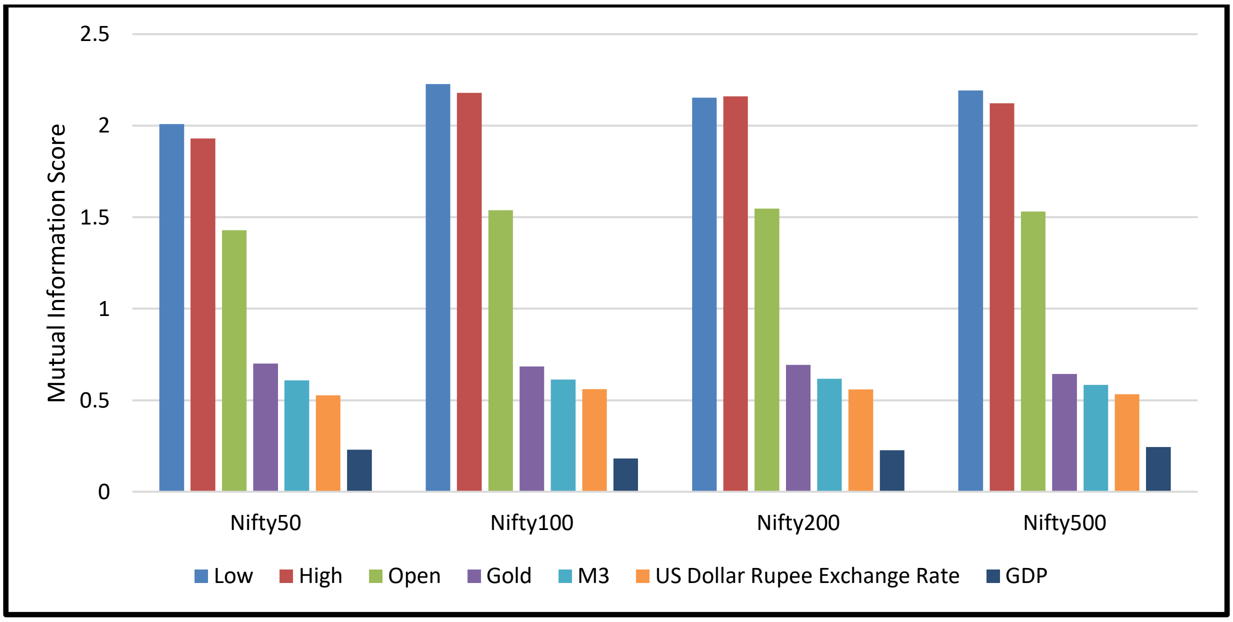


Figure 5.1: Mutual Information score for broad market indices

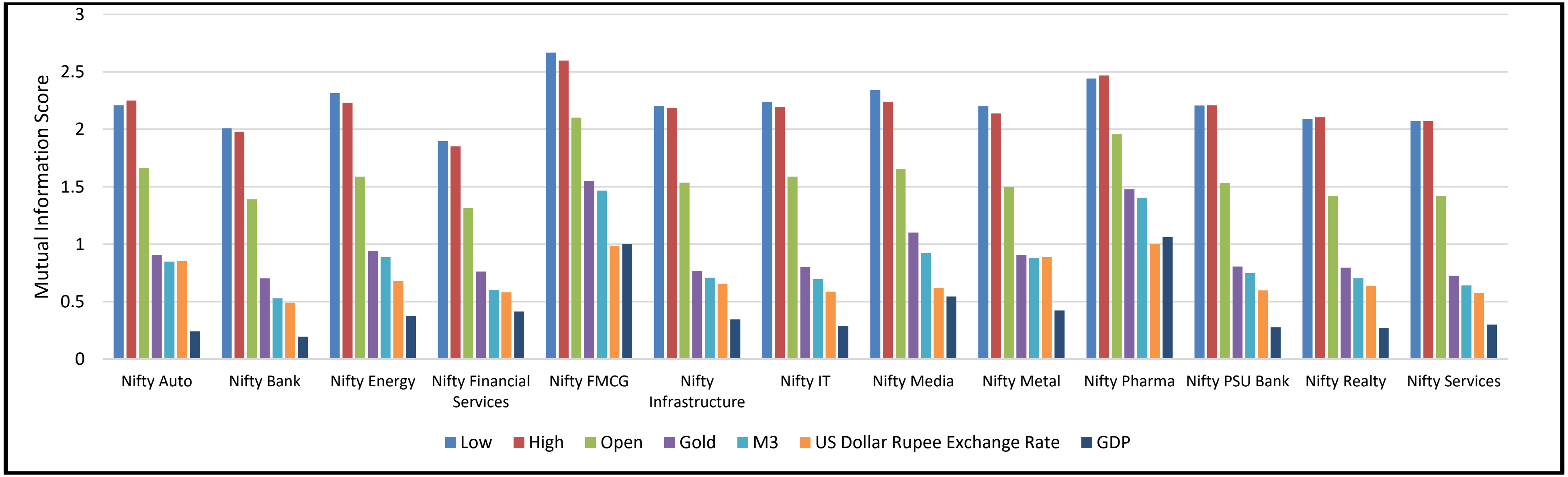


Figure 5.2: Mutual Information score for sectoral indices

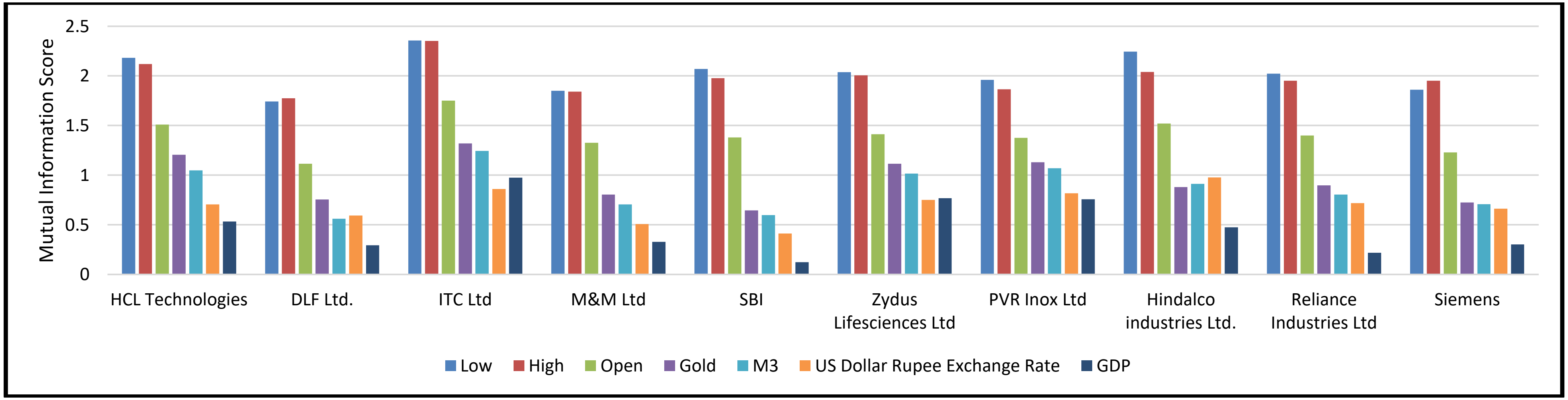


Figure 5.3: Mutual Information score for selected stocks

5.2.2 Event-wise Stock Market Prediction

This section compares the performance of LSTM and ANN models in predicting the Nifty broad market indices, sectoral indices, and stocks based on R-squared, RMSE, and MAE across seven events. These events include demonetization, implementation of the GST regime, the general elections of 2019, COVID-19, the invasion of Ukraine by Russia, the Hamas attack on Israel, and the general elections of 2024. We use line graphs to visualize each index and stock, comparing the predicted performance of LSTM (represented in blue) and ANN (represented in green) against their actual performance (represented in red).

5.2.2.1 Demonetization

Table 5.1: Performance evaluation of LSTM and ANN during demonetization

	LSTM			ANN		
Broad Market Indices	**R-Squared**	**RMSE**	**MAE**	**R-Squared**	**RMSE**	**MAE**
Nifty 50	0.9875	0.0355	0.0306	0.9493	0.0445	0.0370
Nifty 100	0.9647	0.0351	0.0308	0.9468	0.0403	0.0353
Nifty 200	0.9751	0.0281	0.0242	0.9444	0.0409	0.0360
Nifty 500	0.9699	0.0297	0.0258	0.9408	0.0425	0.0376
Sectoral Indices						
Auto	0.9689	0.0179	0.0127	0.9481	0.0223	0.0165
Bank	0.9807	0.0295	0.0249	0.9618	0.0374	0.0316
Energy	0.9706	0.0540	0.0444	0.9405	0.1045	0.0973
Financial Services	0.9796	0.0341	0.0294	0.9648	0.0389	0.0330
FMCG	0.9804	0.0335	0.0241	0.9661	0.0449	0.0344
Infrastructure	0.9792	0.0247	0.0207	0.9643	0.0270	0.0223
IT	0.8838	0.0154	0.0116	0.8411	0.0191	0.0150
Media	0.9499	0.0306	0.0238	0.9157	0.0379	0.0317
Metal	0.9723	0.0263	0.0198	0.9613	0.0263	0.0210
Pharma	0.9597	0.0183	0.0139	0.9571	0.0206	0.0159
PSU Bank	0.9097	0.0251	0.0189	0.9006	0.0268	0.0205
Realty	0.9864	0.0271	0.0198	0.9859	0.0272	0.0193
Services	0.9827	0.0276	0.0232	0.9597	0.0367	0.0314
Stocks						
HCL Technologies	0.7338	0.0145	0.0128	0.5004	0.0212	0.0199
DLF Ltd	0.9709	0.0251	0.0172	0.9309	0.0326	0.0250
ITC Ltd	0.9621	0.0338	0.0225	0.8980	0.0517	0.0402
M&M Ltd	0.9313	0.0264	0.0193	0.9029	0.0281	0.0202
SBI	0.8176	0.0023	0.0017	0.7883	0.0025	0.0020
Zydus Lifesciences Ltd	0.9350	0.0085	0.0061	0.9061	0.0116	0.0090
P V R Inox Ltd.	0.9658	0.0250	0.0188	0.9497	0.0317	0.0243
Hindalco Industries Ltd.	0.9498	0.0371	0.0297	0.9246	0.0420	0.0320
Reliance Industries Ltd.	0.9035	0.1953	0.1210	0.8985	0.2029	0.1037
Siemens Ltd.	0.9568	0.0188	0.0135	0.9367	0.0248	0.0191

Table 5.1 displays the performance evaluation metrics of LSTM and ANN during demonetization across the broad market indices, sectoral indices, and stock price prediction. The R-squared values closer to 1 indicate that the predictor variables (high price and low price) are a good fit to the model. The R-squared value for most indices and stocks was greater than 0.90, except for the IT index, HCL Technologies, and SBI. Thus, 90% of the variance in the closing price was explained by the predictor variables. Furthermore, HCL Technologies had the lowest R-squared values compared to other stocks. This observation indicates increased volatility or unique market factors influencing the stock.

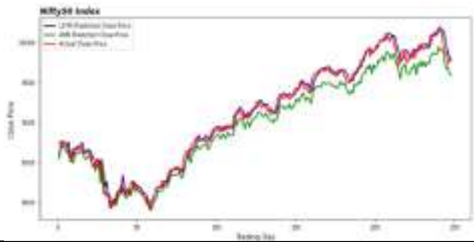
The RMSE values for LSTM ranged from 0.0281 to 0.0355, while ANN's RMSE values spanned 0.0403 to 0.0445, highlighting LSTM's ability to capture market dynamics more effectively.

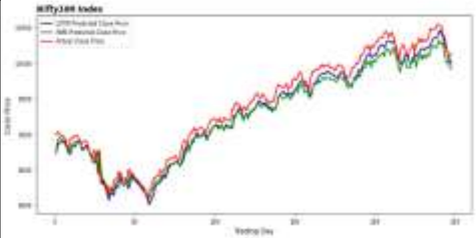
Among the sectoral indices, the Nifty IT index achieved the lowest RMSE and MAE values using both models. LSTM recorded an RMSE of 0.0154 and MAE of 0.0116, while ANN followed closely with an RMSE of 0.0191 and MAE of 0.0150. The Nifty Auto index was another strong performer, with LSTM achieving an RMSE of 0.0179 and an MAE of 0.0127, reinforcing LSTM's consistent advantage. Interestingly, ANN's next best performance was observed in predicting the Nifty Pharma index, although it lagged behind LSTM in overall accuracy. While most sectoral indices exhibited RMSE values close to zero, the Nifty Energy index showcased a higher RMSE (LSTM: 0.0540, ANN: 0.1045) and MAE (LSTM: 0.0444, ANN: 0.0973) using both models.

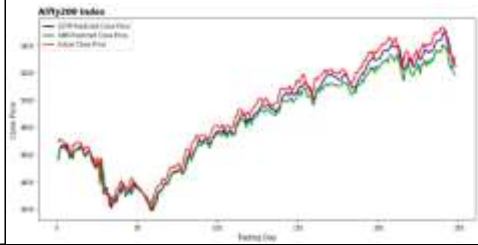
State Bank of India (SBI) had the best predictive performance among the selected stocks. LSTM achieved an impressive RMSE of 0.0023, narrowly beating ANN's 0.0025. Zydus Lifesciences Ltd. also showcased strong performance, with LSTM recording an RMSE of 0.0085 and ANN reporting 0.0116. On the other hand, Reliance Industries Ltd. demonstrated significantly higher RMSE values across both models, signaling greater difficulty in predicting its price movements accurately.

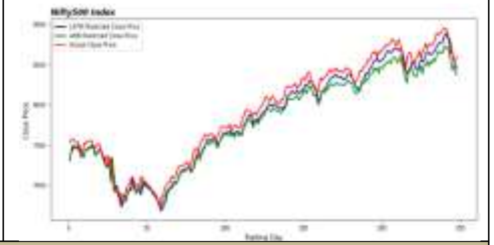
The LSTM model outperformed ANN, as reflected in its consistently lower RMSE and MAE and higher R-squared values. The actual and predicted performance of LSTM and ANN during demonetization from 1st October 2016 to 30th September 2017 across the broad market indices, sectoral indices, and stocks are displayed in Figure 5.4. The ANN predicted close prices are represented in green; the LSTM predicted close prices in blue against the daily actual closing prices in red.

Broad Market Indices

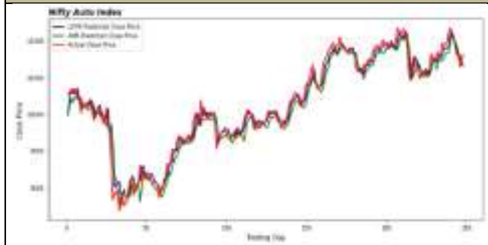


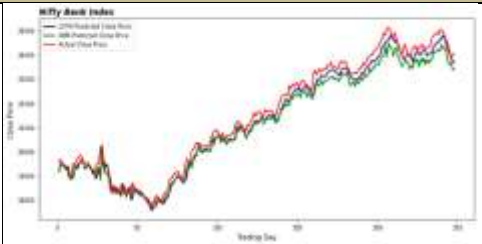


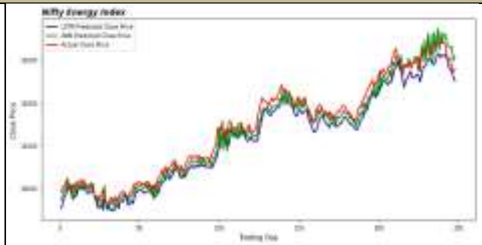


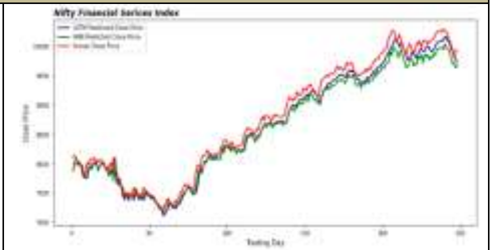


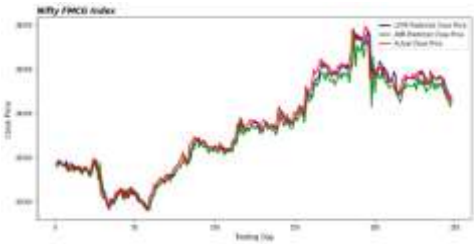
Sectoral Indices

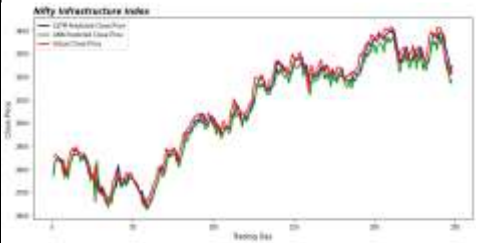


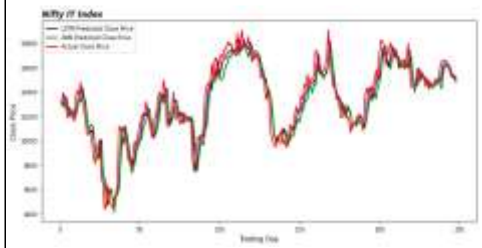


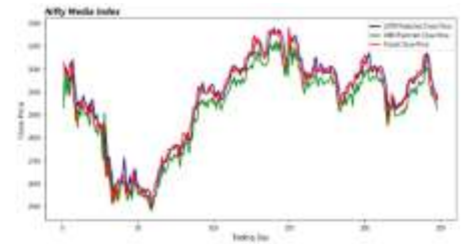


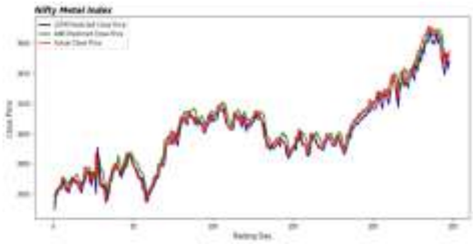


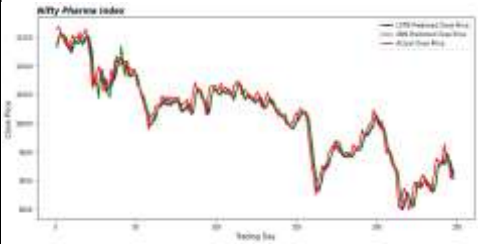


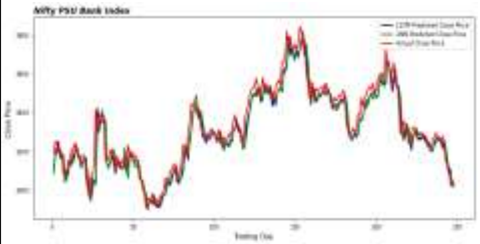


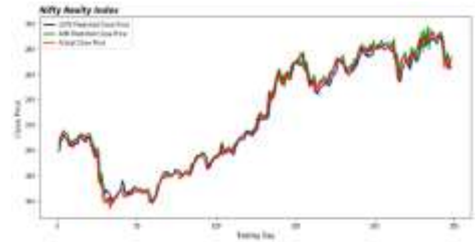


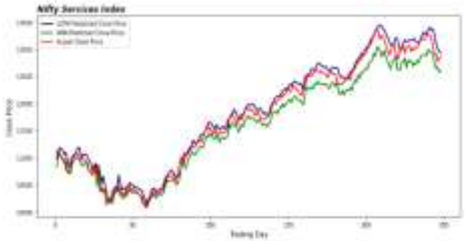



Nifty Metal Index
Close Price
Trading Day

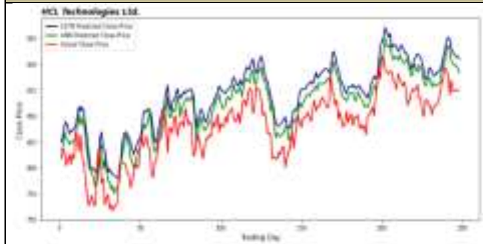

Nifty Pharma Index
Close Price
Trading Day

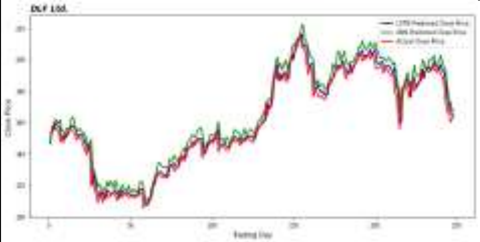

Nifty PSU Bank Index
Close Price
Trading Day

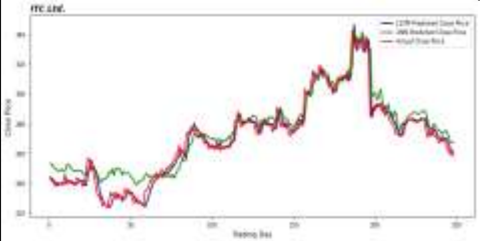

Nifty Realty Index
Close Price
Trading Day

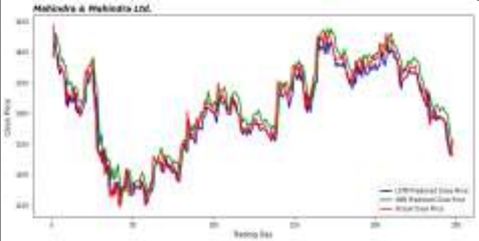

Nifty Services Index
Close Price
Trading Day

Stocks


HCL Technologies Ltd.
Close Price
Trading Day


DLF Ltd.
Close Price
Trading Day


ITC Ltd.
Close Price
Trading Day


Mahindra & Mahindra Ltd.
Close Price
Trading Day

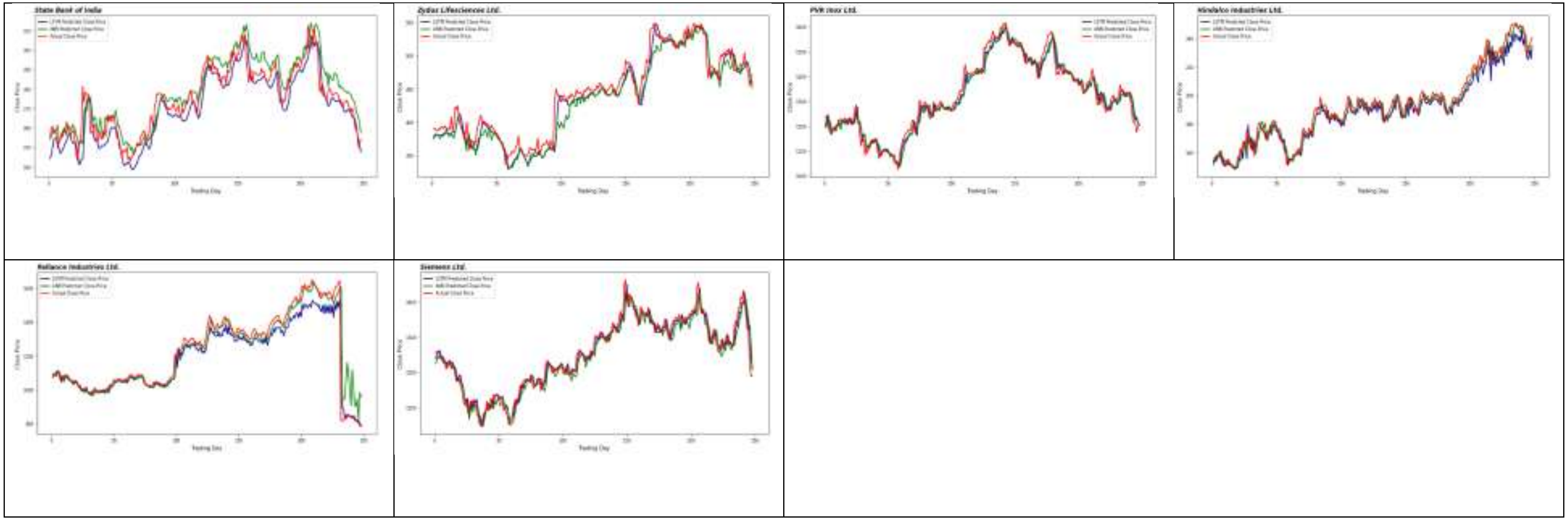


Figure 5.4: Actual and predicted performance of LSTM and ANN during demonetization from 01st October 2016 – 30th September 2017 across the broad market indices, sectoral indices, and selected stocks. ANN predicted close prices (green), the LSTM predicted close prices (blue), and the daily actual close prices (red)

5.2.2.2 Implementation of the GST Regime

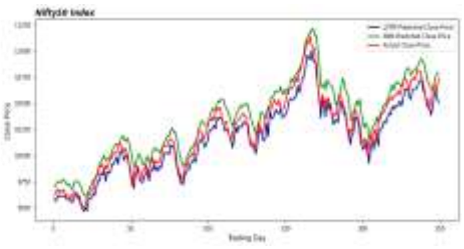
Table 5.2: Performance evaluation of LSTM and ANN during the implementation of the GST regime

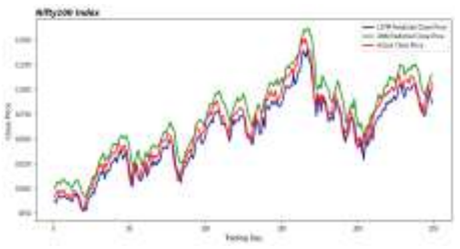
	LSTM			ANN		
Broad Market Indices	**R-Squared**	**RMSE**	**MAE**	**R-Squared**	**RMSE**	**MAE**
Nifty 50	0.9240	0.0234	0.0198	0.8641	0.0274	0.0211
Nifty 100	0.9332	0.0214	0.0183	0.8704	0.0262	0.0204
Nifty 200	0.9425	0.0203	0.0170	0.9185	0.0229	0.0174
Nifty 500	0.9234	0.0193	0.0166	0.9056	0.0208	0.0152
Sectoral Indices						
Auto	0.7561	0.0301	0.0252	0.6808	0.0304	0.0270
Bank	0.8863	0.0213	0.0175	0.6543	0.0445	0.0408
Energy	0.9318	0.0382	0.0315	0.8297	0.0470	0.0387
Financial Services	0.8673	0.0255	0.0217	0.6914	0.0428	0.0386
FMCG	0.9073	0.0304	0.0224	0.5553	0.0537	0.0474
Infrastructure	0.9450	0.0196	0.0153	0.5947	0.0412	0.0357
IT	0.9853	0.0212	0.0169	0.9804	0.0226	0.0176
Media	0.9474	0.0221	0.0175	0.6568	0.0566	0.0516
Metal	0.9720	0.0271	0.0208	0.8853	0.0631	0.0565
Pharma	0.9168	0.0180	0.0138	0.7748	0.0254	0.0208
PSU Bank	0.9474	0.0346	0.0202	0.9349	0.0386	0.0224
Realty	0.8774	0.0578	0.0457	0.8339	0.0639	0.0571
Services	0.9465	0.0221	0.0184	0.6935	0.0520	0.0484
Stocks						
HCL Technologies	0.8946	0.0126	0.0089	0.7146	0.0195	0.0157
DLF Ltd	0.9377	0.0372	0.0287	0.9270	0.0528	0.0428
ITC Ltd	0.9452	0.0298	0.0215	0.9134	0.0452	0.0374
M&M Ltd	0.9243	0.0724	0.0216	0.8189	0.1154	0.0506
SBI	0.8955	0.0032	0.0020	0.7787	0.0041	0.0030
Zydus Lifesciences Ltd	0.9332	0.0088	0.0075	0.9230	0.0097	0.0077
P V R Inox Ltd.	0.8404	0.0232	0.0177	0.7297	0.0268	0.0197
Hindalco Industries Ltd.	0.9684	0.0376	0.0285	0.9123	0.0895	0.0821
Reliance Industries Ltd.	0.9593	0.0825	0.0272	0.9274	0.1036	0.0388
Siemens Ltd.	0.9600	0.0190	0.0141	0.6355	0.0525	0.0470

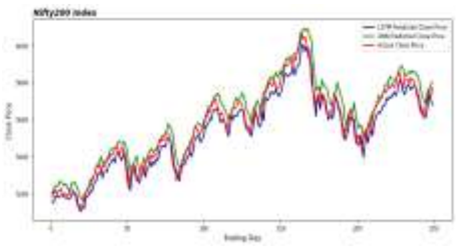
Table 5.2 compares the Nifty prediction performance of LSTM and ANN based on R-squared, RMSE, and MAE values. LSTM has a lower MAE compared to ANN in terms of the prediction performance of the four broad market indices.

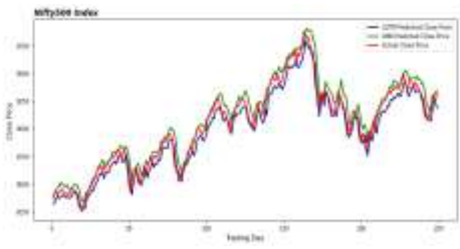
Amongst the sectoral indices, the highest RMSE and MAE were witnessed in the case of the Nifty Realty index using both LSTM and ANN. The RMSE was 0.0578, and the MAE was 0.0457 using LSTM. The RMSE was 0.0639, and the MAE was 0.0571 using ANN. From the entire experimental setup, SBI and Zydus Lifesciences Ltd. had the lowest RMSE of 0.0032 and 0.0088, respectively, by applying LSTM. LSTM consistently outperforms ANN with lower RMSE and MAE across all indices and selected stocks

Broad Market Indices

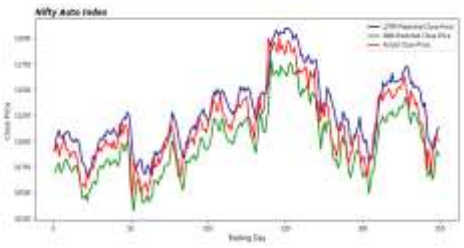


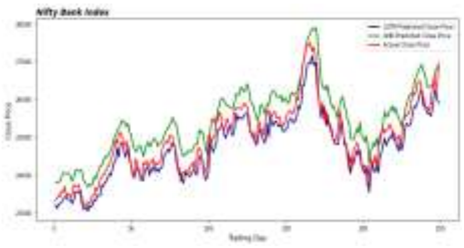


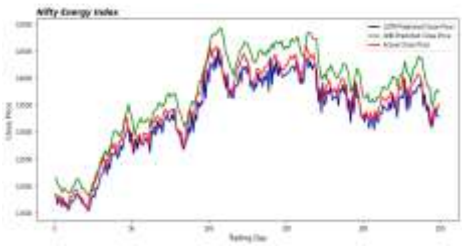


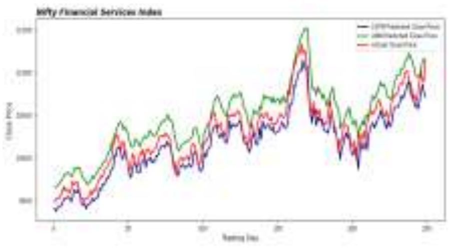


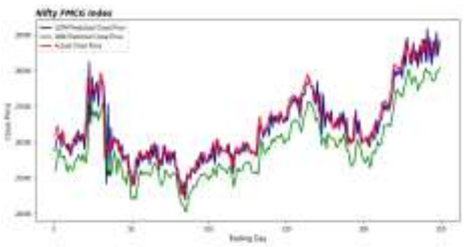
Sectoral Indices

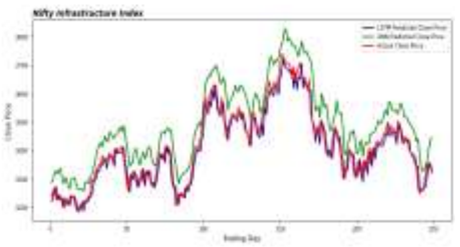


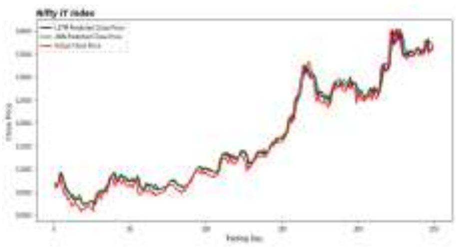


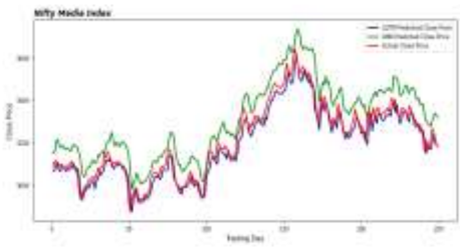


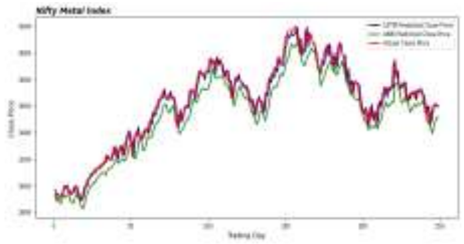


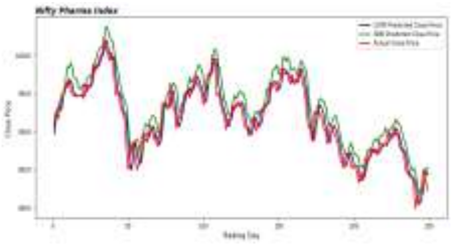


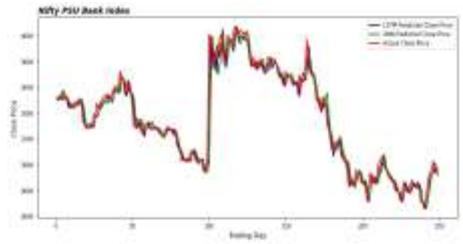


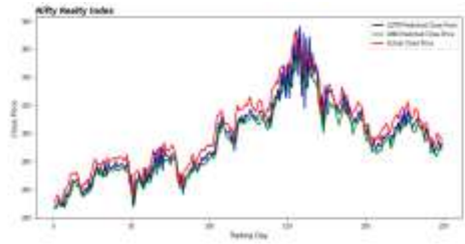


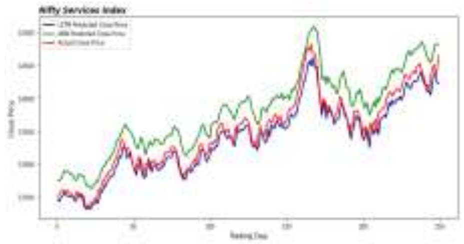



Nifty Metal Index
Close Price
Trading Day

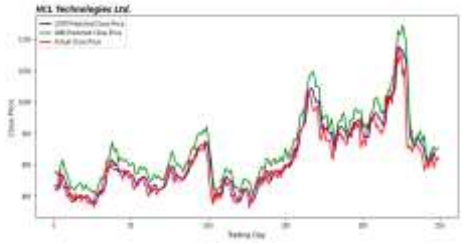

Nifty Pharma Index
Close Price
Trading Day

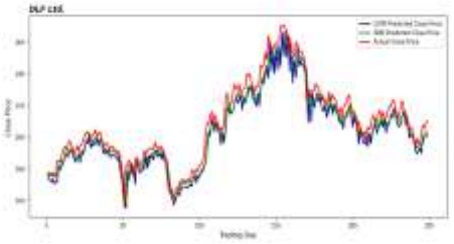

Nifty PSU Bank Index
Close Price
Trading Day

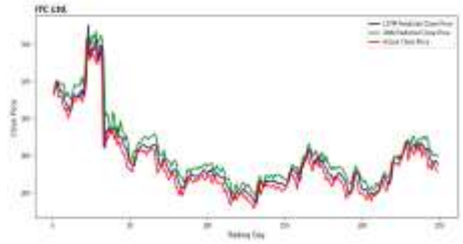

Nifty Realty Index
Close Price
Trading Day

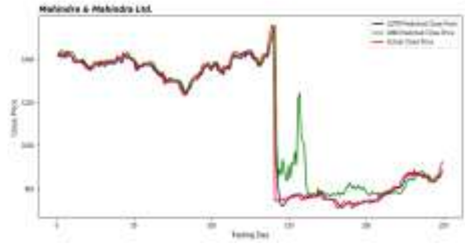

Nifty Services Index
Close Price
Trading Day

Stocks


HCL Technologies Ltd.
Close Price
Trading Day


DLF Ltd.
Close Price
Trading Day


ITC Ltd.
Close Price
Trading Day


Mahindra & Mahindra Ltd.
Close Price
Trading Day

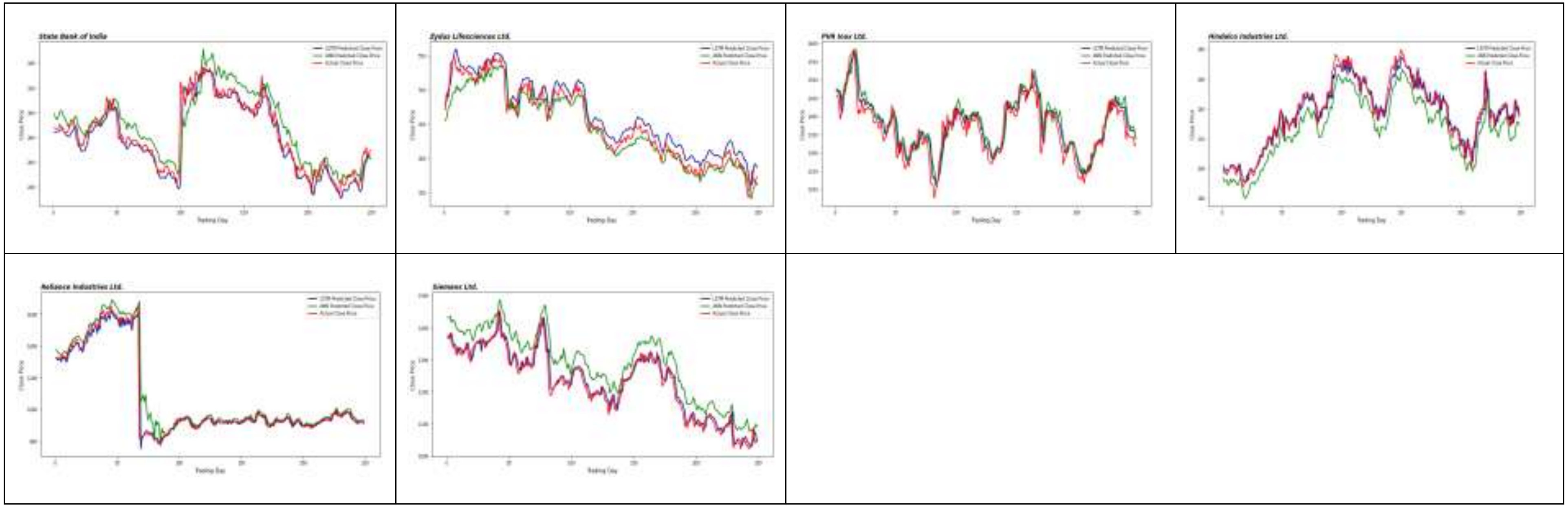


Figure 5.5: Actual and predicted performance of LSTM and ANN during the implementation of the GST regime from 1st June 2017 – 31st May 2018 across the broad market indices, sectoral indices, and selected stocks. ANN predicted close prices (green), the LSTM predicted close prices (blue), and the daily actual close prices (red).

5.2.2.3 The General Elections 2019

Table 5.3: Performance evaluation of LSTM and ANN during the general elections 2019

	LSTM			ANN		
Broad Market Indices	**R-Squared**	**RMSE**	**MAE**	**R-Squared**	**RMSE**	**MAE**
Nifty 50	0.9574	0.0370	0.0231	0.9393	0.0452	0.0323
Nifty 100	0.9530	0.0360	0.0229	0.9351	0.0404	0.0254
Nifty 200	0.9536	0.0352	0.0226	0.9120	0.0550	0.0451
Nifty 500	0.9552	0.0342	0.0219	0.9249	0.0458	0.0312
Sectoral Indices						
Auto	0.9602	0.0316	0.0211	0.5415	0.1036	0.0458
Bank	0.9494	0.0362	0.0234	0.9229	0.0435	0.0263
Energy	0.9667	0.0269	0.0193	0.9371	0.0388	0.0322
Financial Services	0.9348	0.0428	0.0298	0.9190	0.0460	0.0291
FMCG	0.8529	0.0388	0.0321	0.6082	0.0514	0.0462
Infrastructure	0.9571	0.0328	0.0229	0.9260	0.0442	0.0323
IT	0.9107	0.0405	0.0332	0.8892	0.0434	0.0264
Media	0.9571	0.0351	0.0250	0.6086	0.1582	0.1011
Metal	0.9721	0.0214	0.0165	0.9556	0.0317	0.0264
Pharma	0.9524	0.0209	0.0147	0.7356	0.1003	0.0657
PSU Bank	0.9838	0.0293	0.0218	0.8231	0.0899	0.0429
Realty	0.9601	0.0231	0.0170	0.9399	0.0265	0.0189
Services	0.9460	0.0376	0.0228	0.9267	0.0432	0.0265
Stocks						
HCL Technologies	0.8108	0.1061	0.0583	0.3940	0.4795	0.2727
DLF Ltd	0.9588	0.0340	0.0241	0.9473	0.0368	0.0264
ITC Ltd	0.9818	0.0334	0.0219	0.5776	0.1571	0.0844
M&M Ltd	0.9106	0.0256	0.0206	0.7735	0.0989	0.0660
SBI	0.9480	0.0473	0.0366	0.9360	0.0488	0.0375
Zydus Lifesciences Ltd	0.8463	0.0063	0.0053	0.8692	0.0528	0.0483
P V R Inox Ltd.	0.9153	0.0491	0.0410	0.8596	0.0590	0.0447
Hindalco Industries Ltd.	0.9668	0.0211	0.0161	0.9482	0.0285	0.0226
Reliance Industries Ltd.	0.9530	0.0358	0.0255	0.9313	0.0508	0.0400
Siemens Ltd.	0.9422	0.0621	0.0482	0.9569	0.0653	0.0524

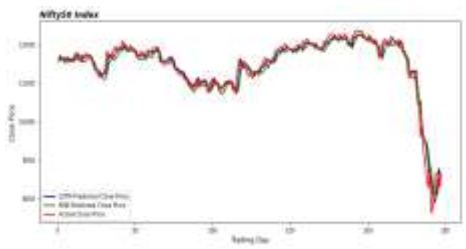
In most sectoral predictions, the differences in RMSE values between LSTM and ANN are relatively small. However, sectors like Nifty Auto, Nifty Media, Nifty Pharma, and Nifty PSU Bank stand out with significantly larger differences in terms of RMSE.

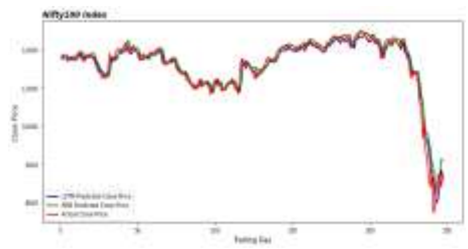
Amongst the individual stocks analyzed, Zydus Lifesciences Ltd. achieved the lowest RMSE of 0.0063 on applying LSTM. It is worth noting that all MAE values remained below 0.05 when applying LSTM for predictions, with one exception: HCL Technologies Ltd., which recorded a slightly higher MAE of 0.0583. The R-squared value is lower for this company compared to the stocks (0.81 for LSTM and 0.39 for ANN). Meanwhile, the R-squared values obtained on making predictions by applying the LSTM model were above 0.90 (except for Nifty FMCG:

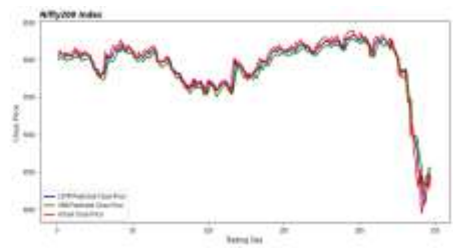
0.8529 and Zydus Lifesciences Ltd.: 0.8463), indicating that the predictor variables could explain around 90% of the variance in the close price.

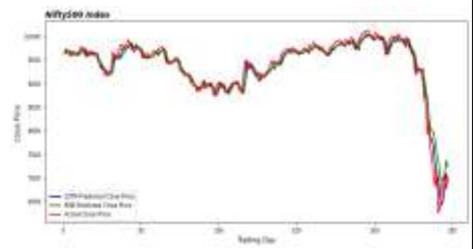
Overall, the LSTM model consistently outperformed ANN during the 2019 elections, highlighting its capability in financial predictions.

Broad Market Indices

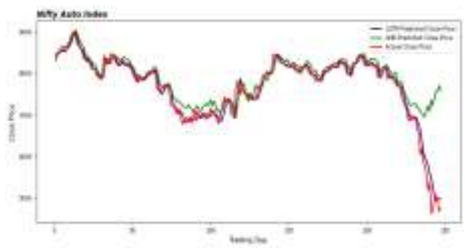


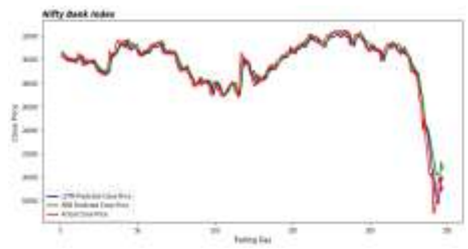


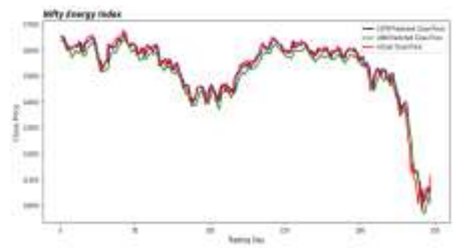


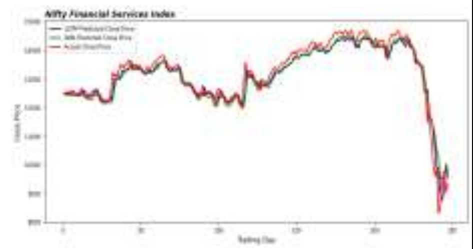


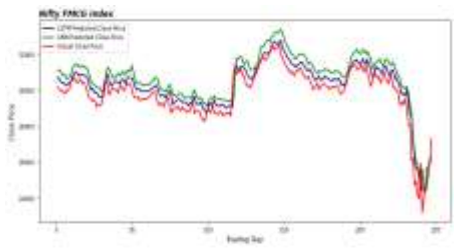
Sectoral Indices

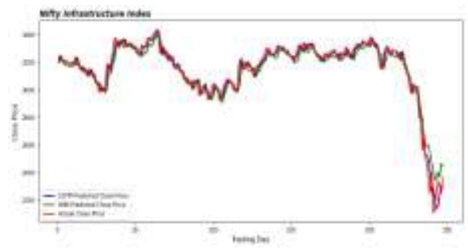


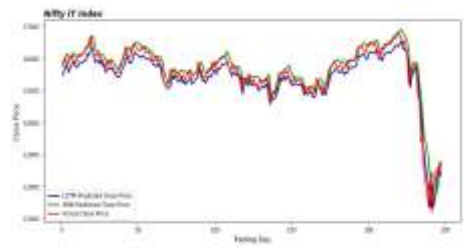


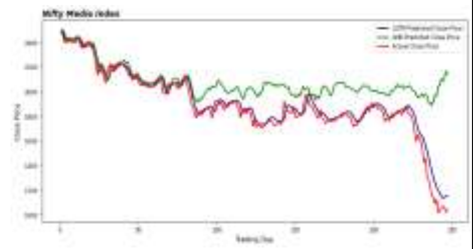


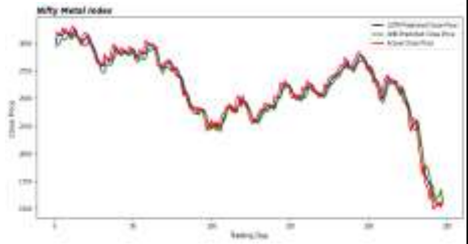


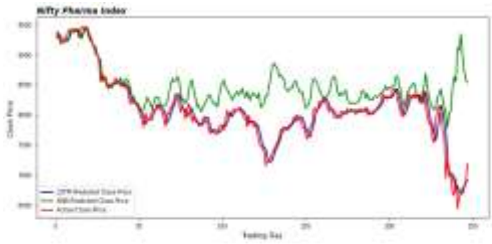


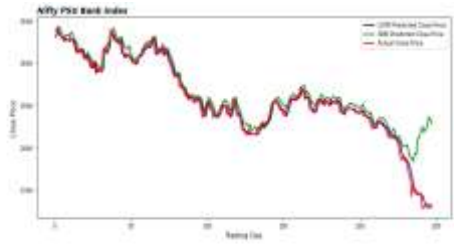


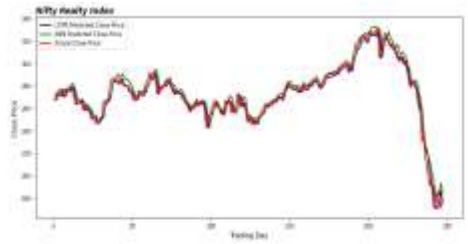


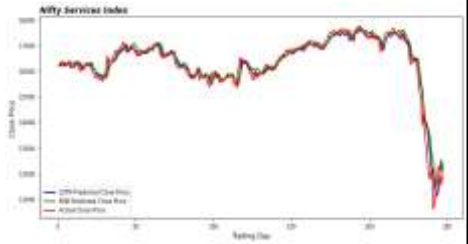



Nifty Metal Index

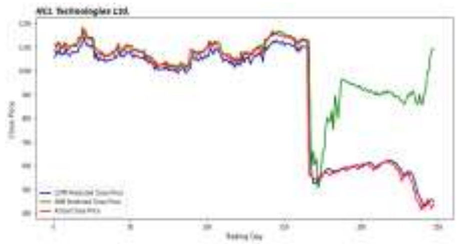

Nifty Pharma Index

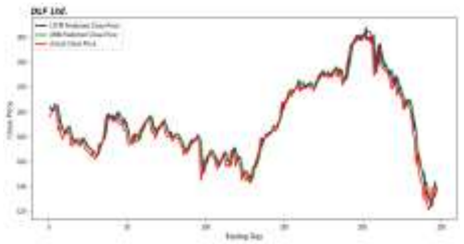

Nifty PSU Bank Index

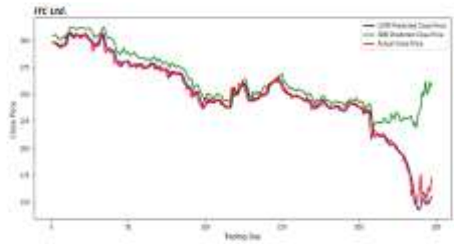

Nifty Realty Index

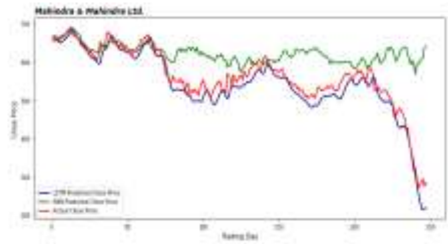

Nifty Services Index

Stocks


HCL Technologies Ltd.


DLF Ltd.


ITC Ltd.


Mahindra & Mahindra Ltd.

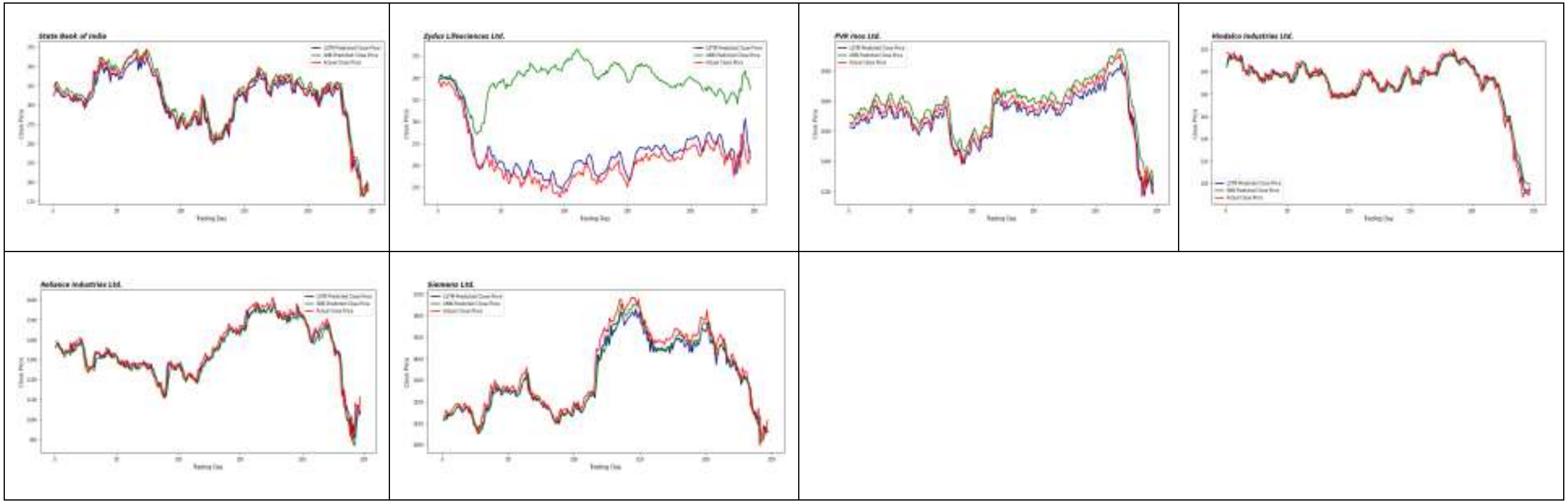


Figure 5.6: Actual and predicted performance of LSTM and ANN during the general elections 2019 from 1st April 2019 – 31st March 2020 across the broad market indices, sectoral indices, and selected stocks. ANN predicted close prices (green), the LSTM predicted close prices (blue), and the daily actual close prices (red).

5.2.2.4 The COVID-19 Pandemic

Table 5.4: Performance evaluation of LSTM and ANN during the COVID-19 pandemic

	LSTM			ANN		
Broad Market Indices	**R-Squared**	**RMSE**	**MAE**	**R-Squared**	**RMSE**	**MAE**
Nifty 50	0.9807	0.0421	0.0312	0.9290	0.0724	0.0514
Nifty 100	0.9615	0.0416	0.0306	0.9399	0.0649	0.0444
Nifty 200	0.9819	0.0416	0.0302	0.9426	0.0659	0.0438
Nifty 500	0.9829	0.0416	0.0301	0.9397	0.0697	0.0478
Sectoral Indices						
Auto	0.9845	0.0324	0.0233	0.6305	0.1574	0.0879
Bank	0.9736	0.0389	0.0283	0.9342	0.0599	0.0373
Energy	0.9726	0.0317	0.0235	0.7902	0.0879	0.0780
Financial Services	0.9718	0.0380	0.0277	0.9624	0.0428	0.0290
FMCG	0.9563	0.0325	0.0226	0.9288	0.0357	0.0252
Infrastructure	0.9802	0.0360	0.0252	0.8830	0.0833	0.0631
IT	0.9870	0.0684	0.0497	0.9452	0.1317	0.0909
Media	0.9550	0.0244	0.0195	0.7186	0.2028	0.1535
Metal	0.9848	0.0238	0.0181	0.9507	0.0390	0.0262
Pharma	0.9798	0.0517	0.0405	0.9661	0.0646	0.0445
PSU Bank	0.9626	0.0262	0.0202	0.9432	0.3839	0.3441
Realty	0.9811	0.0276	0.0201	0.9713	0.0312	0.0235
Services	0.9784	0.0416	0.0307	0.9530	0.0602	0.0366
Stocks						
HCL Technologies	0.9870	0.0292	0.0218	0.9281	0.0684	0.0416
DLF Ltd	0.9743	0.0342	0.0248	0.9405	0.0481	0.0374
ITC Ltd	0.8633	0.0426	0.0317	0.4935	0.2744	0.2459
M&M Ltd	0.9788	0.0186	0.0141	0.5070	0.0893	0.0508
SBI	0.9781	0.0317	0.0235	0.9410	0.0473	0.0321
Zydus Lifesciences Ltd	0.9782	0.0298	0.0213	0.9694	0.0387	0.0285
P V R Inox Ltd.	0.9862	0.0392	0.0280	0.8955	0.0580	0.0405
Hindalco Industries Ltd.	0.9682	0.0281	0.0210	0.9243	0.0416	0.0303
Reliance Industries Ltd.	0.9835	0.0706	0.0561	0.9771	0.1069	0.0883
Siemens Ltd.	0.9738	0.0357	0.0253	0.9717	0.0357	0.0257

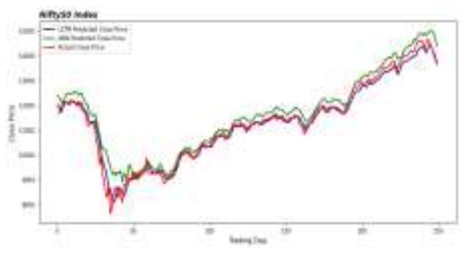
Table 5.4 highlights the performance of LSTM and ANN models for financial predictions during the COVID-19 pandemic across indices, sectors, and selected stocks, measured by R-squared, RMSE, and MAE.

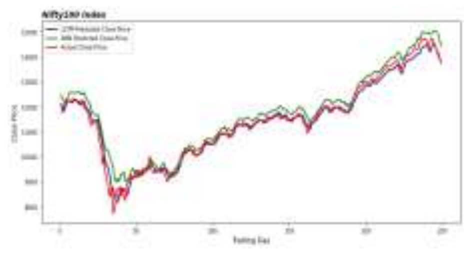
LSTM demonstrates significantly lower RMSE and MAE than ANN across all broad market indices, indicating its ability to adapt better to the highly dynamic and volatile market conditions during the pandemic. The R-squared values achieved for this category of indices range from 96.15% to 98.29% when applying LSTM. The R-squared values achieved by applying ANN ranged between 92.90% to 94.26%. These R-squared values indicate that the predictor variables are a good fit to the model.

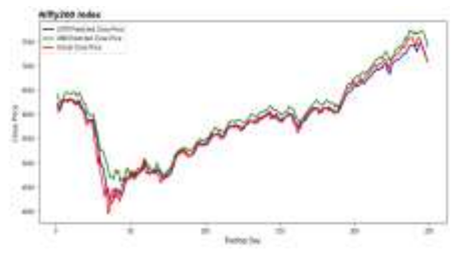
In the sectoral indices, LSTM exhibited a notable edge over ANN in several key sectors, including Nifty Auto, Nifty Energy, Nifty Infrastructure, Nifty IT, Nifty Media, and Nifty PSU Bank, where the RMSE differences were particularly substantial.

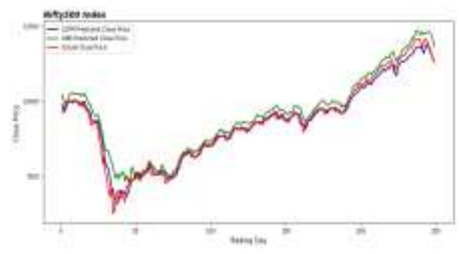
Similarly, in the stock-level predictions, the RMSE values ranged from 0.0186 to 0.0706 when using LSTM and 0.0357 to 0.2744 when using ANN. The RMSE for LSTM predicted prices were lower than ANN predicted values, further underscoring LSTM's superior accuracy during this unpredictable period.

Broad Market Indices

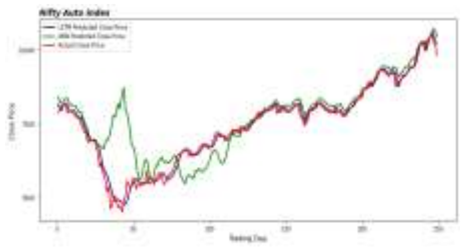


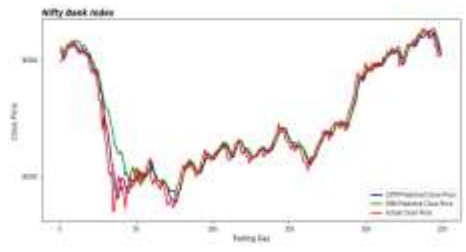


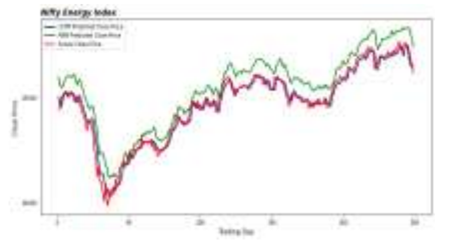


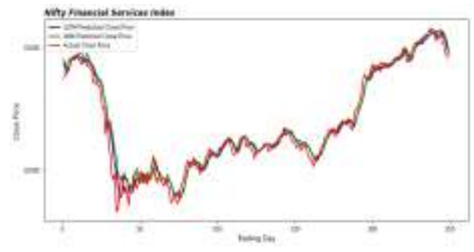


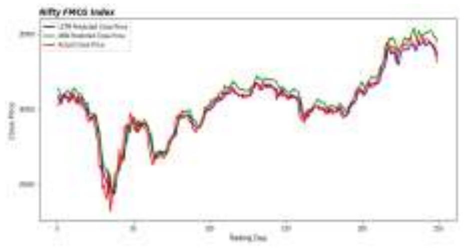
Sectoral Indices

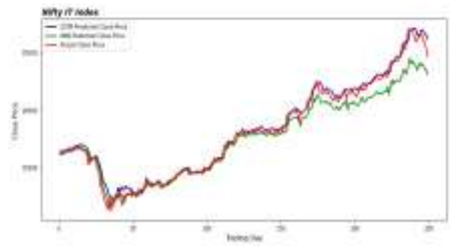


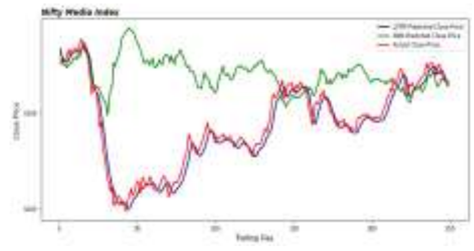


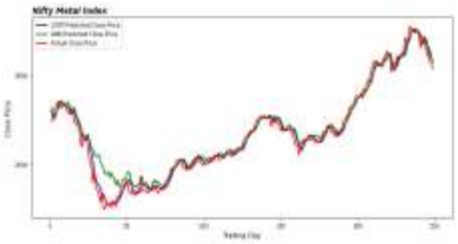


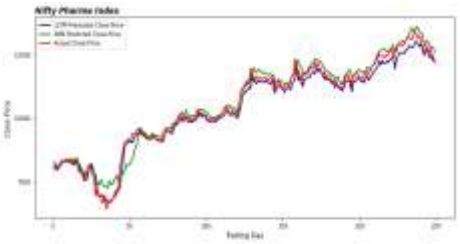


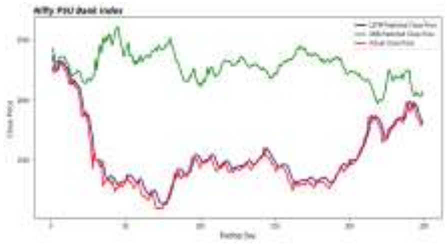


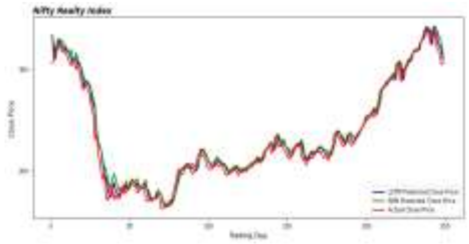


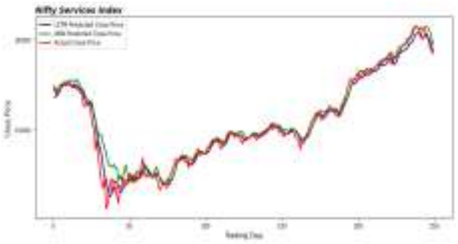



Nifty Metal Index

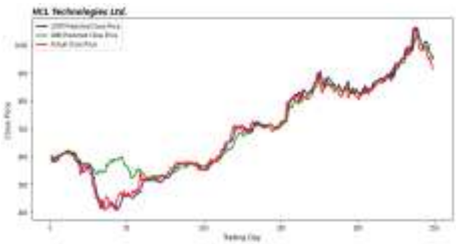

Nifty Pharma Index

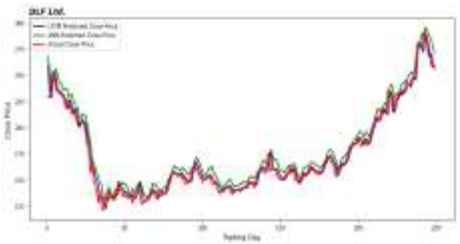

Nifty PSU Bank Index

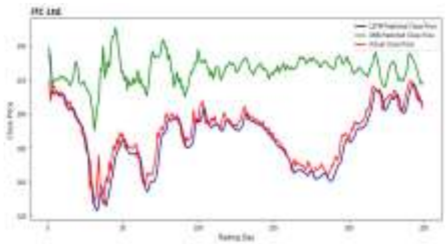

Nifty Realty Index

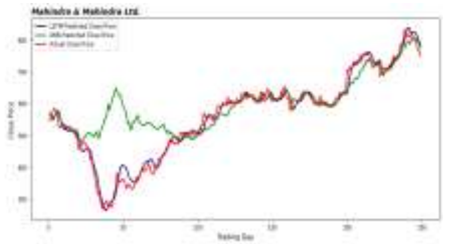

Nifty Services Index

Stocks


HCL Technologies Ltd.


DLF Ltd.


ITC Ltd.


Mahindra & Mahindra Ltd.

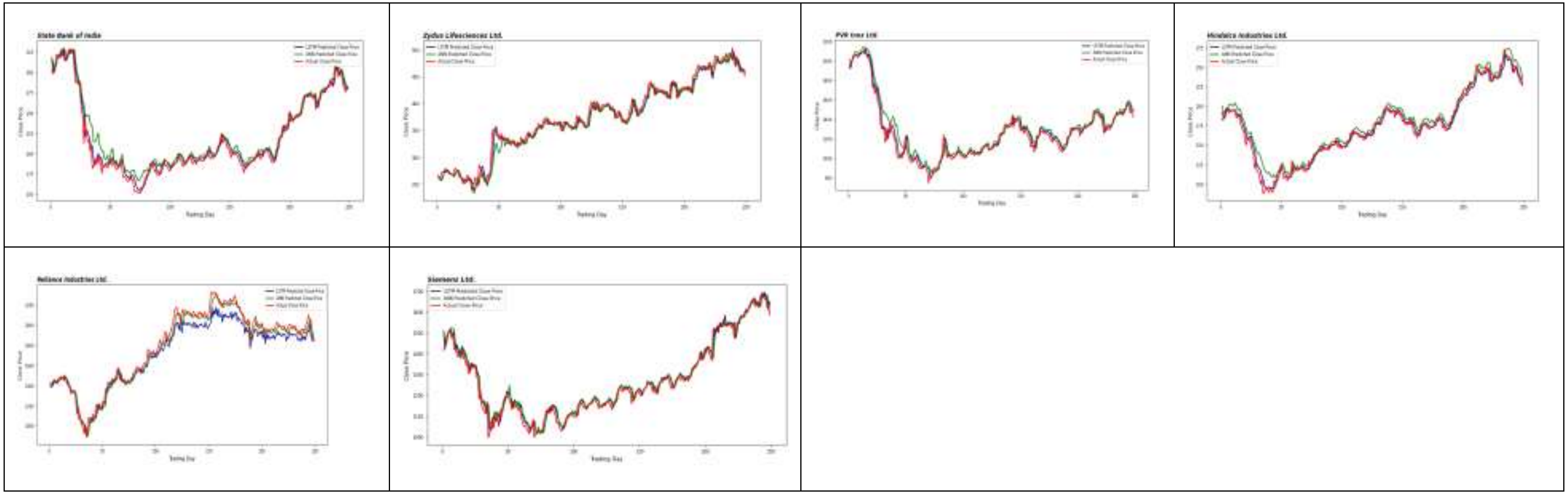


Figure 5.7: Actual and predicted performance of LSTM and ANN during the COVID-19 pandemic from 1st February 2020 – 31st January 2021 across the broad market indices, sectoral indices, and selected stocks. ANN predicted close prices (green), the LSTM predicted close prices (blue), and the daily actual close prices (red).

5.2.2.5 The Invasion of Ukraine by Russia

Table 5.5: Performance evaluation of LSTM and ANN during the invasion of Ukraine by Russia

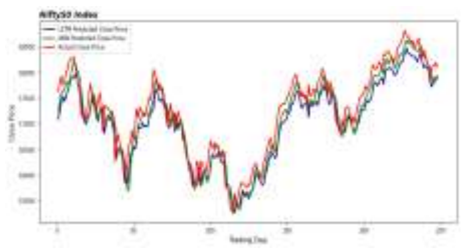
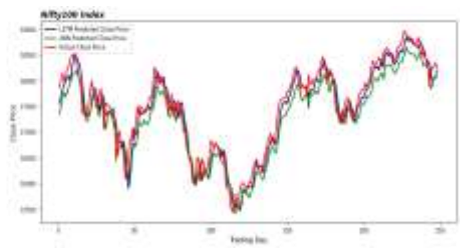
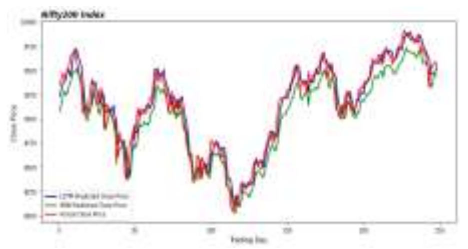
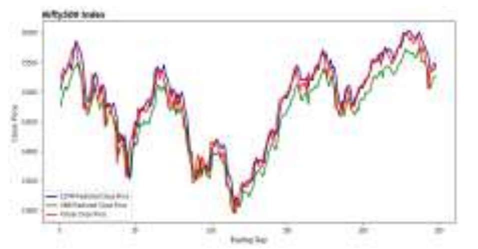
	LSTM			ANN		
Broad Market Indices	**R-Squared**	**RMSE**	**MAE**	**R-Squared**	**RMSE**	**MAE**
Nifty 50	0.9124	0.0279	0.0241	0.8676	0.0602	0.0366
Nifty 100	0.9297	0.0259	0.0217	0.8748	0.0271	0.0234
Nifty 200	0.9369	0.0220	0.0180	0.8758	0.0271	0.0235
Nifty 500	0.9278	0.0207	0.0165	0.8808	0.0266	0.0231
Sectoral Indices						
Auto	0.9587	0.0271	0.0203	0.9561	0.0278	0.0226
Bank	0.9546	0.0258	0.0209	0.9309	0.0325	0.0275
Energy	0.8614	0.0266	0.0207	0.6555	0.0512	0.0465
Financial Services	0.9410	0.0254	0.0205	0.8957	0.0333	0.0288
FMCG	0.9654	0.0451	0.0388	0.9444	0.0528	0.0466
Infrastructure	0.9018	0.0254	0.0217	0.8449	0.0294	0.0258
IT	0.9307	0.0324	0.0277	0.8273	0.0510	0.0454
Media	0.9079	0.0154	0.0115	0.7722	0.0196	0.0148
Metal	0.8942	0.0356	0.0280	0.8666	0.0391	0.0340
Pharma	0.8903	0.0190	0.0151	0.4083	0.0427	0.0384
PSU Bank	0.9822	0.0242	0.0187	0.9819	0.0249	0.0180
Realty	0.9047	0.0230	0.0173	0.7797	0.0357	0.0316
Services	0.9115	0.0287	0.0246	0.8945	0.0305	0.0258
Stocks						
HCL Technologies	0.9557	0.0228	0.0169	0.9358	0.0276	0.0231
DLF Ltd	0.7539	0.0441	0.0362	0.7241	0.0481	0.0426
ITC Ltd	0.9913	0.0242	0.0183	0.9905	0.0259	0.0202
M&M Ltd	0.9814	0.0350	0.0271	0.9746	0.0469	0.0385
SBI	0.9529	0.0323	0.0263	0.9247	0.0341	0.0283
Zydus Lifesciences Ltd	0.9356	0.0176	0.0137	0.9145	0.0208	0.0166
P V R Inox Ltd.	0.8759	0.0378	0.0302	0.8533	0.0489	0.0429
Hindalco Industries Ltd.	0.9556	0.0406	0.0326	0.9497	0.0409	0.0338
Reliance Industries Ltd.	0.6243	0.0398	0.0343	0.3113	0.0890	0.0829
Siemens Ltd.	0.8607	0.0325	0.0255	0.8455	0.0574	0.0508

The LSTM model consistently achieved lower RMSE values across all broad indices, with a particularly notable difference observed in the Nifty 50 index. The RMSE for LSTM for this index was 0.0279, which is less than half of ANN's RMSE of 0.0602, highlighting LSTM's superior predictive accuracy.

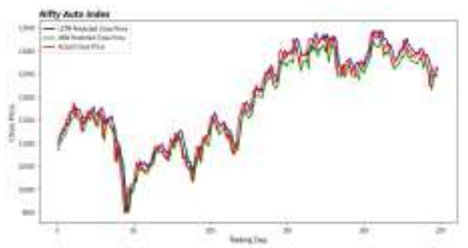
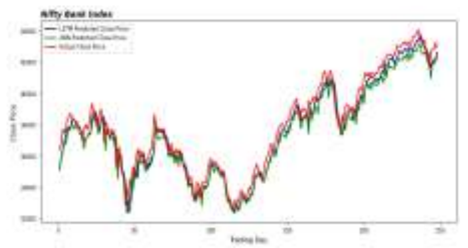
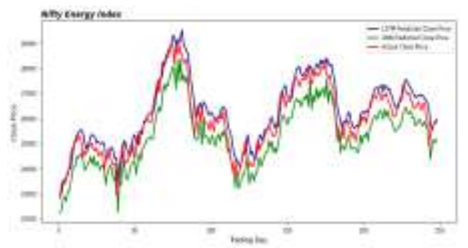
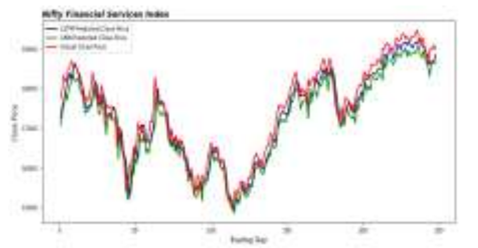
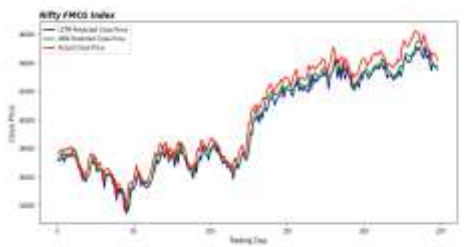
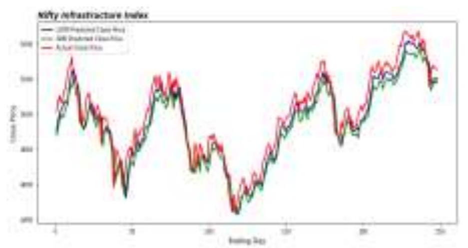
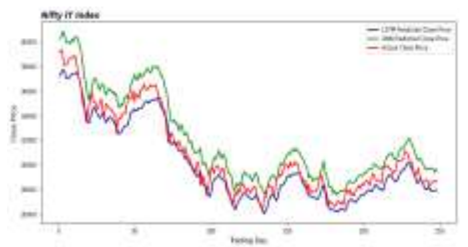
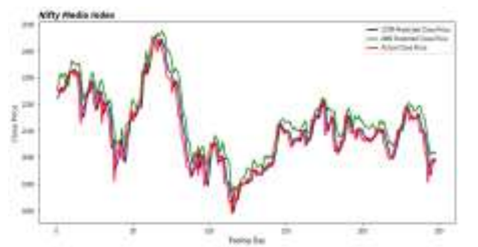
The Nifty FMCG sector recorded the highest RMSE in the sectoral indices, at 0.0451 for LSTM and 0.0528 for ANN. On the other hand, the Nifty Media sector stood out with the lowest MAE, not only among all sectoral indices but also across broad indices and individual stocks. The MAE for LSTM in this sector was as low as 0.0115 compared to ANN's 0.0148.

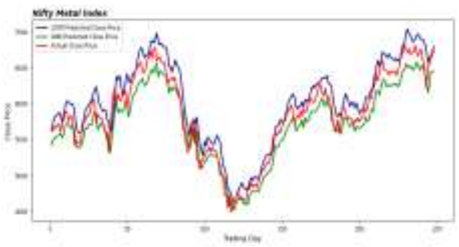
Among the individual stocks, Zydus Lifesciences Ltd. emerged as the best-performing stock in terms of predictive accuracy, with an RMSE of 0.0176 for LSTM and 0.0208 for ANN, while ITC Ltd. had the highest R-squared values of 0.99 using both LSTM and ANN. This indicates that around 99% of the variance in the closing price of ITC Ltd. can be explained using the predictor variables.

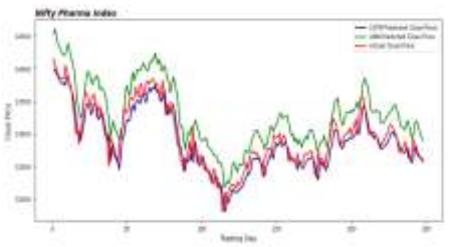
Broad Market Indices

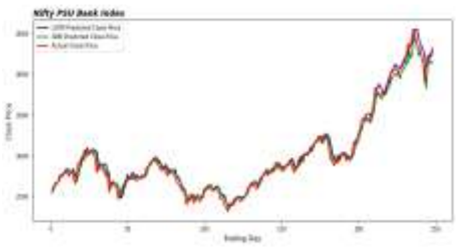





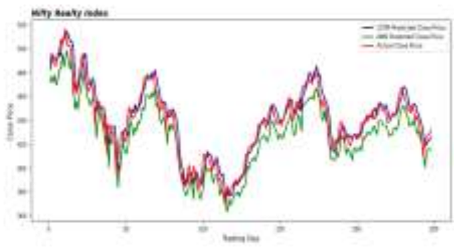
Sectoral Indices


Nifty Metal Index

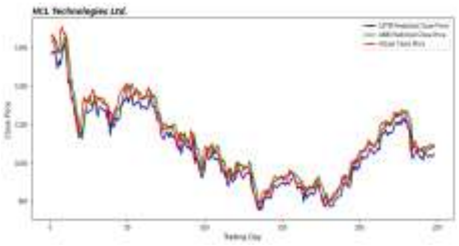

Nifty Pharma Index

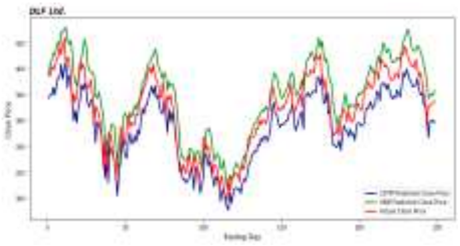

Nifty PSU Bank Index

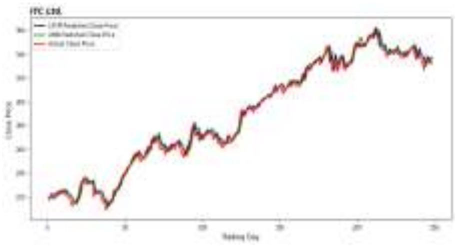

Nifty Realty Index

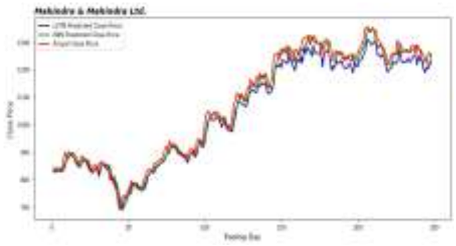

Nifty Services Index

Stocks


HCL Technologies Ltd.


DLF Ltd.


ITC Ltd.


Mahindra & Mahindra Ltd.

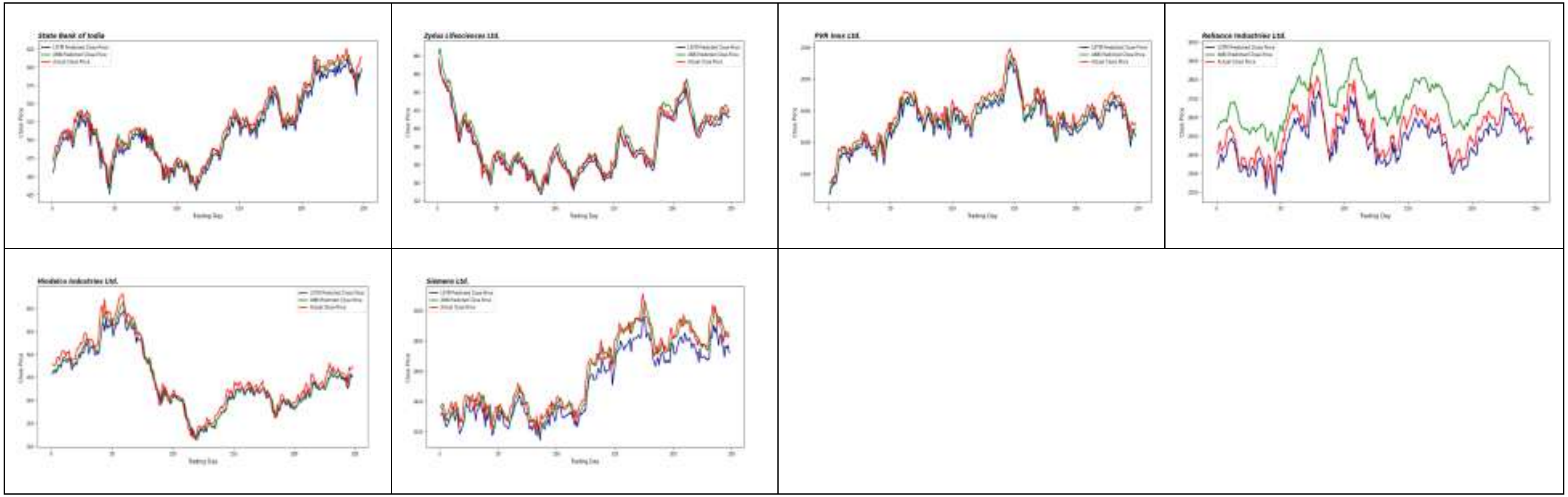


Figure 5.8: Actual and predicted performance of LSTM and ANN during the invasion of Ukraine by Russia from 1st January 2022 – 31st December 2022 across the broad market indices, sectoral indices, and selected stocks. ANN predicted close prices (green), the LSTM predicted close prices (blue), and the daily actual close prices (red).

5.2.2.6 The Hamas Attack on Israel

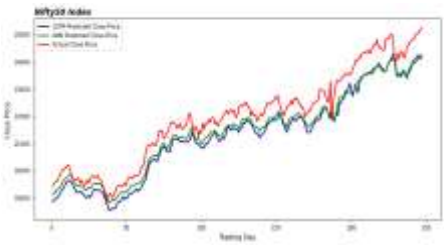
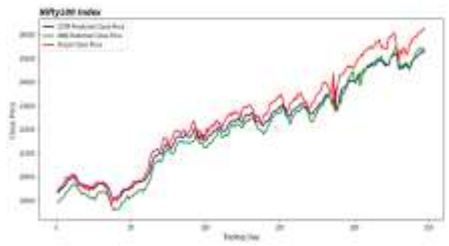
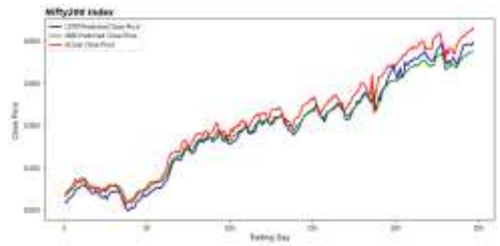
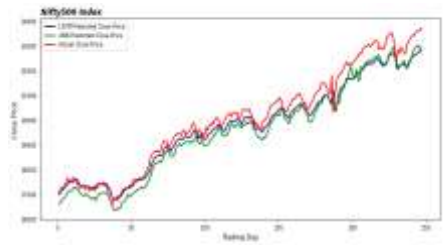
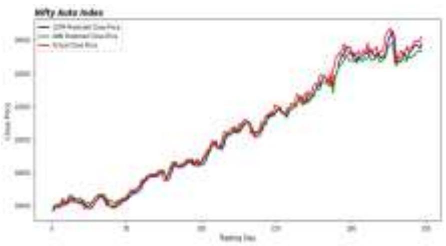
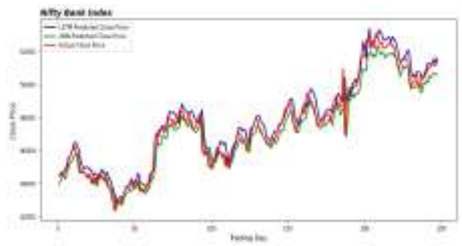
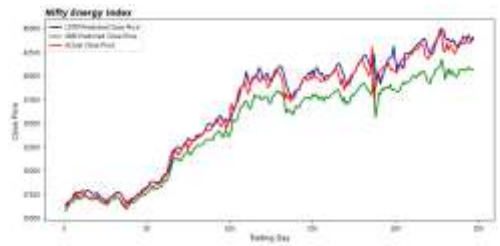
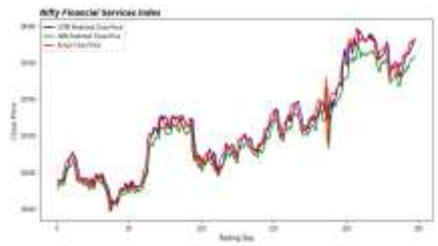
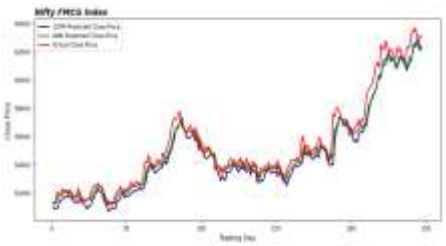
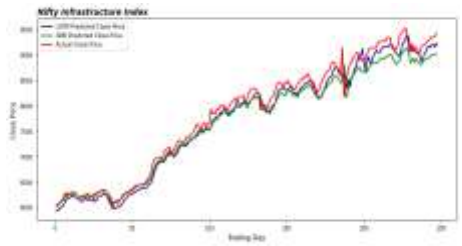
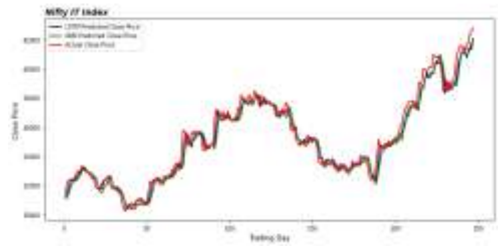
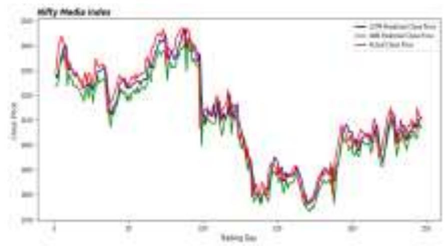
Table 5.6: Performance evaluation of LSTM and ANN during the Hamas attack on Israel

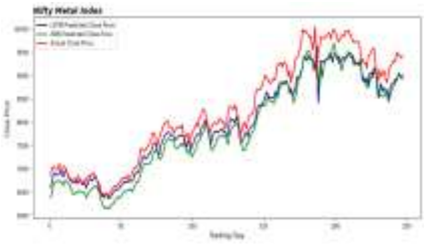
	LSTM			ANN		
Broad Market Indices	**R-Squared**	**RMSE**	**MAE**	**R-Squared**	**RMSE**	**MAE**
Nifty 50	0.8848	0.0437	0.0404	0.8338	0.0469	0.0421
Nifty 100	0.9491	0.0391	0.0314	0.9172	0.0410	0.0370
Nifty 200	0.9557	0.0354	0.0314	0.9508	0.0403	0.0324
Nifty 500	0.9504	0.0416	0.0335	0.9239	0.0544	0.0507
Sectoral Indices						
Auto	0.9902	0.0325	0.0258	0.9870	0.0352	0.0261
Bank	0.9503	0.0191	0.0136	0.9435	0.0220	0.0166
Energy	0.9838	0.0350	0.0249	0.8976	0.0980	0.0835
Financial Services	0.9649	0.0211	0.0162	0.9324	0.0276	0.0223
FMCG	0.9614	0.0202	0.0150	0.9419	0.0204	0.0155
Infrastructure	0.9785	0.0413	0.0346	0.9639	0.0523	0.0418
IT	0.9701	0.0180	0.0131	0.9562	0.0208	0.0150
Media	0.9120	0.0377	0.0311	0.8730	0.0427	0.0367
Metal	0.9440	0.0629	0.0527	0.8970	0.0782	0.0726
Pharma	0.9826	0.0314	0.0264	0.9605	0.0471	0.0360
PSU Bank	0.9777	0.0412	0.0299	0.8453	0.1062	0.0951
Realty	0.9167	0.1189	0.1088	0.8025	0.1932	0.1684
Services	0.9797	0.0219	0.0173	0.9463	0.0313	0.0249
Stocks						
HCL Technologies	0.9527	0.0568	0.0506	0.8237	0.0685	0.0617
DLF Ltd	0.8890	0.1143	0.1047	0.8631	0.1485	0.1358
ITC Ltd	0.9379	0.0178	0.0130	0.9114	0.0197	0.0152
M&M Ltd	0.9830	0.0419	0.0305	0.9761	0.0630	0.0442
SBI	0.9616	0.0409	0.0303	0.9281	0.0637	0.0544
Zydus Lifesciences Ltd	0.9974	0.0526	0.0379	0.9605	0.0893	0.0655
P V R Inox Ltd.	0.9198	0.0200	0.0155	0.8477	0.0374	0.0333
Hindalco Industries Ltd.	0.9671	0.0251	0.0180	0.9482	0.0388	0.0326
Reliance Industries Ltd.	0.9718	0.0243	0.0181	0.9656	0.0273	0.0210
Siemens Ltd.	0.9859	0.0619	0.0448	0.9674	0.1057	0.0725

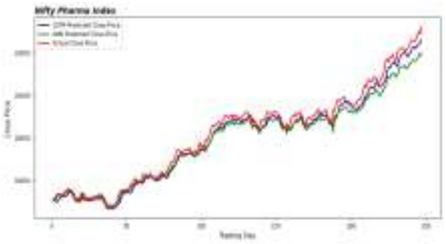
The performance evaluation metrics for LSTM and ANN during the Hamas attack on Israel are displayed in Table 5.6. The R-squared values across the indices and stocks are above 0.80 using both models. The highest R-squared values were achieved in predicting the Nifty Auto index, with 0.99 using LSTM and 0.98 using ANN. This signals that the predictor variables, which are the low and high close prices, are a good fit to the model.

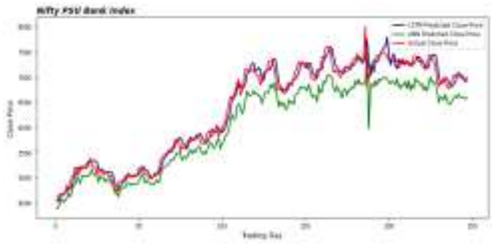
The predictive performance of the LSTM and ANN models shows only a slight difference in terms of RMSE for the broad market indices, with LSTM consistently outperforming ANN overall. Significant differences were observed in the RMSE of Nifty Energy, Nifty PSU Bank, and Nifty Realty index, with Nifty Realty emerging as the most challenging sector for both models. It recorded the highest RMSE and MAE values amongst all sectors. The highest RMSE (LSTM: 0.1189, ANN: 0.1932) and MAE (LSTM: 0.1088, ANN: 0.1684) on running both models were seen in the case of the Nifty Realty index.

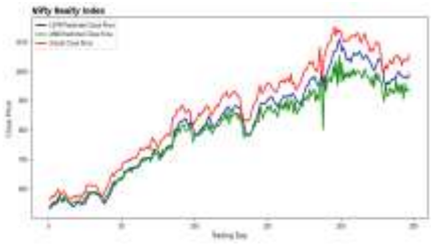
DLF Ltd. and Siemens Ltd. exhibited the highest RMSE using both models at the stock level. Nevertheless, LSTM consistently outperformed ANN. The RMSE for DLF Ltd. was 0.1143, and for Siemens Ltd. 0.0619.

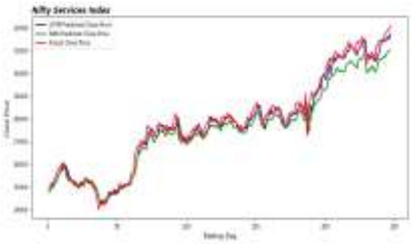
<table>
<tr><th colspan="4">Broad Market Indices</th></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr><th colspan="4">Sectoral Indices</th></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
</table>


Nifty Metal Index

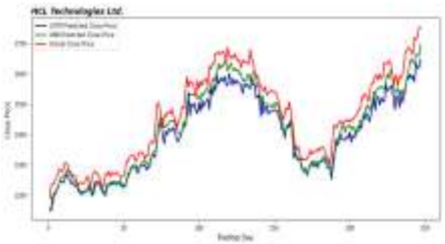

Nifty Pharma Index

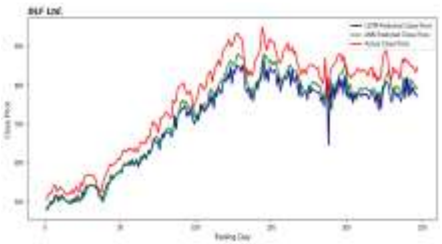

Nifty PSU Bank Index

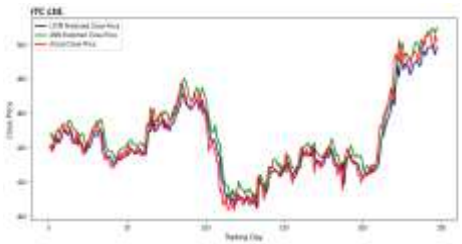

Nifty Realty Index

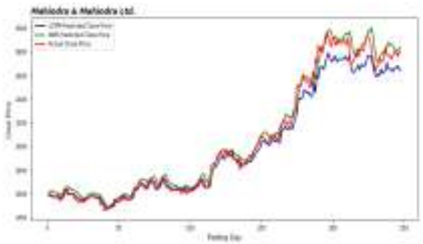

Nifty Services Index

Stocks


HCL Technologies Ltd.


DLF Ltd.


ITC Ltd.


Mahindra & Mahindra Ltd.

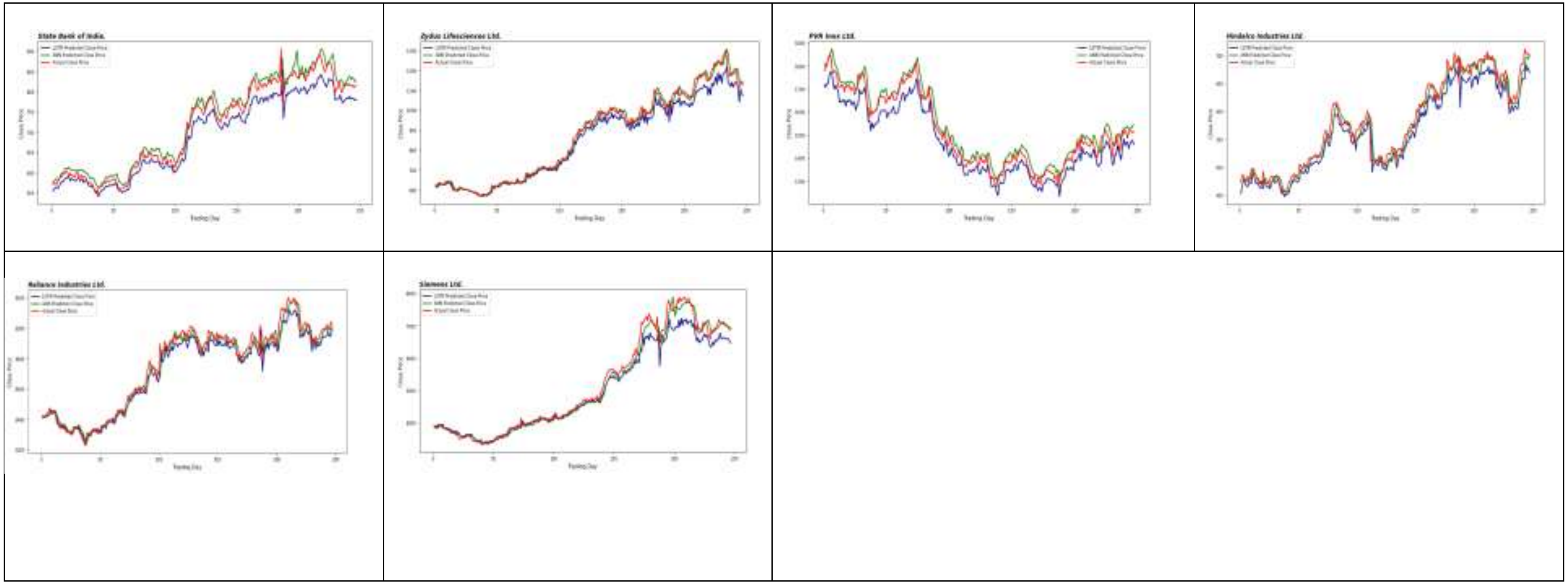


Figure 5.9: Actual and predicted performance of LSTM and ANN during the Hamas attack on Israel from 1st September 2023 – 31st August 2024 across the broad market indices, sectoral indices, and selected stocks. ANN predicted close prices (green), the LSTM predicted close prices (blue), and the daily actual close prices (red).

5.2.2.7 The General Elections 2024

Table 5.7: Performance evaluation of LSTM and ANN during the general elections 2024

	LSTM			ANN		
Broad Market Indices	**R-Squared**	**RMSE**	**MAE**	**R-Squared**	**RMSE**	**MAE**
Nifty 50	0.9855	0.0470	0.0364	0.9678	0.0544	0.0476
Nifty 100	0.8033	0.0608	0.0545	0.5876	0.0846	0.0752
Nifty 200	0.8697	0.0434	0.0330	0.7803	0.0628	0.0570
Nifty 500	0.7733	0.0624	0.0539	0.7191	0.0640	0.0571
Sectoral Indices						
Auto	0.8563	0.0471	0.0404	0.7651	0.0479	0.0377
Bank	0.8156	0.0414	0.0295	0.6548	0.0526	0.0418
Energy	0.6922	0.0449	0.0342	0.5354	0.1303	0.1239
Financial Services	0.9137	0.0418	0.0300	0.8852	0.0485	0.0360
FMCG	0.9404	0.0403	0.0317	0.8729	0.0497	0.0406
Infrastructure	0.6328	0.0483	0.0418	0.6001	0.0762	0.0670
IT	0.9620	0.0593	0.0475	0.9086	0.0953	0.0843
Media	0.8741	0.0438	0.0349	0.6619	0.0609	0.0490
Metal	0.5981	0.0455	0.0355	0.4129	0.0569	0.0453
Pharma	0.9501	0.0499	0.0417	0.8729	0.0652	0.0570
PSU Bank	0.7925	0.0270	0.0172	0.7701	0.0844	0.0786
Realty	0.6338	0.0691	0.0618	0.5829	0.0859	0.0760
Services	0.9447	0.0470	0.0353	0.8941	0.0639	0.0545
Stocks						
HCL Technologies	0.9667	0.0397	0.0304	0.9632	0.0418	0.0345
DLF Ltd	0.6081	0.0379	0.0301	0.4792	0.0767	0.0691
ITC Ltd	0.9460	0.0332	0.0265	0.9284	0.0349	0.0277
M&M Ltd	0.8511	0.0574	0.0440	0.7402	0.0893	0.0783
SBI	0.5311	0.0580	0.0438	0.3962	0.0783	0.0652
Zydus Lifesciences Ltd	0.8503	0.0428	0.0330	0.8435	0.0511	0.0409
P V R Inox Ltd.	0.9431	0.0297	0.0235	0.9166	0.0329	0.0261
Hindalco Industries Ltd.	0.7321	0.0518	0.0404	0.5799	0.0767	0.0676
Reliance Industries Ltd.	0.7945	0.1698	0.0600	0.7468	0.1858	0.0836
Siemens Ltd.	0.7797	0.0583	0.0450	0.6278	0.0711	0.0567

Table 5.7 presents the performance evaluation of LSTM and ANN during the general elections of 2024 across the broad market, sectoral indices, and individual stocks.

The RMSE values for the broad market indices using LSTM ranged between 0.0434 and 0.0624, and using ANN, they ranged between 0.0544 and 0.0846.

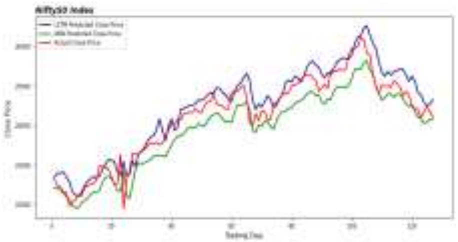
Amongst the sectoral indices, the largest disparity was observed in the energy sector, with LSTM achieving an MAE of 0.0342, significantly lower than ANN's 0.1239, followed by Nifty PSU Bank (LSTM: 0.0172, ANN: 0.0786) and Nifty IT (LSTM: 0.0475, ANN: 0.0843).

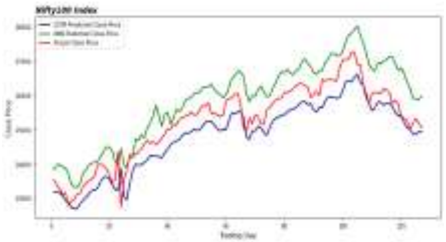
In stock-level predictions, LSTM consistently delivers lower RMSE and MAE values. Significant prediction differences using LSTM and ANN were witnessed in the RMSE of DLF

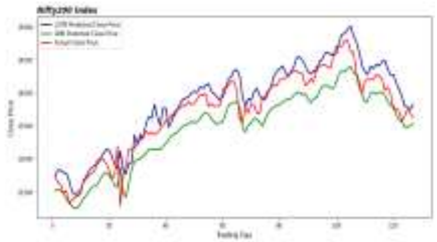
Ltd. and M&M Ltd. The highest MAE among all stocks was seen in Reliance Industries Ltd. (LSTM: 0.0600, ANN: 0.0836).

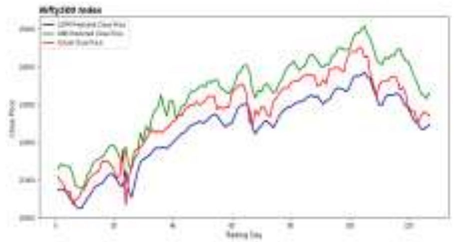
The comparative analysis of LSTM and ANN models highlights LSTM's consistent superiority in prediction accuracy across broad market indices, sectoral indices, and selected stocks, as reflected in lower RMSE and MAE values and higher R-squared values.

Broad Market Indices

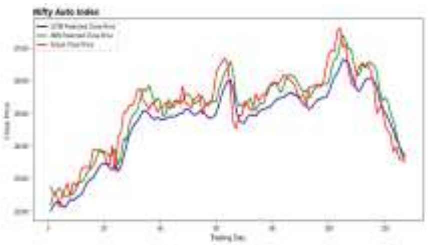


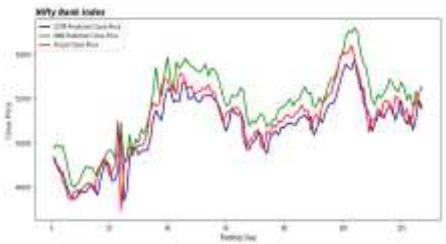


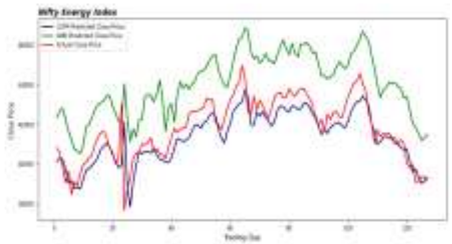


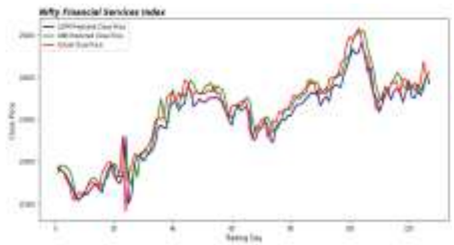


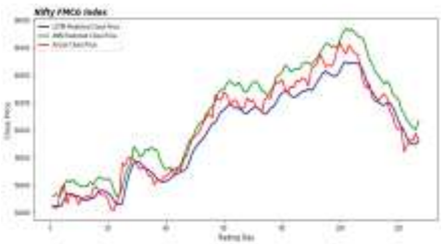
Sectoral Indices

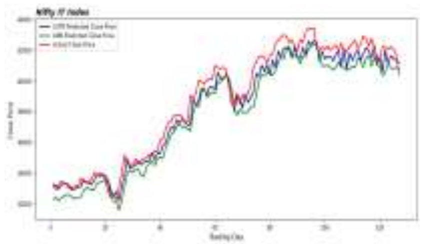


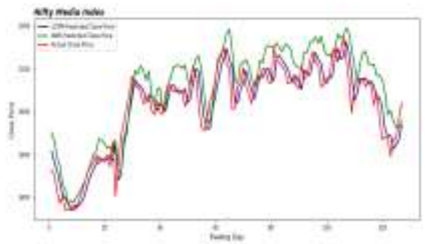


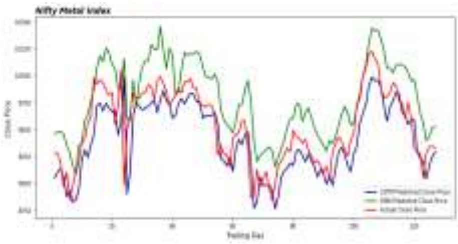


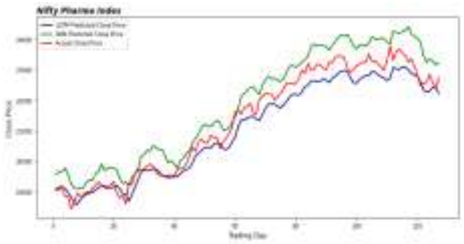


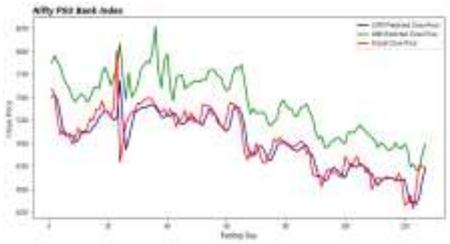


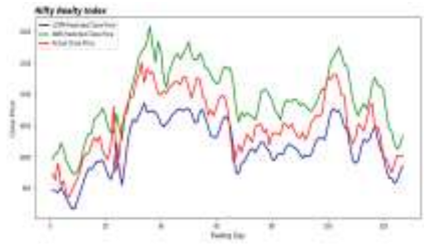


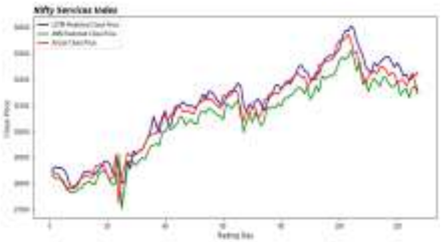



Nifty Metal Index

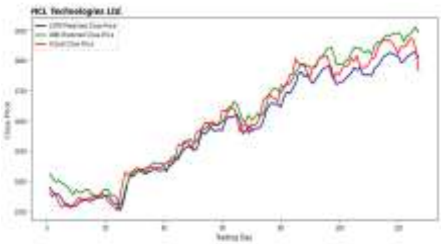

Nifty Pharma Index

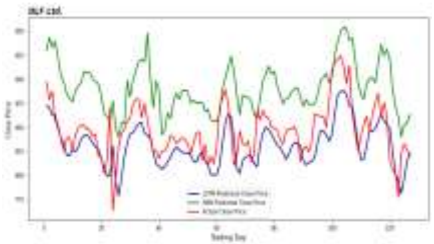

Nifty PSU Bank Index

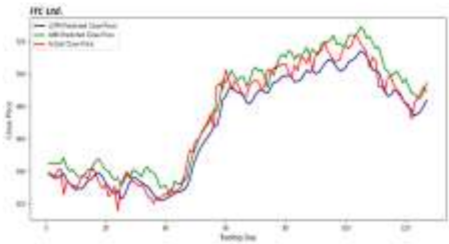

Nifty Realty Index

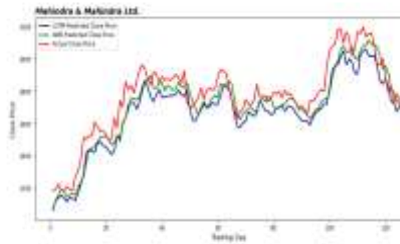

Nifty Services Index

Stocks


HCL Technologies Ltd.


DLF Ltd.


ITC Ltd.


Mahindra & Mahindra Ltd.

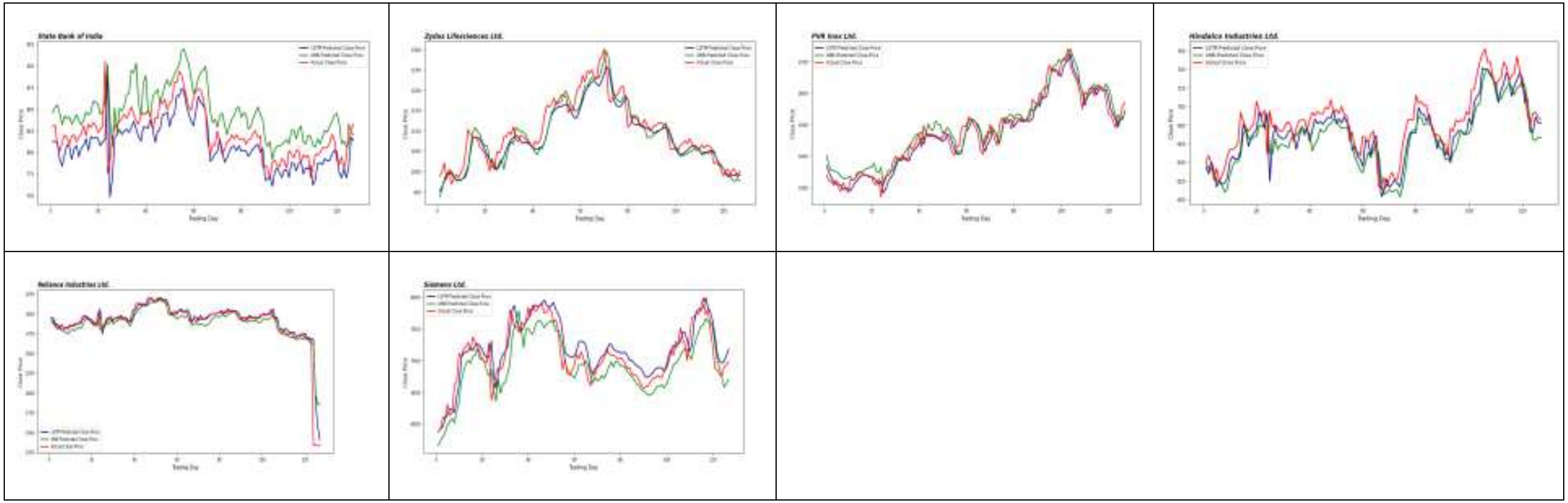


Figure 5.10: Actual and predicted performance of LSTM and ANN during the general elections 2024 from 1st May 2024 – 31st October 2024 across the broad market indices, sectoral indices, and selected stocks. ANN predicted close prices (green), the LSTM predicted close prices (blue), and the daily actual close prices (red).

Chapter 6

Portfolio Optimization

6.1 Introduction

This section details the results and discussions on portfolio formation and fund allocation using machine learning. Firstly, the number of clusters to be formed is decided by the highest silhouette score obtained using K-means clustering. The cluster formation is visualized on a scatter plot. One cluster is selected based on the centroid details such as returns, standard deviation, and the silhouette score of each cluster. The correlation of stock returns is visualized to understand the nature of the selected stocks.

Secondly, from the stocks in the selected cluster, three portfolios are created with different weights, namely the EWP, MSRP, and MVP. The weights allocated to MSRP and MVP are simulated using Monte Carlo. A dummy investment plan of Rs. 1,00,000 is created, and funds are allocated per the selected weights for each portfolio. The performance of these portfolios is compared with that of the Nifty benchmark indices, Nifty50 and Nifty100, in terms of return on investment for that period. Moreover, to evaluate the performance of the created portfolios throughout the selected period, we present a daily cumulative return chart and compare it with the benchmark indices. The portfolios are constructed for several periods covering seven events. The following section presents and discusses the event-wise results.

6.2 Results and Discussion

6.2.1 Demonetization

Table 6.1 highlights that the highest Silhouette score of 0.394 was obtained with five clusters, indicating an optimal clustering structure. The formation of these five clusters is visualized in the scatter plot in Figure 6.1, while Table 6.2 provides the centroid details for each cluster. Notably, Cluster C4 shows the highest Silhouette score among the five clusters, confirming that the stocks are well-clustered. Additionally, this cluster's centroid represents the highest mean returns with a moderate risk level, suggesting a favorable risk-return trade-off.

Table 6.1: Silhouette Score for various clusters on the demonetization dataset

Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score
2	0.314	12	0.287	22	0.255
3	0.329	13	0.266	23	0.242
4	0.370	14	0.276	24	0.256
5	**0.394**	15	0.277	25	0.231
6	0.337	16	0.245	26	0.242
7	0.331	17	0.266	27	0.227
8	0.335	18	0.256	28	0.231
9	0.302	19	0.263	29	0.235
10	0.266	20	0.258	30	0.249
11	0.277	21	0.260		

Furthermore, the scatter plot in Figure 6.1 reveals several overlaps among stocks in Cluster 1, 2, and 5, suggesting potential similarities in their characteristics. Apart from conventional return-based metrics, the model has clustered stocks while considering their validation ratios.

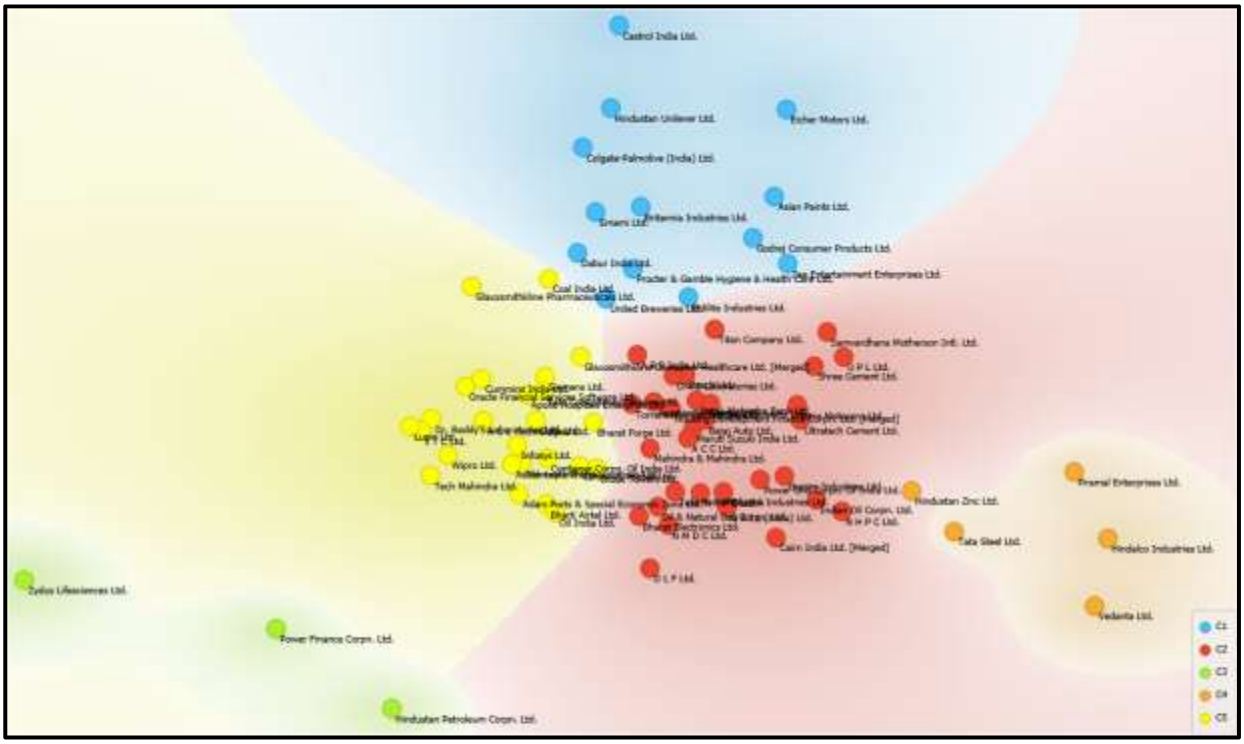


Figure 6.1: Cluster formation on the demonetization dataset

In contrast, Cluster C3 comprises stocks with low returns and high standard deviation, signifying higher volatility and unfavorable investment prospects. These findings provide insights into stock behavior during the demonetization period, helping investors navigate market uncertainties effectively.

Table 6.2: Centroid details for the demonetization dataset

Cluster	Silhouette	Daily Average Returns	Standard Deviation
C1	0.592828	0.590047	0.108678
C2	0.608239	0.639764	0.154326
C3	0.601011	0.235761	0.84702
C4	**0.650538**	**0.930647**	**0.35486**
C5	0.636242	0.438635	0.143682

The correlation heatmap in Figure 6.2 provides insights into the relationships between the stock returns of five companies that belong to cluster C4. These include Hindalco Industries Ltd., Hindustan Zinc Ltd., Piramal Enterprises Ltd., Tata Steel Ltd., and Vedanta Ltd. A higher correlation value indicates that the stocks move in tandem, whereas a lower value, closer to 0, indicates a weaker relationship. Hindalco Industries Ltd., Hindustan Zinc Ltd., Tata Steel Ltd., and Vedanta Ltd., operating in the metal and mining sector, have higher correlations, likely influenced by common market industry factors. In contrast, Piramal Enterprises Ltd. operates in the diversified financial services sector, which behaves differently from the metals sector. This is apparent in its low correlation values.

Figure 6.2: Correlation of stock returns on the demonetization dataset

Table 6.3: Fund allocation for selected stocks on the demonetization dataset

Selected Shares	Weight (%)	Initial investment as on 01st October 2016 (Rs.)	Number of Shares	Investment value as on 30th September 2017 (Rs.)	ROI
Maximum Sharpe Ratio Portfolio					
Hindalco Industries Ltd.	24.29	24290	157.62	37924.56	56.13%
Hindustan Zinc Ltd.	20.78	20780	81.67	24165.07	16.29%
Piramal Enterprises Ltd.	50.9	50900	27.43	69812.27	37.16%
Tata Steel Ltd.	3.38	3380	89.66	5760.345	70.42%
Vedanta Ltd.	0.65	650	3.64	1142.727	75.80%

	100.00	**100000**		**138804.9**	**38.80%**
Minimum Volatility Portfolio					
Hindalco Industries Ltd.	5.03	5030	32.64	7853.459	56.13%
Hindustan Zinc Ltd.	34.18	34180	134.33	39747.93	16.29%
Piramal Enterprises Ltd.	40.84	40840	22.01	56014.41	37.16%
Tata Steel Ltd.	19.69	19690	522.28	33556.56	70.42%
Vedanta Ltd.	0.26	260	1.45	457.0909	75.80%
	100.00	**100000**		**137629.5**	**37.63%**
Equally Weighted Portfolio					
Hindalco Industries Ltd.	20	20000	129.79	31226.48	56.13%
Hindustan Zinc Ltd.	20	20000	78.60	23258.01	16.29%
Piramal Enterprises Ltd.	20	20000	10.78	27431.15	37.16%
Tata Steel Ltd.	20	20000	530.50	34084.88	70.42%
Vedanta Ltd.	20	20000	111.89	35160.84	75.80%
	100.00	**100000**		**151161.4**	**51.16%**
Nifty Benchmark Indices					
Nifty 50					**12.02%**
Nifty 100					**12.79%**

Table 6.3 displays the funds allocated as per the weights simulated by Monte Carlo for the MSRP and MVP, as well as the stocks in the EWP, which have been allocated an equal amount from the total investment. The portfolios are held for one year, from 01st October 2016 to 30th September 2017. We calculate the number of shares purchased on 01st October 2016, based on the total amount allocated for each stock. Lastly, the investment value is calculated by considering the same number of shares held until the 30th September 2017. The ROI for the MSRP, MVP, and EWP are 38.80%, 37.63%, and 51.16%, respectively. The constructed portfolios outperformed the Nifty50 and Nifty100 indices, which had an ROI of 12.02% and 12.79%, respectively, for the same period. The cumulative returns graph in Figure 6.3 indicates that the MSRP and the MVP performed closely to the benchmark indices over the first three months. The Equally Weighted Portfolio outperformed the benchmark indices daily throughout the period. On the day of the event, i.e., the announcement of demonetization, the benchmark indices continued to show cumulative negative returns, while the EWP continued to show a positive cumulative return.

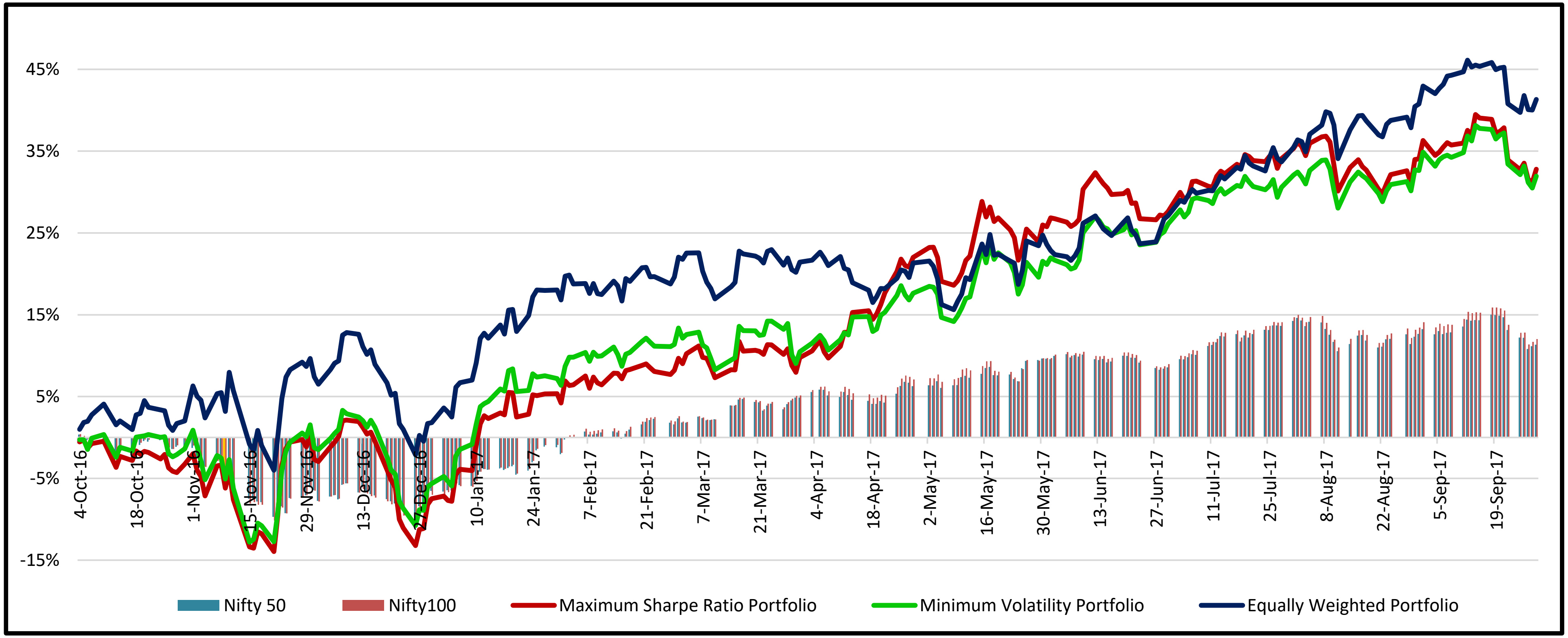


Figure 6.3: Cumulative returns for the period 01st October 2016 to 30th September 2017

6.2.2 Implementation of the GST Regime

Table 6.4: Silhouette score for various clusters on the GST dataset

Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score
2	0.338	12	0.330	22	0.326
3	0.357	13	0.325	23	0.306
4	0.341	14	0.322	24	0.310
5	0.345	15	0.324	25	0.318
6	0.357	16	0.324	26	0.299
7	**0.369**	17	0.338	27	0.292
8	0.357	18	0.339	28	0.279
9	0.329	19	0.343	29	0.292
10	0.332	20	0.314	30	0.271
11	0.347	21	0.327		

Table 6.4 shows that the highest silhouette score was achieved with seven clusters, indicating an optimal clustering structure. Figure 6.4 illustrates the cluster formation for stock classification using seven clusters. Additionally, Table 6.5 reveals that cluster C4 exhibits the highest mean returns while maintaining a moderate level of risk, making it a potentially attractive option for portfolio selection.

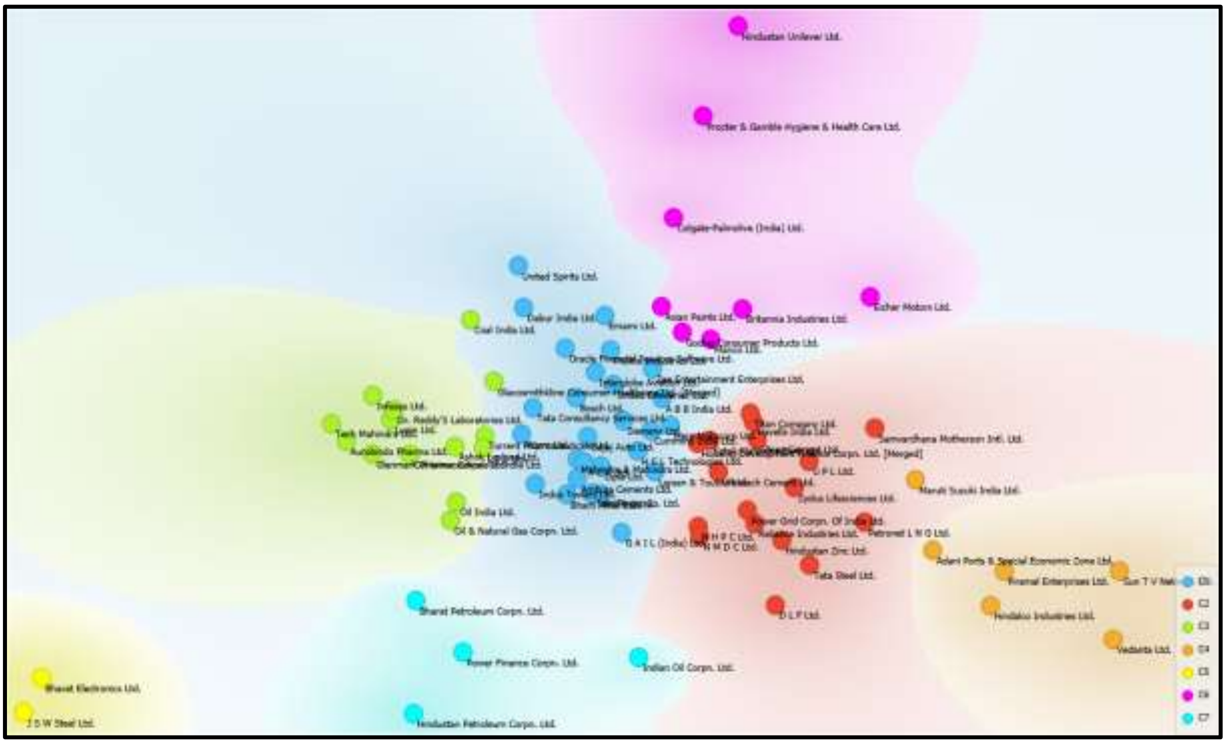


Figure 6.4: Cluster formation on the GST dataset

Table 6.5: Centroid details for the GST dataset

Cluster	Silhouette	Daily Average Returns	Standard Deviation
C1	0.606647	0.453571	0.116461
C2	0.608032	0.635377	0.176263
C3	0.624552	0.277583	0.171805
C4	**0.626572**	**0.883304**	**0.261959**
C5	0.693914	0.014179	0.987703
C6	0.572242	0.560091	0.097929
C7	0.597273	0.427974	0.595801

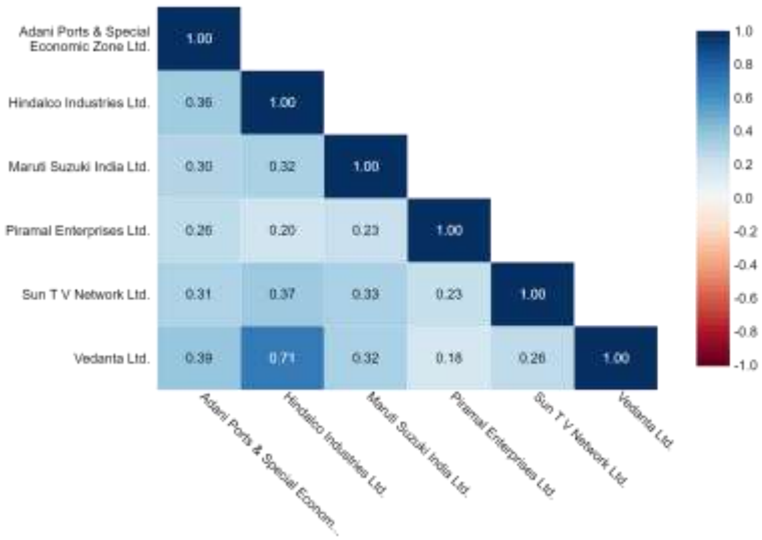


Figure 6.5: Correlation of stock returns on the GST dataset

Cluster C4 comprises six stocks: Adani Ports & Special Economic Zone Ltd., Hindalco Industries Ltd., Maruti Suzuki India Ltd., Piramal Enterprises Ltd., Sun T V Network Ltd., and Vedanta Ltd. All stock returns have a low positive correlation (below 0.4) except for Vedanta Ltd. and Hindalco Industries Ltd. (0.71), belonging to the metal sector. Adani Ports & Special Economic Zone Ltd belongs to the infrastructure sector, Maruti Suzuki India Ltd. belongs to the auto sector, Piramal Enterprises Ltd belongs to the financial services sector, and Sun T V Network Ltd belongs to the media sector.

Table 6.6: Fund allocation for selected stocks on the GST dataset

Selected Shares	Weight (%)	Initial investment as on 01st June 2017 (Rs.)	Number of Shares	Investment value as on 31st May 2018 (Rs.)	ROI
Maximum Sharpe Ratio Portfolio					
Adani Ports & Special Economic Zone Ltd.	7.12	7120	20.39	7995.7274	12.30%
Hindalco Industries Ltd.	0.77	770	3.90	914.24081	18.73%
Maruti Suzuki India Ltd.	46.44	46440	6.50	55476.39	19.46%
Piramal Enterprises Ltd.	18.49	18490	6.57	15568.574	-15.80%
Sun T V Network Ltd.	11.39	11390	14.17	13004.186	14.17%

Vedanta Ltd.	15.79	15790	68.28	16954.192	7.37%
	100	**100000**		**109913.31**	**9.91%**
Minimum Volatility Portfolio					
Adani Ports & Special Economic Zone Ltd.	16.96	16960	48.57	19046.002	12.30%
Hindalco Industries Ltd.	5.36	5360	27.17	6364.0659	18.73%
Maruti Suzuki India Ltd.	54.98	54980	7.69	65678.121	19.46%
Piramal Enterprises Ltd.	14.55	14550	5.17	12251.095	-15.80%
Sun T V Network Ltd.	5.91	5910	7.35	6747.5625	14.17%
Vedanta Ltd.	2.24	2240	9.69	2405.1546	7.37%
	100	**100000**		**112492**	**12.49%**
Equally Weighted Portfolio					
Adani Ports & Special Economic Zone Ltd.	16.67	16670	47.74	18720.334	12.30%
Hindalco Industries Ltd.	16.67	16670	84.51	19792.72	18.73%
Maruti Suzuki India Ltd.	16.67	16670	2.33	19913.683	19.46%
Piramal Enterprises Ltd.	16.67	16670	5.92	14036.135	-15.80%
Sun T V Network Ltd.	16.66	16660	20.73	19021.048	14.17%
Vedanta Ltd.	16.66	16660	72.04	17888.337	7.37%
	100	**100000**		**109372.26**	**9.37%**
Nifty Benchmark Indices					
Nifty 50					**11.65%**
Nifty 100					**11.66%**

Table 6.6 shows that relying solely on the ROI indicates that the benchmark indices outperform MSRP and EWP by a small margin. However, considering the overall performance throughout the investment holding period, the three proposed portfolios consistently outperform the benchmark indices, as seen in Figure 6.6. The proposed portfolios remain positive even when the benchmark indices show cumulative negative returns. This is especially seen during the week prior to the implementation of the GST regime, where Nifty50 and Nifty100 are on cumulative negative returns, while all the proposed portfolios are in positive returns. However, during the last week of the study period, the EWP and the MSRP's returns coincided with that of the benchmark indices, while the MVP (12.91%) ends at 2.62% higher than Nifty100 (10.29%) and 2.86% higher than Nifty50 (10.05%).

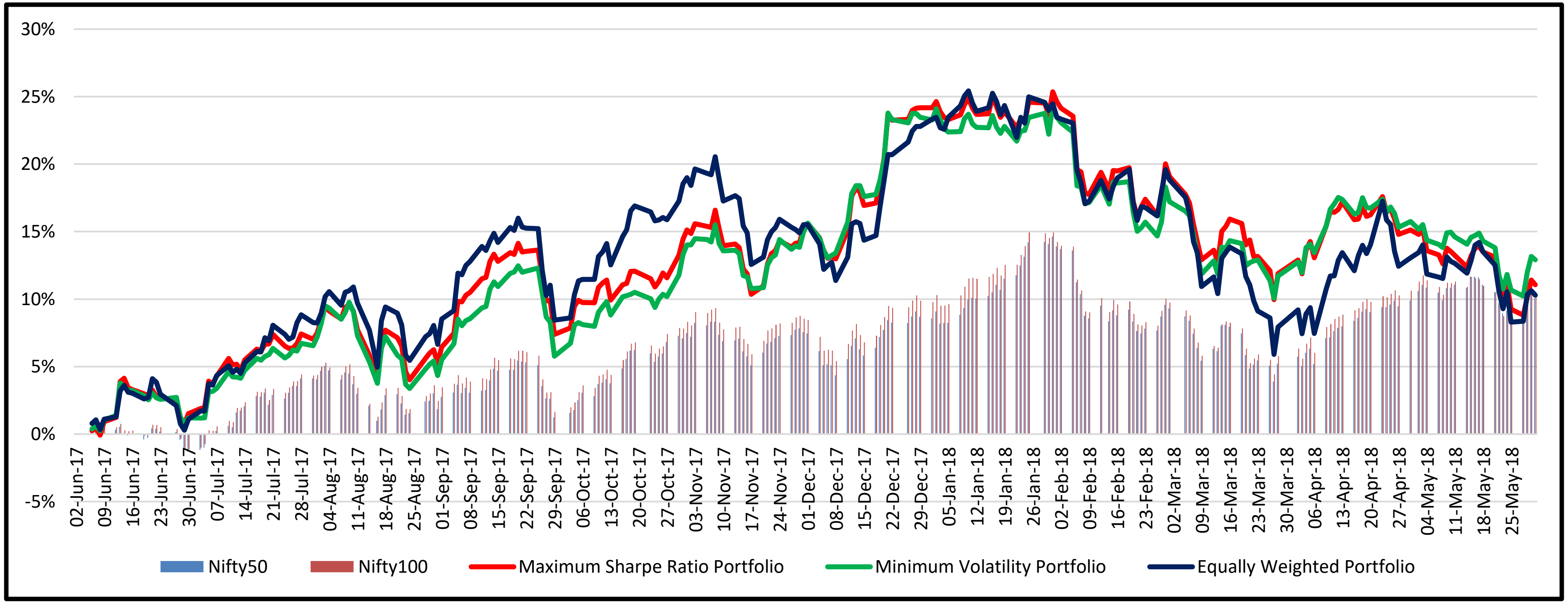


Figure 6.6: Cumulative returns for the period 01st June 2017 to 31st May 2018

6.2.3 The General Elections 2019

Table 6.7: Silhouette score for various clusters on the general elections 2019 dataset

Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score
2	0.371	12	0.288	22	0.272
3	**0.395**	13	0.296	23	0.269
4	0.309	14	0.271	24	0.254
5	0.315	15	0.271	25	0.263
6	0.334	16	0.293	26	0.245
7	0.329	17	0.285	27	0.223
8	0.338	18	0.278	28	0.251
9	0.318	19	0.289	29	0.236
10	0.274	20	0.279	30	0.237
11	0.312	21	0.274		

The highest silhouette score of 0.395 was obtained when creating three clusters, as shown in Table 6.7. The stock classification of the three clusters is displayed in Figure 6.7. As seen in Table 6.8, the mean returns of Cluster C2 are the highest, while its risk is the lowest. Hence, we selected cluster C2 as our investment portfolio.

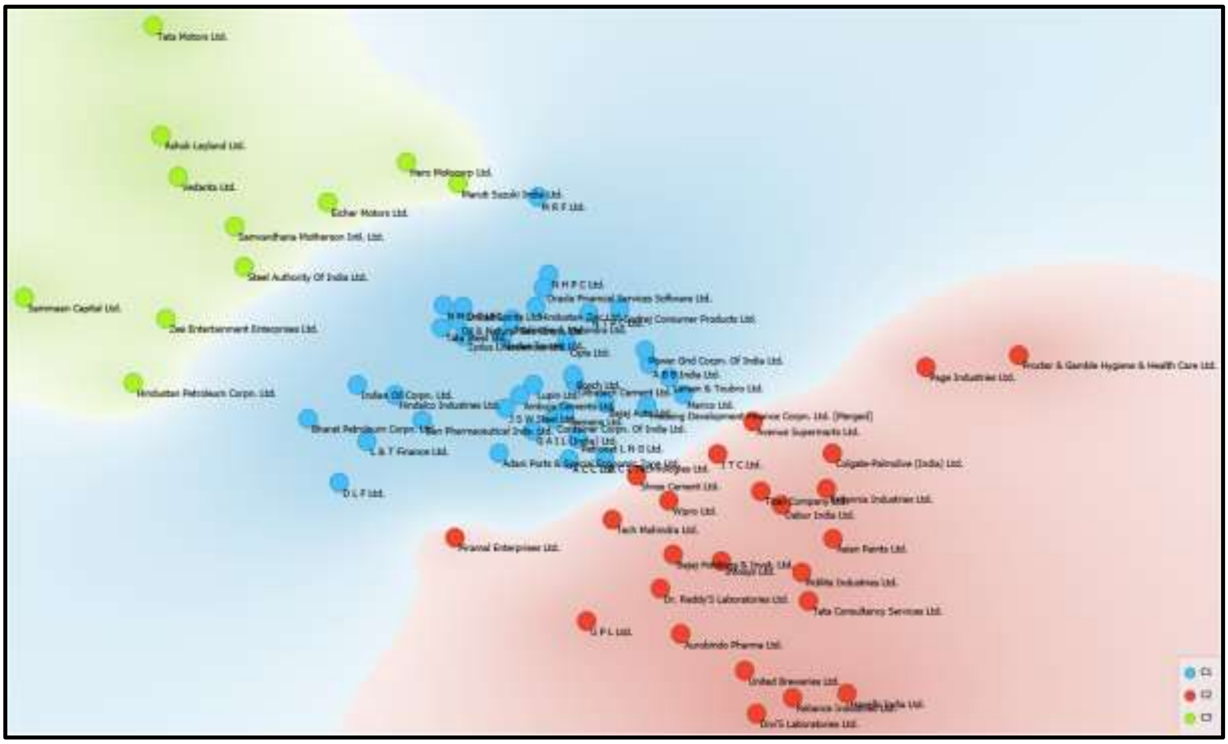


Figure 6.7: Cluster formation on the general elections 2019 dataset

Table 6.8: Centroid details for the general election 2019 dataset

Cluster	Silhouette	Daily Average Returns	Standard Deviation
C1	0.639229	0.549981	0.268196
C2	**0.588889**	**0.795754**	**0.231817**
C3	0.603887	0.277002	0.575999

The correlation of the 24 stocks in Cluster 2 is visualized in Figure 6.8. The figure depicts a low correlation among the selected stocks.

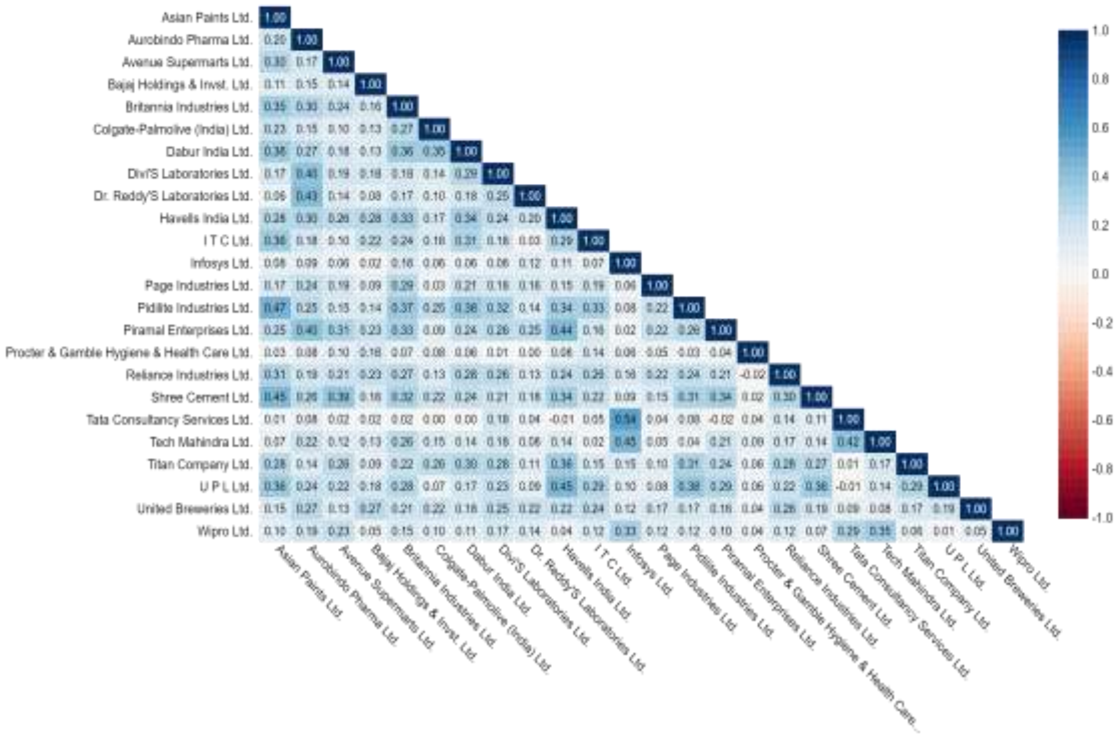


Figure 6.8: Correlation of stock returns on the general elections 2019 dataset

Table 6.9: Fund allocation for selected stocks on the general elections 2019 dataset

Selected Shares	Weight (%)	Initial investment as on 01st April 2019 (Rs.)	Number of Shares	Investment value as on 31st March 2020 (Rs.)	ROI
Maximum Sharpe Ratio Portfolio					
Asian Paints Ltd.	5.68	5680	3.81	6341.129	11.64%
Aurobindo Pharma Ltd.	4.34	4340	5.48	2262.978	-47.86%
Avenue Supermarts Ltd.	1.7	1700	1.14	2490.874	46.52%
Bajaj Holdings & Invst. Ltd.	8.23	8230	2.46	4415.693	-46.35%
Britannia Industries Ltd.	9.89	9890	3.24	8719.68	-11.83%
Colgate-Palmolive (India) Ltd.	4.97	4970	3.95	4948.275	-0.44%
Dabur India Ltd.	1.84	1840	4.59	2065.325	12.25%
Divi's Laboratories Ltd.	4.95	4950	2.87	5713.339	15.42%
Dr. Reddy's Laboratories Ltd.	5.34	5340	1.91	5962.79	11.66%
Havells India Ltd.	7.55	7550	9.73	4676.591	-38.06%
I T C Ltd.	1.54	1540	5.18	889.5475	-42.24%
Infosys Ltd.	9.03	9030	11.96	7671.494	-15.04%

Page Industries Ltd.	0.18	180	0.01	119.5932	-33.56%
Pidilite Industries Ltd.	0.87	870	0.69	942.0572	8.28%
Piramal Enterprises Ltd.	0.16	160	0.06	55.80739	-65.12%
Procter & Gamble Hygiene & Health Care Ltd.	9.02	9020	0.84	8583.803	-4.84%
Reliance Industries Ltd.	7.34	7340	5.27	5873.424	-19.98%
Shree Cement Ltd.	0.94	940	0.05	877.4692	-6.65%
Tata Consultancy Services Ltd.	5.81	5810	2.86	5222.18	-10.12%
Tech Mahindra Ltd.	1.7	1700	2.16	1223.87	-28.01%
Titan Company Ltd.	0.86	860	0.77	716.7242	-16.66%
U P L Ltd.	3.51	3510	3.76	1226.603	-65.05%
United Breweries Ltd.	1.67	1670	1.18	1086.05	-34.97%
Wipro Ltd.	2.88	2880	11.01	2165.091	-24.82%
	100.00	**100000**		**84250.39**	**-15.75%**
Minimum Volatility Portfolio					
Asian Paints Ltd.	0.79	790	0.53	881.9528	11.64%
Aurobindo Pharma Ltd.	0.68	680	0.86	354.5681	-47.86%
Avenue Supermarts Ltd.	1.4	1400	0.94	2051.308	46.52%
Bajaj Holdings & Invst. Ltd.	6.45	6450	1.92	3460.659	-46.35%
Britannia Industries Ltd.	2.41	2410	0.79	2124.816	-11.83%
Colgate-Palmolive (India) Ltd.	9.02	9020	7.17	8980.571	-0.44%
Dabur India Ltd.	1.49	1490	3.72	1672.464	12.25%
Divi's Laboratories Ltd.	0.96	960	0.56	1108.042	15.42%
Dr. Reddy's Laboratories Ltd.	4.78	4780	1.71	5337.479	11.66%
Havells India Ltd.	7.75	7750	9.99	4800.474	-38.06%
I T C Ltd.	7.88	7880	26.51	4551.711	-42.24%
Infosys Ltd.	1.3	1300	1.72	1104.423	-15.04%
Page Industries Ltd.	0.97	970	0.04	644.4745	-33.56%
Pidilite Industries Ltd.	0.01	10	0.01	10.82824	8.28%
Piramal Enterprises Ltd.	1.18	1180	0.44	411.5795	-65.12%
Procter & Gamble Hygiene & Health Care Ltd.	10.19	10190	0.94	9697.223	-4.84%
Reliance Industries Ltd.	3.59	3590	2.58	2872.696	-19.98%
Shree Cement Ltd.	1.7	1700	0.09	1586.912	-6.65%
Tata Consultancy Services Ltd.	8.96	8960	4.41	8053.482	-10.12%
Tech Mahindra Ltd.	7.91	7910	10.07	5694.596	-28.01%
Titan Company Ltd.	10.67	10670	9.52	8892.381	-16.66%
U P L Ltd.	0.48	480	0.51	167.7406	-65.05%

United Breweries Ltd.	1.55	1550	1.10	1008.01	-34.97%
Wipro Ltd.	7.88	7880	30.12	5923.929	-24.82%
	100	**100000**		**81392.32**	**-18.61%**
Equally Weighted Portfolio					
Asian Paints Ltd.	4.17	4170	2.79	4655.371	11.64%
Aurobindo Pharma Ltd.	4.17	4170	5.26	2174.336	-47.86%
Avenue Supermarts Ltd.	4.17	4170	2.79	6109.967	46.52%
Bajaj Holdings & Invst. Ltd.	4.17	4170	1.24	2237.356	-46.35%
Britannia Industries Ltd.	4.17	4170	1.37	3676.549	-11.83%
Colgate-Palmolive (India) Ltd.	4.17	4170	3.31	4151.772	-0.44%
Dabur India Ltd.	4.17	4170	10.40	4680.655	12.25%
Divi's Laboratories Ltd.	4.17	4170	2.42	4813.055	15.42%
Dr. Reddy's Laboratories Ltd.	4.17	4170	1.49	4656.336	11.66%
Havells India Ltd.	4.17	4170	5.38	2582.965	-38.06%
I T C Ltd.	4.17	4170	14.03	2408.71	-42.24%
Infosys Ltd.	4.17	4170	5.52	3542.65	-15.04%
Page Industries Ltd.	4.17	4170	0.16	2770.576	-33.56%
Pidilite Industries Ltd.	4.17	4170	3.33	4515.378	8.28%
Piramal Enterprises Ltd.	4.17	4170	1.55	1454.48	-65.12%
Procter & Gamble Hygiene & Health Care Ltd.	4.17	4170	0.39	3968.344	-4.84%
Reliance Industries Ltd.	4.16	4160	2.99	3328.807	-19.98%
Shree Cement Ltd.	4.16	4160	0.22	3883.268	-6.65%
Tata Consultancy Services Ltd.	4.16	4160	2.05	3739.116	-10.12%
Tech Mahindra Ltd.	4.16	4160	5.30	2994.882	-28.01%
Titan Company Ltd.	4.16	4160	3.71	3466.945	-16.66%
U P L Ltd.	4.16	4160	4.45	1453.751	-65.05%
United Breweries Ltd.	4.16	4160	2.94	2705.369	-34.97%
Wipro Ltd.	4.16	4160	15.90	3127.353	-24.82%
	100	**100000**		**83097.99**	**-16.90%**
Nifty Benchmark Indices					
Nifty50					**-26.32%**
Nifty100					**-26.21%**

The performance of the portfolios during the general elections of 2019 was evaluated for the period from 01st April 2019 to 31st March 2020. For most of the period, the cumulative returns were negative. Although the ROIs of the proposed portfolios (MSRP: -15.75%, MVP: -18.61%, EWP: -16.90%) are higher than the benchmark indices (Nifty50: -26.32%, Nifty100: -26.21%), the daily performance displayed in Figure 6.9 says otherwise.

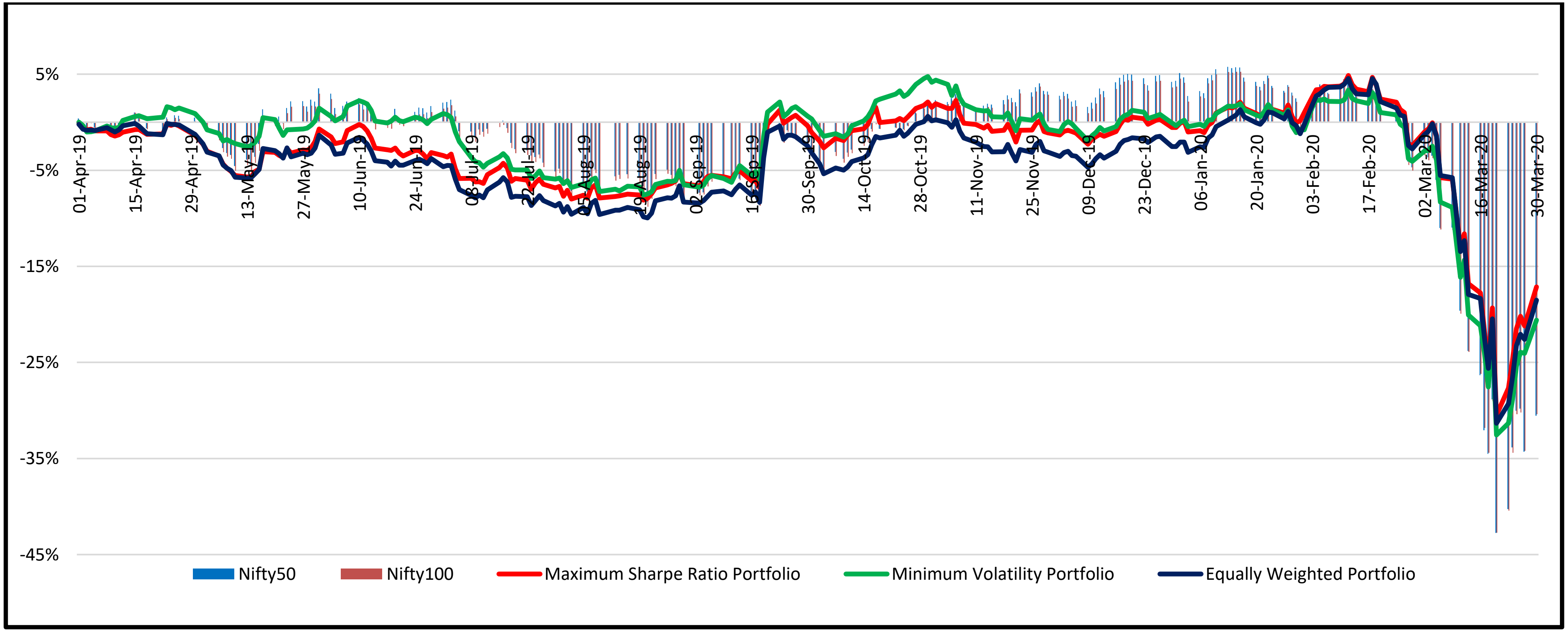


Figure 6.9: Cumulative returns for the period 01st April 2019 to 31st March 2020

6.2.4 The COVID-19 Pandemic

Table 6.10: Silhouette score for various clusters on the COVID-19 pandemic dataset

Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score
2	**0.418**	12	0.284	22	0.238
3	0.359	13	0.285	23	0.265
4	0.334	14	0.303	24	0.262
5	0.294	15	0.286	25	0.248
6	0.334	16	0.277	26	0.234
7	0.331	17	0.247	27	0.255
8	0.319	18	0.246	28	0.252
9	0.312	19	0.231	29	0.249
10	0.284	20	0.246	30	0.244
11	0.286	21	0.240		

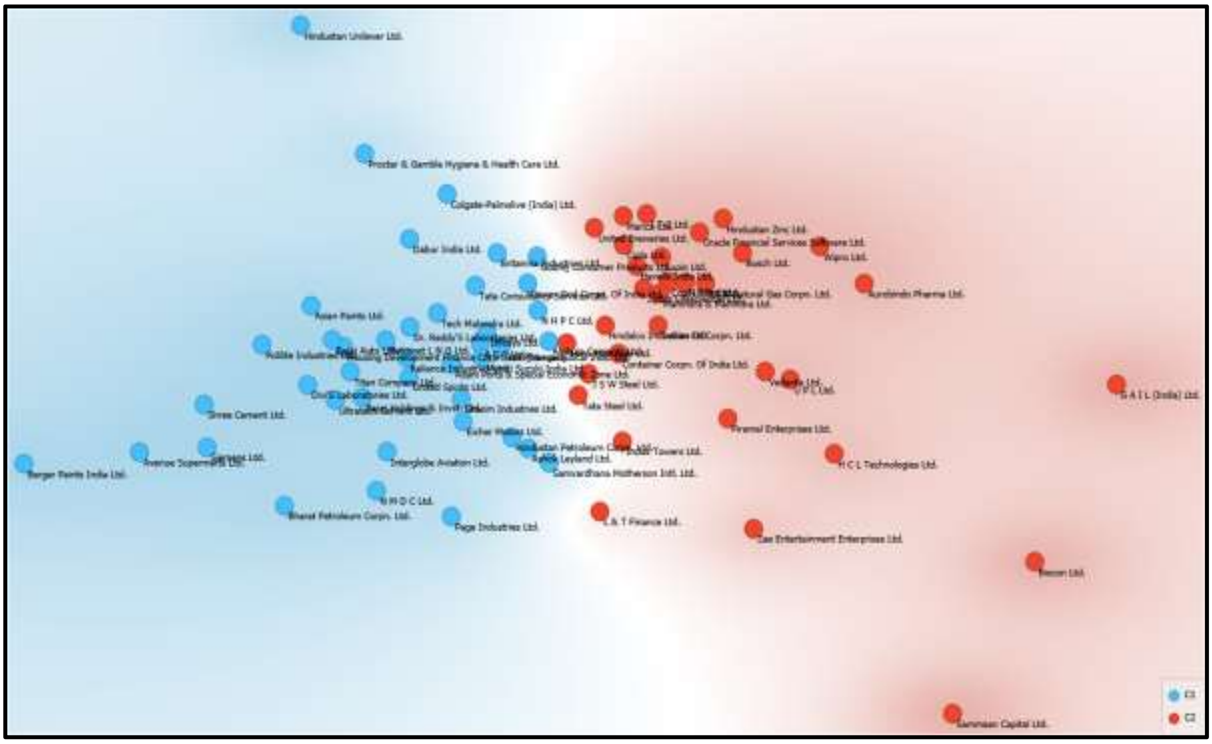


Figure 6.10: Cluster formation on the COVID-19 pandemic dataset

Table 6.10 displays the Silhouette scores obtained on forming several clusters, with the highest score (0.418) achieved on forming 2 clusters. From the scatter plot (in Figure 6.10), C1 represents higher returns with lower standard deviation, while C2 represents lower returns with higher risk. Therefore, C1 is selected for portfolio construction on the COVID-19 dataset. The correlation of stock returns for the COVID-19 pandemic dataset is represented as a heat map in Figure 6.11.

Table 6.11: Centroid details for the COVID-19 pandemic dataset

Cluster	Silhouette	Daily Average Returns	Standard Deviation
C1	**0.628224**	**0.688259**	**0.149928**
C2	0.618113	0.408698	0.268592

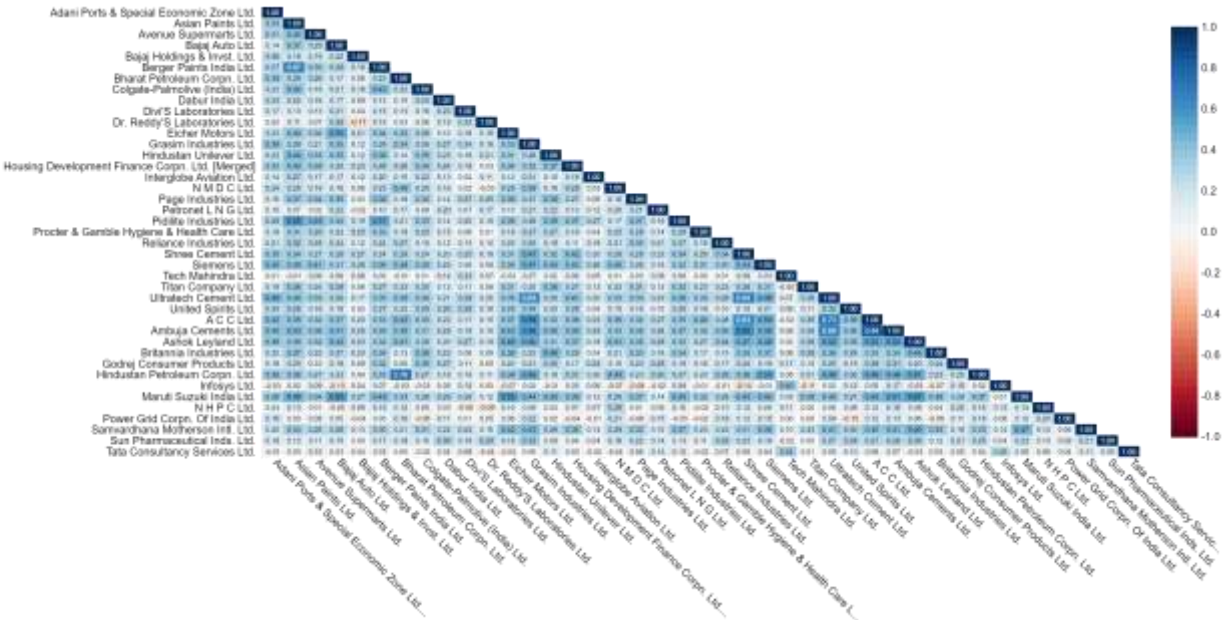


Figure 6.11: Correlation of stock returns on the COVID-19 pandemic dataset

Table 6.12: Fund allocation for selected stocks on the COVID-19 pandemic dataset

Selected Shares	Weight (%)	Initial investment as on 01st February 2020 (Rs.)	Number of Shares	Investment value as on 31st January 2021 (Rs.)	ROI
Maximum Sharpe Ratio Portfolio					
A C C Ltd.	1.37	1370	0.96	1540.218	12.42%
Adani Ports & Special Economic Zone Ltd.	0.6	600	1.66	844.1071	40.68%
Ambuja Cements Ltd.	0.6	600	3.07	747.4277	24.57%
Ashok Leyland Ltd.	0.25	250	3.24	359.0408	43.62%
Asian Paints Ltd.	1.19	1190	0.68	1632.009	37.14%
Avenue Supermarts Ltd.	4.69	4690	2.26	6001.561	27.97%
Bajaj Auto Ltd.	0.81	810	0.26	1032.258	27.44%
Bajaj Holdings & Invst. Ltd.	4.17	4170	1.18	3839.951	-7.91%
Berger Paints India Ltd.	4.37	4370	7.97	5631.603	28.87%
Bharat Petroleum Corpn. Ltd.	0.61	610	1.38	527.6122	-13.51%
Britannia Industries Ltd.	0.74	740	0.24	836.394	13.03%
Colgate-Palmolive (India) Ltd.	2.42	2420	1.85	2970.097	22.73%
Dabur India Ltd.	2.37	2370	4.87	2507.162	5.79%
Divi's Laboratories Ltd.	5.05	5050	2.54	8564.972	69.60%

Dr. Reddy's Laboratories Ltd.	5.11	5110	1.64	7531.034	47.38%
Eicher Motors Ltd.	2.98	2980	0.15	4076.929	36.81%
Godrej Consumer Products Ltd.	3.2	3200	4.82	3585.705	12.05%
Grasim Industries Ltd.	0.21	210	0.28	291.9232	39.01%
Hindustan Petroleum Corpn. Ltd.	0.2	200	0.89	194.5705	-2.71%
Hindustan Unilever Ltd.	0.78	780	0.38	851.0492	9.11%
Housing Development Finance Corpn. Ltd. [Merged]	5.08	5080	2.24	5325.349	4.83%
Infosys Ltd.	0.58	580	0.74	921.9359	58.95%
Interglobe Aviation Ltd.	2.01	2010	1.48	2290.599	13.96%
Maruti Suzuki India Ltd.	0	0	0.00	0	0
N H P C Ltd.	4.56	4560	190.79	4579.079	0.42%
N M D C Ltd.	4.98	4980	47.47	4994.242	0.29%
Page Industries Ltd.	3.22	3220	0.14	3693.557	14.71%
Petronet L N G Ltd.	4.81	4810	18.28	4331.193	-9.95%
Pidilite Industries Ltd.	4.07	4070	2.74	4576.181	12.44%
Power Grid Corpn. Of India Ltd.	3.74	3740	20.50	3782.034	1.12%
Procter & Gamble Hygiene & Health Care Ltd.	0.73	730	0.06	688.6717	-5.66%
Reliance Industries Ltd.	0.44	440	0.32	585.8662	33.15%
Samvardhana Motherson Intl. Ltd.	0.15	150	1.17	169.5322	13.02%
Shree Cement Ltd.	4.76	4760	0.21	4835.861	1.59%
Siemens Ltd.	0.85	850	0.60	952.1164	12.01%
Sun Pharmaceutical Inds. Ltd.	2.06	2060	4.87	2856.131	38.65%
Tata Consultancy Services Ltd.	3.89	3890	1.80	5590.758	43.72%
Tech Mahindra Ltd.	1.65	1650	2.04	1964.349	19.05%
Titan Company Ltd.	1.2	1200	1.02	1452.683	21.06%
Ultratech Cement Ltd.	5.12	5120	1.20	6394.823	24.90%
United Spirits Ltd.	4.38	4380	7.03	4069.529	-7.09%
	100	**100000**		**117620.1**	**17.62%**
Minimum Volatility Portfolio					
A C C Ltd.	1.27	1270	0.89	1427.793	12.42%
Adani Ports & Special Economic Zone Ltd.	1.26	1260	3.48	1772.625	40.68%
Ambuja Cements Ltd.	0.69	690	3.53	859.5418	24.57%
Ashok Leyland Ltd.	0.27	270	3.50	387.7641	43.62%
Asian Paints Ltd.	1.63	1630	0.93	2235.441	37.14%
Avenue Supermarts Ltd.	1.59	1590	0.77	2034.644	27.97%
Bajaj Auto Ltd.	0.69	690	0.22	879.3313	27.44%
Bajaj Holdings & Invst. Ltd.	1.75	1750	0.50	1611.49	-7.91%
Berger Paints India Ltd.	0.95	950	1.73	1224.262	28.87%

Bharat Petroleum Corpn. Ltd.	1.52	1520	3.43	1314.706	-13.51%
Britannia Industries Ltd.	0.73	730	0.24	825.0914	13.03%
Colgate-Palmolive (India) Ltd.	0.35	350	0.27	429.5595	22.73%
Dabur India Ltd.	3.92	3920	8.06	4146.867	5.79%
Divi's Laboratories Ltd.	4.7	4700	2.37	7971.36	69.60%
Dr. Reddy's Laboratories Ltd.	5.54	5540	1.77	8164.761	47.38%
Eicher Motors Ltd.	2.23	2230	0.11	3050.856	36.81%
Godrej Consumer Products Ltd.	5	5000	7.52	5602.663	12.05%
Grasim Industries Ltd.	3.36	3360	4.42	4670.772	39.01%
Hindustan Petroleum Corpn. Ltd.	0.22	220	0.98	214.0276	-2.71%
Hindustan Unilever Ltd.	1.3	1300	0.63	1418.415	9.11%
Housing Development Finance Corpn. Ltd. [Merged]	5.44	5440	2.40	5702.736	4.83%
Infosys Ltd.	5.38	5380	6.90	8551.75	58.95%
Interglobe Aviation Ltd.	0.32	320	0.24	364.6725	13.96%
Maruti Suzuki India Ltd.	0.37	370	0.05	391.3984	5.78%
N H P C Ltd.	1.68	1680	70.29	1687.029	0.42%
N M D C Ltd.	0.68	680	6.48	681.9447	0.29%
Page Industries Ltd.	2.03	2030	0.09	2328.547	14.71%
Petronet L N G Ltd.	5.79	5790	22.00	5213.64	-9.95%
Pidilite Industries Ltd.	0.86	860	0.58	966.9573	12.44%
Power Grid Corpn. Of India Ltd.	4.82	4820	26.43	4874.172	1.12%
Procter & Gamble Hygiene & Health Care Ltd.	4.1	4100	0.34	3867.882	-5.66%
Reliance Industries Ltd.	4.6	4600	3.33	6124.965	33.15%
Samvardhana Motherson Intl. Ltd.	0.14	140	1.09	158.23	13.02%
Shree Cement Ltd.	2.64	2640	0.12	2682.074	1.59%
Siemens Ltd.	0.38	380	0.27	425.6521	12.01%
Sun Pharmaceutical Inds. Ltd.	2.28	2280	5.39	3161.154	38.65%
Tata Consultancy Services Ltd.	4.77	4770	2.20	6855.505	43.72%
Tech Mahindra Ltd.	5.75	5750	7.12	6845.458	19.05%
Titan Company Ltd.	5.78	5780	4.93	6997.088	21.06%
Ultratech Cement Ltd.	1.96	1960	0.46	2448.018	24.90%
United Spirits Ltd.	1.26	1260	2.02	1170.686	-7.09%
	100	**100000**		**121741.5**	**21.74%**
Equally Weighted Portfolio					
A C C Ltd.	2.44	2440	1.71	2743.161	12.42%
Adani Ports & Special Economic Zone Ltd.	2.44	2440	6.73	3432.702	40.68%
Ambuja Cements Ltd.	2.44	2440	12.49	3039.539	24.57%
Ashok Leyland Ltd.	2.44	2440	31.63	3504.238	43.62%

Asian Paints Ltd.	2.44	2440	1.39	3346.304	37.14%
Avenue Supermarts Ltd.	2.44	2440	1.18	3122.347	27.97%
Bajaj Auto Ltd.	2.44	2440	0.78	3109.519	27.44%
Bajaj Holdings & Invst. Ltd.	2.44	2440	0.69	2246.878	-7.91%
Berger Paints India Ltd.	2.44	2440	4.45	3144.419	28.87%
Bharat Petroleum Corpn. Ltd.	2.44	2440	5.50	2110.449	-13.51%
Britannia Industries Ltd.	2.44	2440	0.79	2757.84	13.03%
Colgate-Palmolive (India) Ltd.	2.44	2440	1.87	2994.643	22.73%
Dabur India Ltd.	2.44	2440	5.02	2581.213	5.79%
Divi's Laboratories Ltd.	2.44	2440	1.23	4138.323	69.60%
Dr. Reddy's Laboratories Ltd.	2.44	2440	0.78	3596.032	47.38%
Eicher Motors Ltd.	2.44	2440	0.12	3338.157	36.81%
Godrej Consumer Products Ltd.	2.44	2440	3.67	2734.1	12.05%
Grasim Industries Ltd.	2.44	2440	3.21	3391.87	39.01%
Hindustan Petroleum Corpn. Ltd.	2.44	2440	10.86	2373.761	-2.71%
Hindustan Unilever Ltd.	2.44	2440	1.18	2662.256	9.11%
Housing Development Finance Corpn. Ltd. [Merged]	2.44	2440	1.08	2557.845	4.83%
Infosys Ltd.	2.44	2440	3.13	3878.489	58.95%
Interglobe Aviation Ltd.	2.44	2440	1.80	2780.628	13.96%
Maruti Suzuki India Ltd.	2.44	2440	0.36	2581.114	5.78%
N H P C Ltd.	2.44	2440	102.09	2450.209	0.42%
N M D C Ltd.	2.44	2440	23.26	2446.978	0.29%
Page Industries Ltd.	2.44	2440	0.10	2798.845	14.71%
Petronet L N G Ltd.	2.44	2440	9.27	2197.112	-9.95%
Pidilite Industries Ltd.	2.44	2440	1.64	2743.46	12.44%
Power Grid Corpn. Of India Ltd.	2.44	2440	13.38	2467.423	1.12%
Procter & Gamble Hygiene & Health Care Ltd.	2.44	2440	0.20	2301.862	-5.66%
Reliance Industries Ltd.	2.44	2440	1.76	3248.894	33.15%
Samvardhana Motherson Intl. Ltd.	2.44	2440	19.03	2757.723	13.02%
Shree Cement Ltd.	2.44	2440	0.11	2478.887	1.59%
Siemens Ltd.	2.44	2440	1.72	2733.134	12.01%
Sun Pharmaceutical Inds. Ltd.	2.44	2440	5.77	3382.99	38.65%
Tata Consultancy Services Ltd.	2.44	2440	1.13	3506.799	43.72%
Tech Mahindra Ltd.	2.43	2430	3.01	2892.95	19.05%
Titan Company Ltd.	2.43	2430	2.07	2941.683	21.06%
Ultratech Cement Ltd.	2.43	2430	0.57	3035.043	24.90%
United Spirits Ltd.	2.43	2430	3.90	2257.752	-7.09%
	100	**100000**		**118807.6**	**18.81%**

Nifty Benchmark Indices	
Nifty50	**16.92%**
Nifty100	**16.78%**

The funds allocated to each of the three portfolios are displayed in Table 6.12. The portfolios are held from 1st February 2020 to 31st January 2021. The returns on investment for the Maximum Sharpe Ratio Portfolio, Minimum Volatility Portfolio, and Equally Weighted Portfolio are 17.62%, 21.74%, and 18.81%, respectively. The constructed portfolios outperformed the Nifty50 and Nifty100 indices, which had an ROI of 16.92% and 16.78%, respectively, for that period. It is also evident from the cumulative returns graph in Figure 6.12 that the proposed portfolios outperformed the benchmark Nifty indices daily. Taking a closer look at the day of the announcement of COVID-19 as a pandemic, these portfolios performed comparatively better than the benchmark indices.

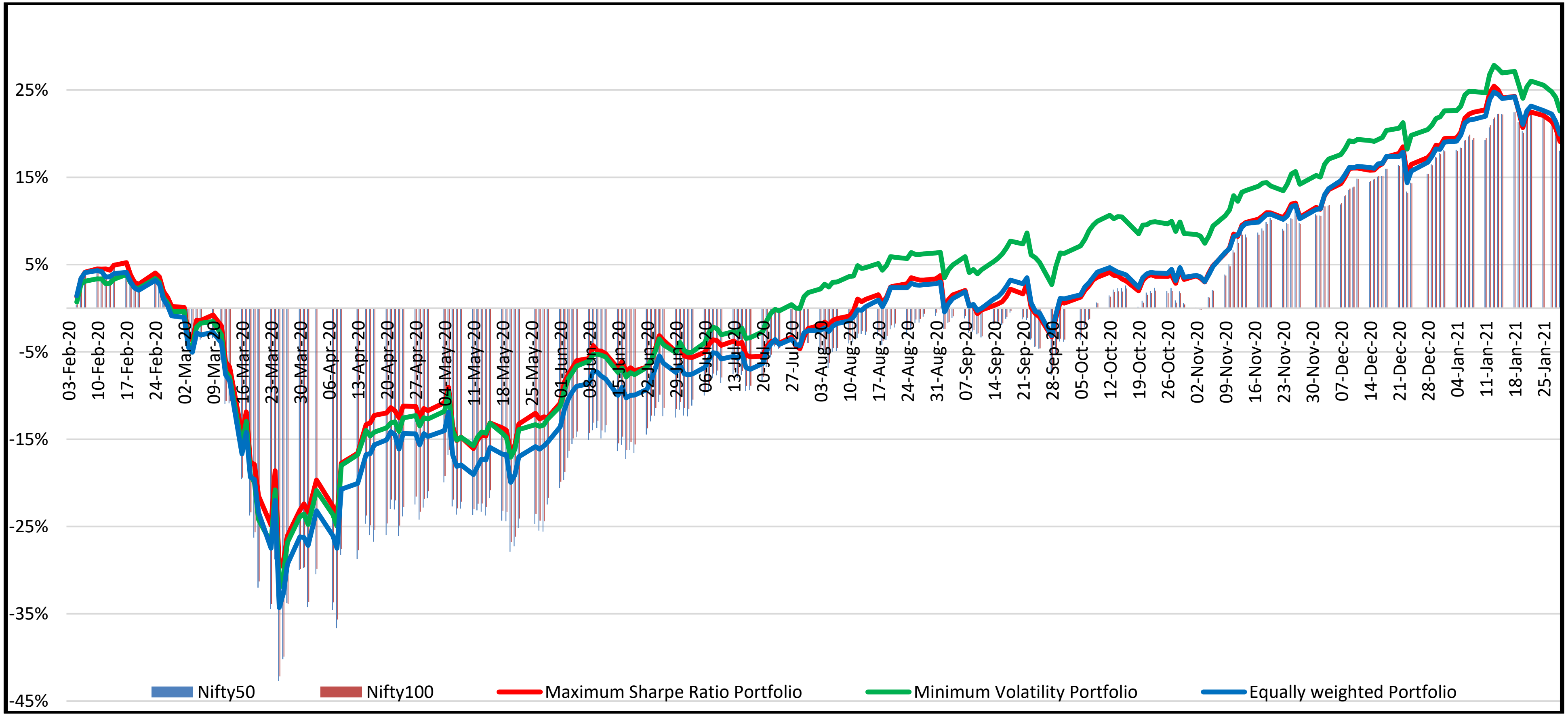


Figure 6.12: Cumulative returns for the period 01st February 2020 to 31st January 2021

6.2.5 The Invasion of Ukraine by Russia

Table 6.13: Silhouette score for various clusters on the Ukraine-Russia dataset

Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score
2	0.309	12	0.301	22	0.309
3	0.300	13	0.306	23	0.298
4	0.276	14	0.318	24	0.279
5	0.285	15	0.308	25	0.289
6	**0.408**	16	0.298	26	0.298
7	0.304	17	0.297	27	0.295
8	0.306	18	0.308	28	0.288
9	0.291	19	0.314	29	0.276
10	0.306	20	0.312	30	0.269
11	0.305	21	0.298		

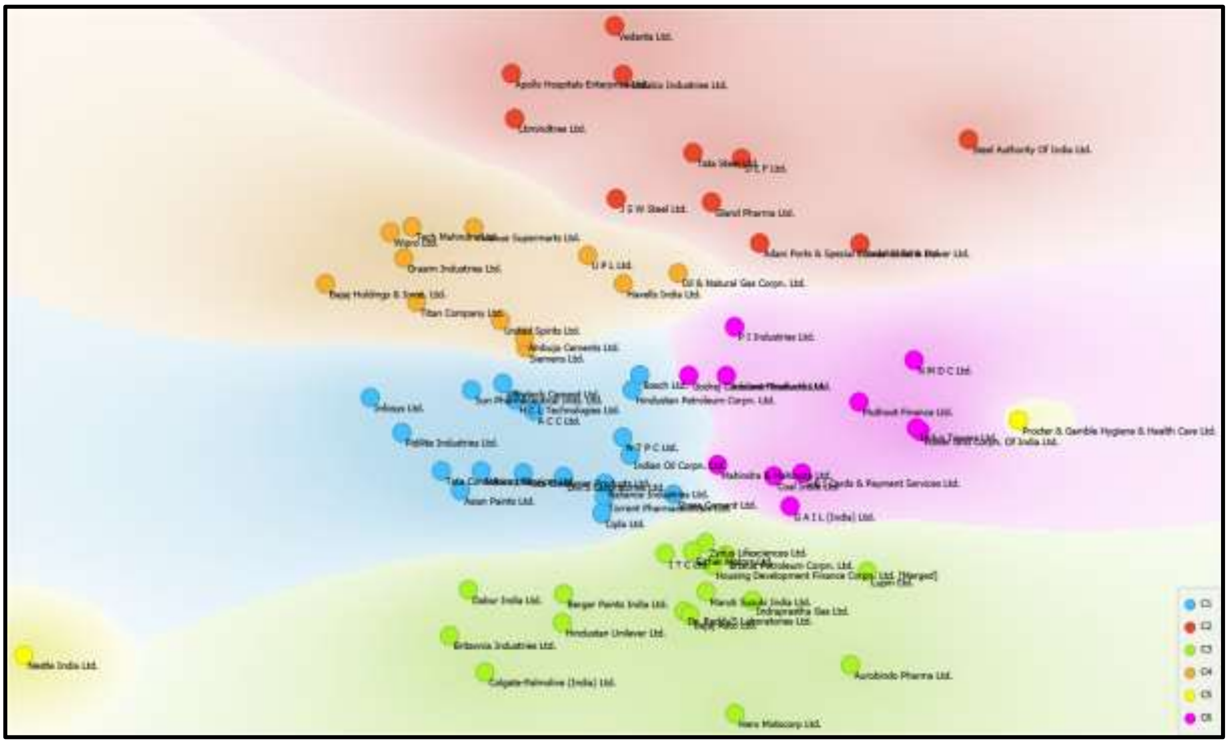


Figure 6.13: Cluster formation on the Ukraine-Russia dataset

Table 6.13 presents the silhouette scores for creating multiple clusters on the Ukraine-Russia dataset. The highest silhouette score (0.408) was obtained when the stocks were grouped into six clusters. The formation of these six clusters is displayed in Figure 6.13. Based on the returns and standard deviation of each of the six clusters presented in Table 6.14, cluster C4, which has high returns and low standard deviation, is selected to form our portfolio. C4 consists of 12 stocks, and the correlation of their returns is displayed in Figure 6.14. Overall, the heatmap indicates a low positive correlation among the stocks.

Table 6.14: Centroid details for the Ukraine-Russia dataset

Cluster	Silhouette	Daily Average Returns	Standard Deviation
C1	0.596161	0.311784	0.242362
C2	0.588222	0.533424	0.710324
C3	0.614509	0.128921	0.217365
C4	**0.593857**	**0.469804**	**0.365381**
C5	0.544693	0.26906	0.0731653
C6	0.572572	0.26784	0.520228

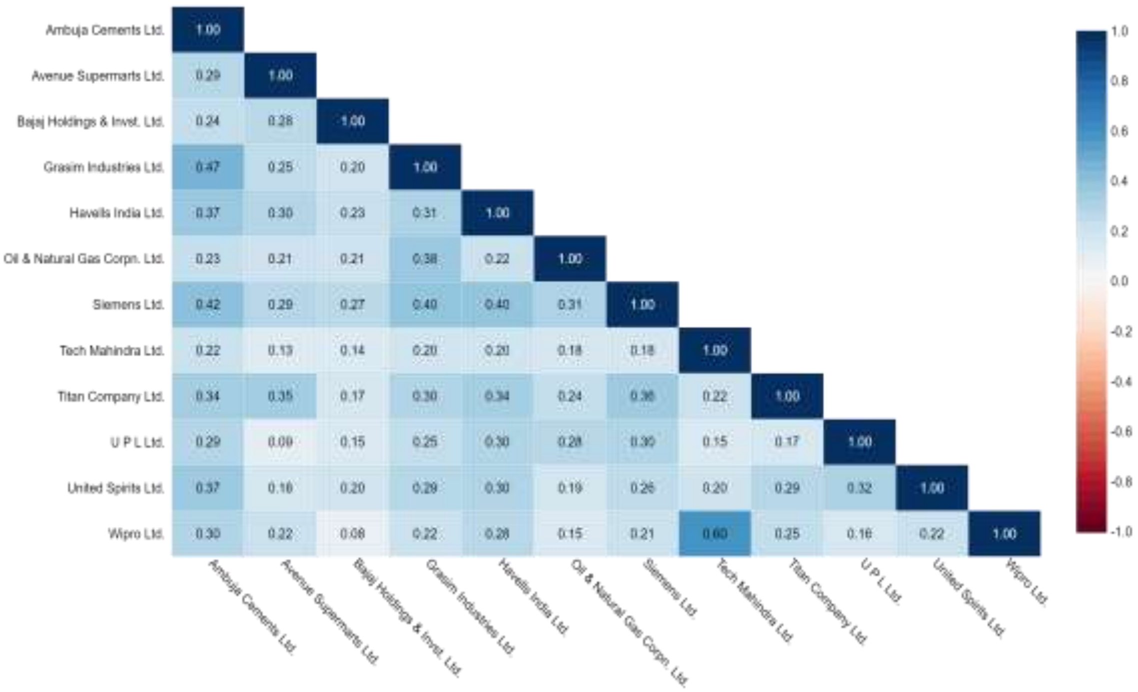


Figure 6.14: Correlation of stock returns on the Ukraine-Russia dataset

Table 6.15: Fund allocation for selected stocks on the Ukraine-Russia dataset

Selected Shares	**Weight (%)**	**Initial investment as on 01st January 2022 (Rs.)**	**Number of Shares**	**Investment value as on 31st December 2022 (Rs.)**	**ROI**
Maximum Sharpe Ratio Portfolio					
Ambuja Cements Ltd.	0.02	20	0.05168	27.08527	35.43%
Avenue Supermarts Ltd.	10.93	10930	2.319856	9438.913	-13.64%
Bajaj Holdings & Invst. Ltd.	18.36	18360	3.431872	19704.44	7.32%
Grasim Industries Ltd.	13.79	13790	8.29923	14303.72	3.73%
Havells India Ltd.	0.35	350	0.248907	273.7848	-21.78%
Oil & Natural Gas Corpn. Ltd.	1.59	1590	11.11499	1631.125	2.59%
Siemens Ltd.	2.47	2470	1.043603	2949.64	19.42%
Tech Mahindra Ltd.	15.44	15440	8.650829	8792.703	-43.05%
Titan Company Ltd.	11.16	11160	4.421816	11485.67	2.92%

U P L Ltd.	6.95	6950	9.094478	6513.01	-6.29%
United Spirits Ltd.	5.91	5910	6.562292	5758.739	-2.56%
Wipro Ltd.	13.03	13030	18.12996	7120.541	-45.35%
	100	**100000**		**87999.37**	**-12.00%**
Minimum Volatility Portfolio					
Ambuja Cements Ltd.	7.18	7180	18.55297	9723.612	35.43%
Avenue Supermarts Ltd.	5.92	5920	1.2565	5112.385	-13.64%
Bajaj Holdings & Invst. Ltd.	21.77	21770	4.069273	23364.14	7.32%
Grasim Industries Ltd.	5.89	5890	3.544776	6109.422	3.73%
Havells India Ltd.	0.8	800	0.568929	625.7938	-21.78%
Oil & Natural Gas Corpn. Ltd.	3.81	3810	26.63404	3908.546	2.59%
Siemens Ltd.	5.8	5800	2.450566	6926.28	19.42%
Tech Mahindra Ltd.	4.62	4620	2.588525	2630.977	-43.05%
Titan Company Ltd.	11.57	11570	4.584266	11907.63	2.92%
U P L Ltd.	5.42	5420	7.092384	5079.211	-6.29%
United Spirits Ltd.	9.61	9610	10.67066	9364.041	-2.56%
Wipro Ltd.	17.61	17610	24.50257	9623.386	-45.35%
	100	**100000**		**94375.42**	**-5.62%**
Equally Weighted Portfolio					
Ambuja Cements Ltd.	8.34	8340	21.55039	11294.56	35.43%
Avenue Supermarts Ltd.	8.34	8340	1.770137	7202.245	-13.64%
Bajaj Holdings & Invst. Ltd.	8.34	8340	1.558922	8950.708	7.32%
Grasim Industries Ltd.	8.34	8340	5.019259	8650.692	3.73%
Havells India Ltd.	8.33	8330	5.923977	6516.078	-21.78%
Oil & Natural Gas Corpn. Ltd.	8.33	8330	58.23139	8545.456	2.59%
Siemens Ltd.	8.33	8330	3.51952	9947.571	19.42%
Tech Mahindra Ltd.	8.33	8330	4.66719	4743.732	-43.05%
Titan Company Ltd.	8.33	8330	3.300513	8573.083	2.92%
U P L Ltd.	8.33	8330	10.90029	7806.241	-6.29%
United Spirits Ltd.	8.33	8330	9.249389	8116.802	-2.56%
Wipro Ltd.	8.33	8330	11.59037	4552.118	-45.35%
	100	**100000**		**94899.28**	**-5.10%**
Nifty Benchmark Indices					
Nifty50					**2.72%**
Nifty100					**2.15%**

Table 6.15 displays the portfolios held from 1st January 2022 to 31st December 2022. The return on investment for the proposed portfolios, MSRP, MVP, and EWP, were -12.00%, -5.62%, and -5.10%, respectively. In comparison, the Nifty50 and Nifty100 index had a positive ROI of 2.72% and 2.15%, respectively. From the cumulative returns graph in Figure 6.15, the benchmark indices continue in cumulative negative returns on the day of the event, while the EWP continues in a positive cumulative return.

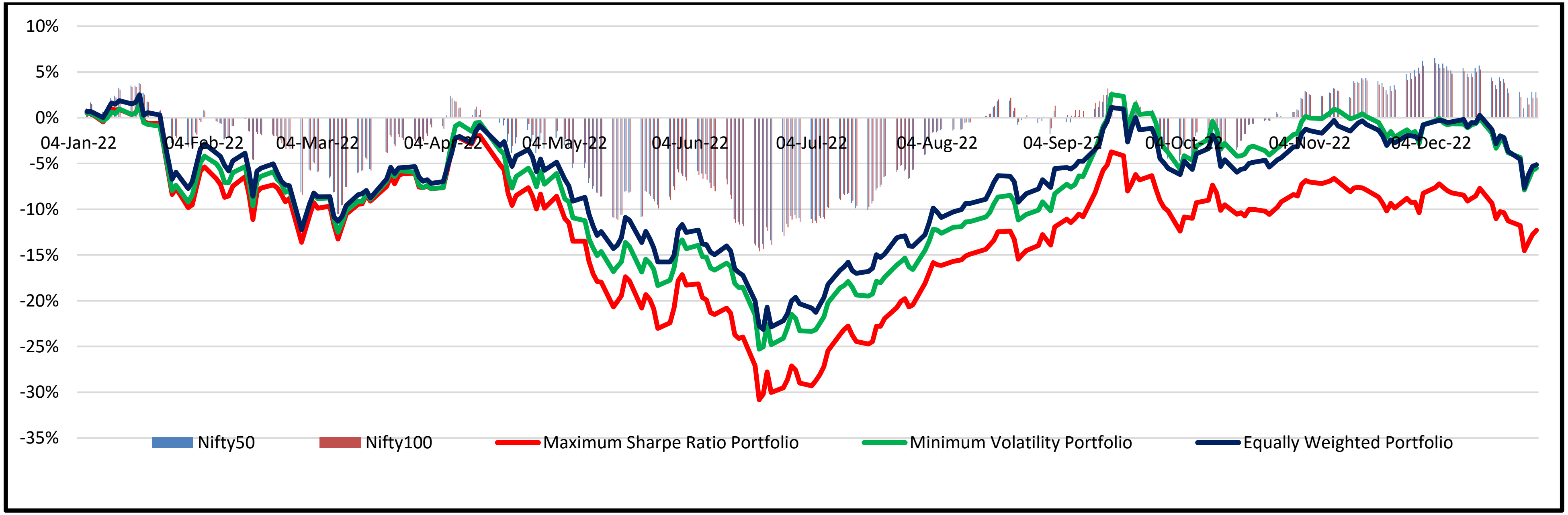


Figure 6.15: Cumulative returns for the period 01st January 2022 to 31st December 2022

6.2.6 The Hamas attack on Israel

Table 6.16: Silhouette score for various clusters on the Hamas-Israel dataset

Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score
2	0.367	12	0.300	22	0.247
3	0.282	13	0.260	23	0.225
4	0.346	14	0.284	24	0.253
5	0.347	15	0.283	25	0.248
6	0.355	16	0.285	26	0.249
7	**0.763**	17	0.266	27	0.242
8	0.365	18	0.257	28	0.240
9	0.280	19	0.257	29	0.228
10	0.321	20	0.249	30	0.228
11	0.323	21	0.229		

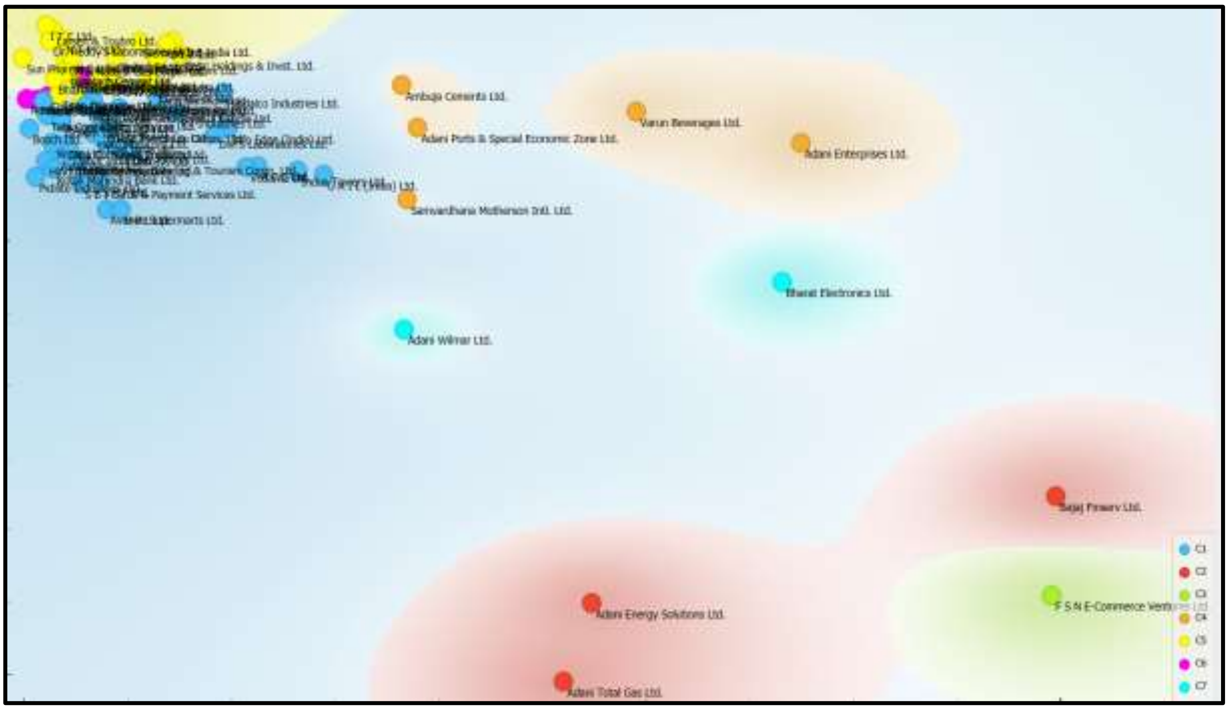


Figure 6.16: Cluster formation on the Hamas-Israel dataset

Table 6.16 shows that the highest silhouette score was obtained when the stocks were grouped into seven clusters. The formation of these seven clusters is portrayed in Figure 6.16. Based on the returns and standard deviation of each of the seven clusters (presented in Table 6.17), cluster C5, which has the highest returns and lower risk, is selected to form our portfolio.

Table 6.17: Centroid details for the Hamas-Israel dataset

Cluster	Silhouette	Daily Average Returns	Standard Deviation
C1	0.599233	0.737556	0.089969
C2	0.607946	0.108302	0.692303
C3	0.500000	0.113646	0.999949
C4	0.532972	0.745427	0.491643
C5	**0.660657**	**0.847873**	**0.071366**
C6	0.634324	0.806702	0.02317
C7	0.624687	0.512687	0.552906

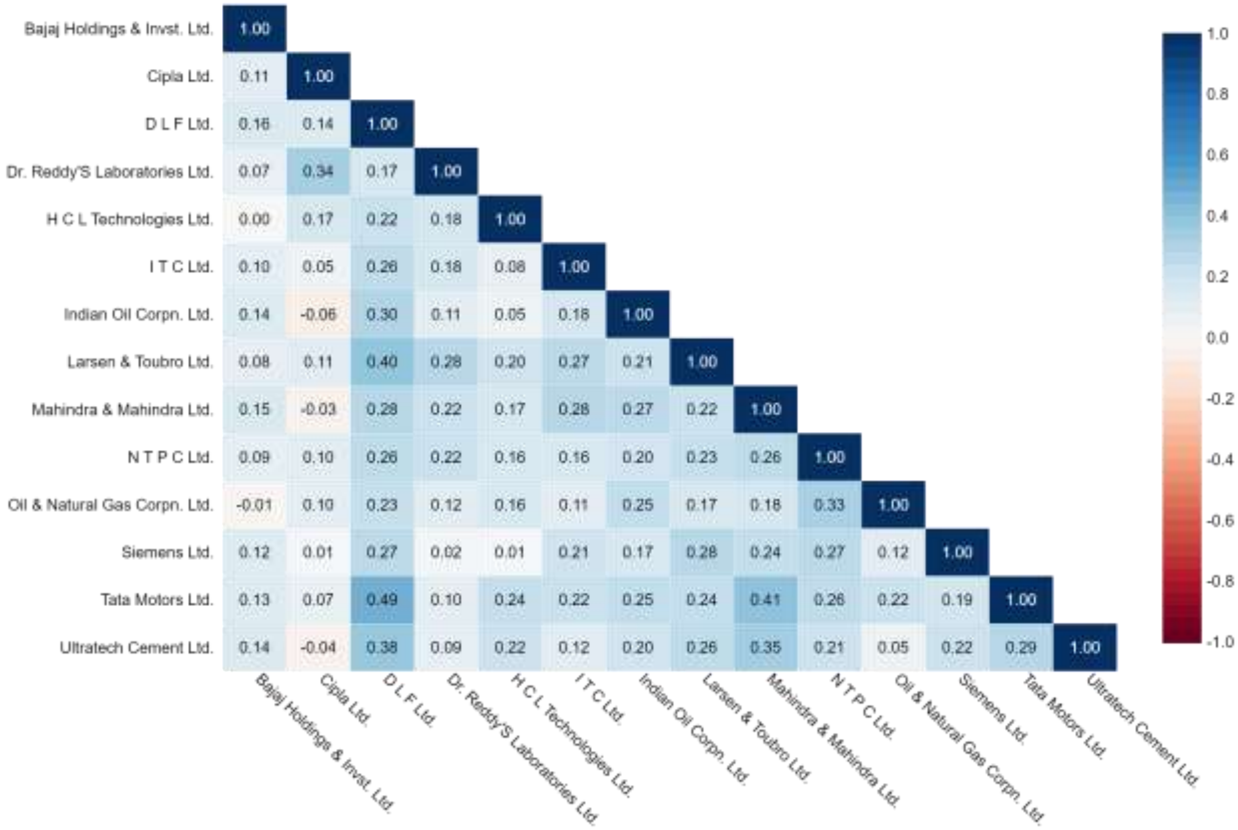


Figure 6.17: Correlation of stock returns on the Hamas-Israel dataset

The heatmap expressing the correlation among the fourteen selected stocks is displayed in Figure 6.17. The heatmap indicates a low correlation among the selected stocks.

Table 6.18: Fund allocation for selected stocks on the Hamas-Israel dataset

Selected Shares	Weight (%)	Initial investment as on 01st September 2023 (Rs.)	Number of Shares	Investment value as on 31st August 2024 (Rs.)	ROI
Maximum Sharpe Ratio Portfolio					
Bajaj Holdings & Invst. Ltd.	7.32	7320	0.98	9898.539	35.23%
Cipla Ltd.	11.68	11680	9.29	15369.94	31.59%
D L F Ltd.	0	0	0.00	0	0
Dr. Reddy's Laboratories Ltd.	14.23	14230	2.54	17842.15	25.38%
H C L Technologies Ltd.	2.38	2380	2.03	3559.898	49.58%
I T C Ltd.	14.78	14780	33.61	16870.78	14.15%
Indian Oil Corpn. Ltd.	8.22	8220	92.26	16326.53	98.62%
Larsen & Toubro Ltd.	10.73	10730	3.97	14707.85	37.07%
Mahindra & Mahindra Ltd.	0.96	960	0.61	1709.524	78.08%
N T P C Ltd.	11.26	11260	51.11	21272.86	88.92%
Oil & Natural Gas Corpn. Ltd.	6.55	6550	37.61	12439.92	89.92%
Siemens Ltd.	5.59	5590	1.43	9824.438	75.75%
Tata Motors Ltd.	1.26	1260	2.10	2329.952	84.92%
Ultratech Cement Ltd.	5.04	5040	0.61	6878.571	36.21%
	100.00	**100000**		**149030.9**	**49.03%**
Minimum Volatility Portfolio					
Bajaj Holdings & Invst. Ltd.	6.79	6790	0.91	9181.842	35.23%
Cipla Ltd.	12.57	12570	10.00	16541.1	31.59%
D L F Ltd.	0.52	520	1.03	871.7556	67.65%
Dr. Reddy's Laboratories Ltd.	12.12	12120	2.16	15196.55	25.38%
H C L Technologies Ltd.	15	15000	12.80	22436.33	49.58%
I T C Ltd.	12.42	12420	28.25	14176.93	14.15%
Indian Oil Corpn. Ltd.	14.01	14010	157.24	27826.6	98.62%
Larsen & Toubro Ltd.	1.82	1820	0.67	2494.714	37.07%
Mahindra & Mahindra Ltd.	0.8	800	0.51	1424.603	78.08%
N T P C Ltd.	7.49	7490	34.00	14150.42	88.92%
Oil & Natural Gas Corpn. Ltd.	7.39	7390	42.43	14035.27	89.92%
Siemens Ltd.	3.49	3490	0.89	6133.683	75.75%
Tata Motors Ltd.	0.93	930	1.55	1719.726	84.92%
Ultratech Cement Ltd.	4.65	4650	0.56	6333.733	36.21%
	100.00	**100000**		**152523.3**	**52.52%**
Equally Weighted Portfolio					
Bajaj Holdings & Invst. Ltd.	7.14	7140	0.96	9655.132	35.23%
Cipla Ltd.	7.14	7140	5.68	9395.663	31.59%
D L F Ltd.	7.14	7140	14.16	11969.88	67.65%
Dr. Reddy's Laboratories Ltd.	7.14	7140	1.27	8952.422	25.38%
H C L Technologies Ltd.	7.14	7140	6.09	10679.7	49.58%
I T C Ltd.	7.14	7140	16.24	8150.025	14.15%
Indian Oil Corpn. Ltd.	7.14	7140	80.13	14181.43	98.62%
Larsen & Toubro Ltd.	7.14	7140	2.64	9786.954	37.07%
Mahindra & Mahindra Ltd.	7.14	7140	4.53	12714.58	78.08%

N T P C Ltd.	7.14	7140	32.41	13489.19	88.92%
Oil & Natural Gas Corpn. Ltd.	7.15	7150	41.06	13579.46	89.92%
Siemens Ltd.	7.15	7150	1.82	12566.14	75.75%
Tata Motors Ltd.	7.15	7150	11.90	13221.55	84.92%
Ultratech Cement Ltd.	7.15	7150	0.86	9738.966	36.21%
	100.00	**100000**		**158081.1**	**58.08%**
Nifty Benchmark Indices					
Nifty50					**31.07%**
Nifty100					**37.13%**

The MSRP, MVP, and EWP have an ROI of 49.03%, 52.52%, and 58.08% for the 12 months during the holding period. Meanwhile, the Nifty 50 index and Nifty 100 index had a lower ROI of 31.07% and 37.13%, respectively. Based on the cumulative returns graph in Figure 6.18, the proposed portfolios performed comparatively better than the benchmark indices.

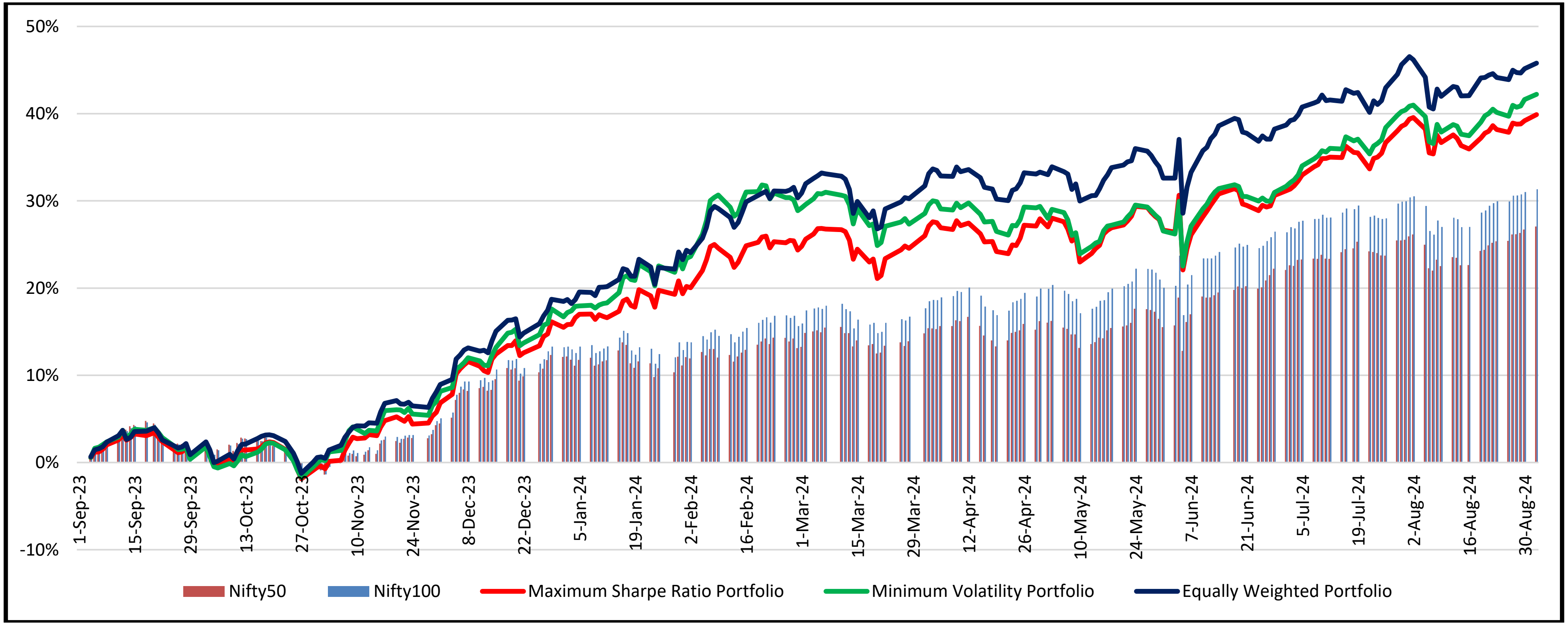


Figure 6.18: Cumulative returns for the period 01st September 2023 to 31st August 2024

6.2.7 The General Elections 2024

Table 6.19: Silhouette score for various clusters on the general elections 2024 dataset

Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score	Number of Clusters	Silhouette Score
2	0.290	12	0.259	22	0.249
3	0.297	13	0.249	23	0.242
4	0.313	14	0.274	24	0.250
5	**0.326**	15	0.257	25	0.261
6	0.281	16	0.260	26	0.233
7	0.282	17	0.266	27	0.236
8	0.283	18	0.268	28	0.210
9	0.283	19	0.263	29	0.214
10	0.287	20	0.245	30	0.207
11	0.284	21	0.260		

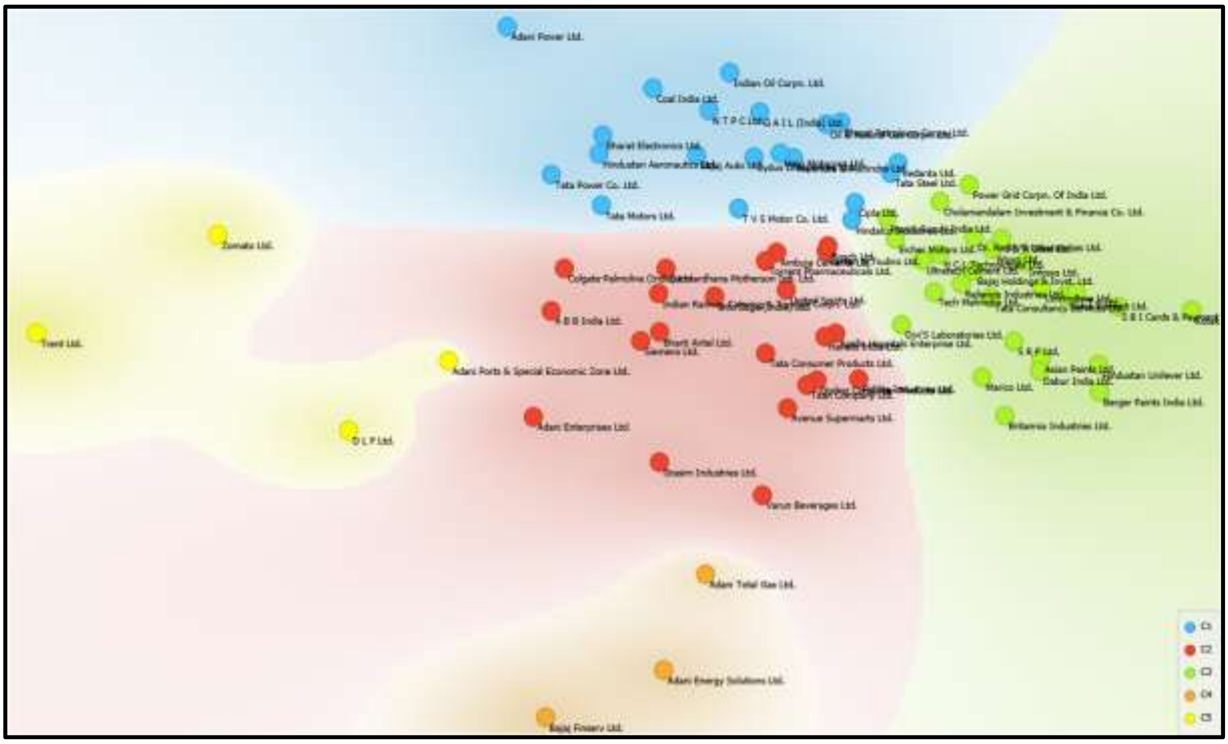


Figure 6.19: Cluster formation on the general elections 2024 dataset

Table 6.19 shows that the highest silhouette score was obtained when the stocks were grouped into five clusters. The formation of these five clusters is displayed in Figure 6.19. Based on the returns and standard deviation of each of the five clusters (presented in Table 6.20), cluster C5, which has the highest returns and moderate risk, is selected to form our portfolio.

Table 6.20: Centroid details for the general elections 2024 dataset

Cluster	Silhouette	Daily Average Returns	Standard Deviation
C1	0.6032	0.645854	0.184346
C2	0.559023	0.546638	0.140665
C3	0.632405	0.40442	0.089085
C4	0.575728	0.416321	0.355772
C5	**0.567545**	**0.79563**	**0.221484**

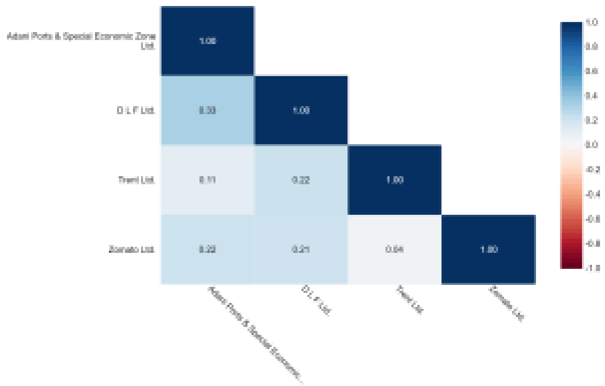


Figure 6.20: Correlation of stock returns on the general election 2024 dataset

The stocks contained in cluster C5 are Adani Ports & Special Economic Zone Ltd., which belongs to the infrastructure sector; D L F Ltd., which is in the realty sector; Trent Ltd., which is in the retailing sector; and Zomato Ltd., which belongs to the IT sector. All selected stocks have a low positive correlation, as seen in Figure 6.20.

Table 6.21: Fund allocation for selected stocks on the general election 2024 dataset

Selected Shares	**Weight (%)**	**Initial investment as on 01st May 2024 (Rs.)**	**Number of Shares**	**Investment value as on 31st October 2024 (Rs.)**	**ROI**
Maximum Sharpe Ratio Portfolio					
Adani Ports & Special Economic Zone Ltd.	12.08	12080	9.12	12545.45702	3.85%
D L F Ltd.	16.87	16870	18.92	15508.06694	-8.07%
Trent Ltd.	41.84	41840	9.49	67631.95628	61.64%
Zomato Ltd.	29.21	29210	151.23	36559.75925	25.16%
	100	**100000**		**132245.2395**	**32.25%**
Minimum Volatility Portfolio					
Adani Ports & Special Economic Zone Ltd.	21.72	21720	16.39	22556.89788	3.85%
D L F Ltd.	28.27	28270	31.70	25987.7328	-8.07%
Trent Ltd.	30.63	30630	6.95	49511.6353	61.64%
Zomato Ltd.	19.38	19380	100.34	24256.35516	25.16%
	100.00	**100000**		**122312.6211**	**22.31%**

Equally Weighted Portfolio					
Adani Ports & Special Economic Zone Ltd.	25	25000	18.87	25963.28025	3.85%
D L F Ltd.	25	25000	28.03	22981.72338	-8.07%
Trent Ltd.	25	25000	5.67	40411.06374	61.64%
Zomato Ltd.	25	25000	129.43	31290.44784	25.16%
	100.00	**100000**		**120646.5152**	**20.65%**
Nifty Benchmark Indices					
Nifty50					**6.88%**
Nifty100					**6.45%**

The portfolio performance during the general elections of 2024 was intended to be evaluated over 1-year, from 01st May 2024 to the 30th April 2025. However, the data for this study was extracted on 01st November 2024, and therefore, the analysis was conducted using the available data, i.e., up to 31st October 2024, covering six months.

The market reacted negatively to the exit polls announced before the election results, as did the proposed portfolios. However, by the end of the third month, each of the three proposed portfolios surpassed the benchmark Nifty indices by a considerable difference, as seen in the cumulative returns graph in Figure 6.21.

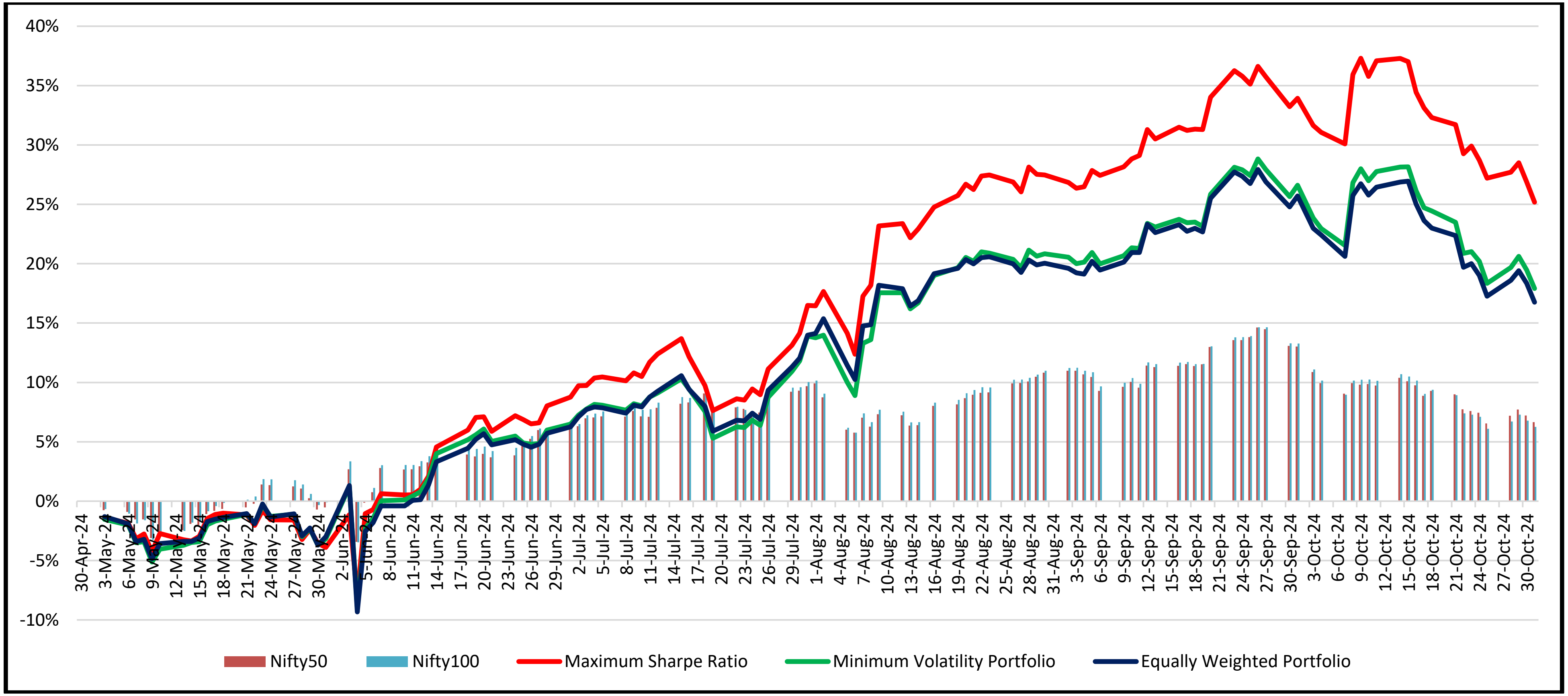


Figure 6.21: Cumulative returns for the period 01st May 2024 to 31st October 2024

Event	MSRP surpasses Nifty50 Index	MSRP surpasses Nifty100 Index	MVP surpasses Nifty50 Index	MVP surpasses Nifty100 Index	EWP surpasses Nifty50 Index	EWP surpasses Nifty100 Index
Demonetization	49%	49%	50%	49%	50%	51%
Implementation of GST	49%	48%	49%	48%	50%	49%
General Elections 2019	52%	52%	52%	51%	53%	52%
COVID-19	47%	48%	49%	49%	48%	49%
The Ukraine-Russia conflict	45%	43%	44%	43%	44%	44%
The Hamas-Israel conflict	54%	53%	54%	53%	54%	55%
General Election 2024	50%	50%	53%	50%	47%	47%

Table 6.22: Percentage of days the portfolios achieved returns higher than those of the Nifty50 and Nifty100 benchmark indices

While it is not possible to conclude that a particular portfolio outperformed the other, Table 6.22 provides strong evidence that the constructed portfolios outperformed the Nifty50 and Nifty100 on nearly half the trading days throughout the study period. The results indicate that an active strategy based on clustering techniques has the potential to compete with established market indices. However, long-term consistency has yet to be studied in order to draw definitive conclusions.

Chapter 7

Findings and Conclusion

This section provides the findings and conclusions drawn for each objective of the study, policy implications, contribution of the study, and scope for future research.

7.1 Findings of the Study

7.1.1 Findings from Objective I Analysis

The study demonstrates that LSTM models consistently outperformed ANN in terms of prediction accuracy, regardless of whether the predictions were for broad market indices, sectoral indices, or selected individual stocks. The LSTM model consistently demonstrated its efficiency in time-series predictions across all scheduled or unforeseen events. Although LSTM is computationally more intensive than ANN, it can retain long-term dependencies in data and can be utilized to generate reasonably accurate forecasts. Moreover, using MI based on entropy as a feature selection technique highlights the relevance of daily high and low stock prices over macroeconomic indicators in stock market predictions. This supports the premise that the stock market data already captures the information carried by macroeconomic variables.

LSTM demonstrated high accuracy in predicting the Nifty50, Nifty100, Nifty200, and Nifty500 indices, achieving a Root Mean Squared Error below 0.05 for all events except for Nifty100 during the general elections of 2024. This indicates that the model's predictions were nearly accurate across the scheduled and unforeseen events.

The R-squared values of the sectoral indices were nearly close to 1 (above 0.80) across all events, except the general elections in 2024 and the Nifty Auto index during the implementation of the GST period. This indicates that the selected predictor variables are a good fit for the model.

A prediction pattern similar to the broad market indices was seen in the stock-wise performance. LSTM had an RMSE below 0.05 across all the events (except for the general elections of 2024), indicating nearly accurate predictions.

LSTM maintained an RMSE below 0.05 for most selected stocks, suggesting strong predictive capabilities. However, accurately predicting the stock prices of Reliance Industries Ltd. and Siemens Ltd. proved to be more challenging. Reliance Industries Ltd. had an RMSE above 0.05 during four events, while Siemens Ltd experienced the same difficulty in three of the seven events.

Model-specific findings:

- Calculation time increased with an increase in the number of hidden layers.
- Calculation time increased with the number of neurons in the hidden layer.
- Systematic evaluation revealed that increasing the number of layers diminished model performance. Both the single-layer ANN and single-layer LSTM models achieved the lowest RMSE values compared to their multilayer counterparts. The results suggest that a simpler model structure with a single layer is sufficient to capture the relevant patterns in the data without overcomplicating the models. These findings align with previous work by Bhandari et al. (2022), as well as the studies of Funahashi (1989) and Hornik, Stinchcombe, and White (1989), which demonstrated that neural networks can approximate any unknown function to a desired level of accuracy using a single hidden layer, provided there are sufficient hidden units.
- The time taken to run an LSTM model was relatively more than that of an ANN model. (LSTM: approximately 100 to 120 seconds for a single hidden layer with 50 neurons and 60 epochs, ANN: 30 to 40 seconds for the same configuration)

7.1.2 Findings from Objective II Analysis

The k-means clustering model, as an unsupervised machine learning algorithm, groups stocks based on underlying patterns in the data without predefined labels. However, the exact rationale behind the clustering process remains opaque, as the model does not inherently provide explanations for why specific stocks are grouped together. We have selected the cluster with the highest mean returns and minimum or moderate risk for investment. Interestingly, the correlation of all stocks in the selected clusters is low.

With regards to portfolio optimization throughout the different events:

1. **Demonetization:**

All portfolios beat the market benchmark indices with high positive returns, and the Equally Weighted Portfolio showcases the highest returns.

2. **GST implementation:**

All portfolios have positive but moderate returns, with the Minimum Volatility Portfolio surpassing the benchmark indices while the Maximum Sharpe Ratio and Equally Weighted Portfolio miss the benchmark.

3. **General elections 2019:**

During this period, the market witnessed a downfall. However, all the machine learning-based portfolios outperformed the Nifty indices.

4. **Covid-19 Pandemic:**

During the year that COVID-19 was announced as a pandemic, the Nifty indices rose by approximately 17% compared to the previous year. On a brighter note, all three portfolios constructed using k-means clustering surpassed the Nifty benchmark indices for this period.

5. **Invasion of Ukraine by Russia:**

The constructed portfolios were unable to beat the market during this period. There were varied responses, with the Maximum Sharpe Ratio Portfolio showing significant negative returns, while Nifty indices showed slight positive returns.

6. **Hamas Attack on Israel:**

Significant positive returns were witnessed during this period, with the Equally Weighted Portfolio showing the highest gain (58.08%) and Nifty indices slightly lower (31.07% for Nifty 50 and 37.13% for Nifty 100).

7. **General Elections 2024:**

The three machine learning-constructed portfolios surpassed the market index by a considerable margin, with the Maximum Sharpe Ratio Portfolio leading with 32.25% returns, followed by the Minimum Volatility Portfolio (22.31%) and the Equally Weighted Portfolio (20.65%). In contrast, the Nifty benchmark indices showed lower expected returns (~6.5%).

7.2 Conclusion of the Study

This empirical study focuses on applying machine learning models to enhance market predictions and portfolio optimization of the Indian stock market from 2012 to 2024. The selected period covers several unforeseen and scheduled events critical to shaping the Indian economy. The scheduled events include the implementation of the GST regime in 2017 and the national general elections of 2019 and 2024. Unforeseen events include demonetization in 2016, the global COVID-19 pandemic in 2020, the invasion of Ukraine by Russia in 2022, and the Hamas attack on Israel in 2023.

Firstly, the study focused on predicting the daily closing index values of the broad market and sectoral indices and the closing prices of individual stocks. The prediction period began at least a month prior to the occurrence of the event in order to ensure that our forecasting analysis accounts for the dynamics leading to these critical events.

From a comprehensive dataset comprising stock market data and macroeconomic variables, the most informative features for predicting the close price were identified using a 75-percentile threshold. A filter-based feature selection technique called Mutual Information based on Entropy revealed the significance of daily high and low prices as key determinants in predicting the target variable. This indicates that the stock market data already captures the information carried by macroeconomic variables; thus, adding macroeconomic variables to predict the Nifty indices or stock prices does not provide significant additional insight. It is important to consider that macroeconomic data reflects past trends rather than immediate market sentiment; therefore, it becomes challenging to predict short-term market movements accurately using exogenous factors.

The study reveals the superiority of LSTM over ANN in each event and across indices and stocks. In most scenarios, the performance of LSTM predictions was almost nearly accurate. With regards to executing both the deep learning models, we noticed that an increase in the number of hidden layers resulted in diminished model performance in terms of MSE. The results suggest that a simpler model structure with a single layer is sufficient to capture the relevant patterns in the data without overcomplicating the models. These findings align with previous work by Bhandari et al. (2022), as well as the studies of Funahashi (1989) and Hornik, Stinchcombe, and White (1989), which demonstrated that neural networks can approximate any unknown function to a desired level of accuracy using a single hidden layer, provided there are sufficient hidden units. The calculation time increased with the number of hidden layers and neurons in the hidden layer. Relatively, the time taken to run an LSTM model was more than that of an ANN model.

Further, in this study, we proposed a portfolio optimization model using unsupervised machine learning and Monte Carlo simulations. We develop a mechanism to select stocks and allocate funds in a portfolio. We tested the robustness of this model by constructing portfolios during the same periods as highlighted in the first objective. The K-mean clustering model formed clusters of stocks constituted in the Nifty100 index on the basis of their returns, standard deviation, and validation ratios. When grouping stocks into different clusters, the clusters with the highest returns and lowest risk or with the highest risk-adjusted returns for each event were selected. Three portfolios were created for each event: the EWP, the MSRP, and the MVP. The funds were allocated in the EWP equally, while Monte Carlo simulations were used to calculate the weights for each stock in MSRP and MVP. Accordingly, each portfolio's performance was evaluated for a period of one year, and the performance of the benchmark indices for the same period was examined and recorded simultaneously. During the period in which the event of demonetization occurred, the returns obtained by holding an EWP was 51.16% as compared to Nifty50 at 12.02%; during the implementation of the GST regime MV was at 12.49% return and Nifty50 index at 11.65%, during General election 2019 MSRP reduced to -15.75% while Nifty50 reduced to a greater extent of -26.32%. During the COVID-19 pandemic, EWP achieved 18.81% return and Nifty50 achieved 16.92%. During the invasion of Ukraine by Russia, EWP reduced to a negative -5.10% return, while Nifty50 remained at a positive 2.72% return. During the attack on Israel by Hamas, the EWP achieved a significant return of 58.08%, while Nifty50 resulted in 31.07% results. During the 6-month evaluation period of the general elections 2024, MSRP achieved 32.35%, and Nifty50 achieved 6.88% returns. EWP consistently performs well, particularly in unexpected events (e.g., Hamas attack on Israel and demonetization). The constructed portfolios outperformed the Nifty benchmark indices in six out of the seven events. Nearly half the trading days in the period under study, the constructed portfolios' return surpassed the Nifty50 and Nifty100 daily returns, indicating that an active strategy based on clustering techniques has the potential to compete with established market indices.

7.3 Policy Implications of the Study

Integrating machine learning models highlights the potential of data-driven approaches in financial policymaking. Governments and regulatory bodies can adopt similar methodologies to enhance predictive accuracy and policy effectiveness.

Insights from the study's portfolio optimization techniques can inform national investment strategies, emphasizing diversification and risk mitigation to safeguard public investments.

The findings can be utilized to develop educational programs aimed at increasing public understanding of market dynamics and portfolio management, encouraging greater participation in equity markets.

7.4 Contribution of the Study

This empirical study contributes to financial time series forecasting, focusing on optimizing predictive techniques through systematic feature selection and hyperparameter tuning. The findings of this research contribute to the existing body of knowledge on predictive modeling in financial markets. It provides valuable insights for investors, traders, policymakers, financial institutions, and researchers seeking more accurate predictions in the face of market volatility.

This study contributes to advancing portfolio management techniques by integrating machine learning approaches. The study explores three robust portfolio allocation strategies—Equally Weighted Portfolio, Maximum Sharpe Ratio Portfolio, and Minimum Volatility Portfolio—within the context of Indian stock markets. This provides practical guidance for investors seeking to optimize risk-adjusted returns.

By combining academic rigor with practical financial strategies, the study bridges theoretical advancements in financial modeling and real-world investment decision-making.

7.5 Scope for Further Research

Due to the chosen time frame, key variables such as the Consumer Price Index, 91-day Treasury Bill rate, and unemployment rate were omitted. Incorporating these variables for stock market prediction models highlights an avenue for further exploration. Long-term predictions can be explored to evaluate the impact of macroeconomic features on the predictive performance of stock market models. Moreover, an interesting extension could be the inclusion of behavioral finance indicators, such as investor sentiment or market mood, to improve the predictive capability of models and provide a holistic understanding of market dynamics.

Future studies could focus on integrating explainable AI frameworks to enhance machine learning models' interpretability for stock market predictions. This would help stakeholders better understand model outputs and their implications.

The current study is limited to stocks within the Nifty index. Future research could expand the analysis to include international indices. This would enable an evaluation across a more diverse range of assets.

Regarding portfolio optimization, this study relies on a specific set of financial indicators for clustering. Future research could examine the impact of alternative metrics, such as earnings yield, dividend payout ratios, or ESG scores, on clustering performance and portfolio outcomes.

While this study focuses on Equally Weighted, Maximum Sharpe Ratio, and Minimum Volatility Portfolios, future research could investigate other allocation techniques, such as Risk Parity Portfolios, Factor-Based Portfolios, or Minimum Drawdown Portfolios. Additionally, integrating machine learning algorithms for dynamic portfolio rebalancing could further enhance performance.

Implementing the predictive models in real-time trading environments could validate their practical utility. This would involve testing their ability to adapt to real-world constraints such as transaction costs.

RESEARCH PAPER PUBLICATIONS

- Published a research paper titled "Financial applications of machine learning: A literature review" in **Expert Systems with Applications**, Volume 219, 2023 **(SCOPUS Indexed Journal)**.
- Published a research paper titled "Predictive Analysis of S&P BSE GREENEX Index: Unlocking Insights for Sustainable Investments" in **Australasian Accounting, Business and Finance Journal**, Volume 18, Issue 3, 2024 **(SCOPUS Indexed Journal)**.

RESEARCH PAPER PRESENTATIONS

- Presented a paper titled "**Stock market prediction using LSTM and ANN**" at the International Finance and Accounting Conference (IFAC - 2023) hosted by IIM Jammu on 08^{th} – 09^{th} September 2023.
- Presented a paper titled "**Predictive Analysis of S&P BSE GREENEX Index**" at the ISDSI-Global Conference 2023 held at IIM Ranchi from 26^{th} to 29^{th} December 2023.

Bibliography

Alqaryouti, O., Farouk, T., & Siyam, N. (2018). Clustering Stock Markets for Balanced Portfolio Construction. *International Conference on Advanced Intelligent Systems and Informatics* (pp. 577-587). Springer Nature. https://link.springer.com/chapter/10.1007/978-3-319-99010-1_53

Atkins, A., Niranjan, M., & Gerding, E. (2018). Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science, 4*, 120-137. https://doi.org/10.1016/j.jfds.2018.02.002

Ayala, J., Garcia-Torres, M., Noguera, J. L., Gomez-Vela, F., & Divina, F. (2021). Technical analysis strategy optimization using a machine learning approach in stock market indices. *Knowledge-Based Systems, 225*. https://doi.org/10.1016/j.knosys.2021.107119

B.S., B., & Mathew, T. (2016). Clustering and Regression Techniques for Stock Prediction. *International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015), 24*, pp. 1248-1255. https://doi.org/10.1016/j.protcy.2016.05.104

Baek, S., Mohanty, S. A., & Glambosky, M. (2020). COVID-19 and stock market volatility: An industry level analysis. *Finance Research Letters, 37*. https://doi.org/10.1016/j.frl.2020.101748

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *North American Journal of Economics, 47*, 552 - 567. https://doi.org/10.1016/j.najef.2018.06.013

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, K. R. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications, 9*(100320), 1-15. https://doi.org/10.1016/j.mlwa.2022.100320

Bustos, O., & Quimbaya, P. A. (2020). Stock market movement forecast: A Systematic review. *Expert Systems With Applications, 156*. https://doi.org/10.1016/j.eswa.2020.113464

Chandra, R., & Chand, S. (2016). Evaluation of co-evolutionary neural network architectures for time series prediction with mobile application in finance. *Applied Soft Computing, 49*, 462-473. http://dx.doi.org/10.1016/j.asoc.2016.08.029

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering, 40*, 16-28. http://dx.doi.org/10.1016/j.compeleceng.2013.11.024

Chen, H., Mai, Y., & Li, S.-P. (2014). Analysis of network clustering behavior of the Chinese stock market. *Physica A, 414*, 360-367. http://dx.doi.org/10.1016/j.physa.2014.07.039

Chen, L.-H., & Huang, L. (2009). Portfolio optimization of equity mutual funds with fuzzy return rates and risks. *Expert Systems with Applications, 36*, 3720–3727. https://doi.org/10.1016/j.eswa.2008.02.027

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing Journal, 100*, 106943. https://doi.org/10.1016/j.asoc.2020.106943

Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems With Applications, 80*, 340-355. http://dx.doi.org/10.1016/j.eswa.2017.02.044

Cheng, J.-H., Chen, H.-P., & Lin, Y.-M. (2010). A hybrid forecast marketing timing model based on probabilistic neural network, rough set and C4.5. *Expert Systems with Applications, 37*(3), 1814-1820. https://doi.org/10.1016/j.eswa.2009.07.019

Cheung, K. (2020, August 20). *10 Applications of Machine Learning in Finance*. Retrieved from https://algorithmxlab.com/blog/applications-machine-learning-finance/

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems With Applications, 83*, 187-205. http://dx.doi.org/10.1016/j.eswa.2017.04.030

Cover, T. M., & Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley & Sons Publishers.

D'Urso, P., Cappelli, C., Lallo, D. D., & Massari, R. (2013). Clustering of financial time series. *Physica A: Statistical Mechanics and its Applications, 392*(9), 2114-2129. https://doi.org/10.1016/j.physa.2013.01.027

Das, S., Behera, R. K., Kumar, M., & Rath, S. K. (2018). Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction. *International Conference on Computational Intelligence and Data Science (ICCIDS 2018), 132*, pp. 956-964. http://10.1016/j.procs.2018.05.111

de Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications, 40*(18), 7596-7606. https://doi.org/10.1016/j.eswa.2013.06.071

Deng, X., & Yuan, Y. (2021). A novel fuzzy dominant goal programming for portfolio selection with systematic risk and non-systematic risk. *Soft Computing, 25*, 14809–14828. https://doi.org/10.1007/s00500-021-06226-x

Domashova, J., & Zabelina, O. (2021). Detection of fraudulent transactions using SAS Viya machine learning algorithms. *Procedia Computer Science, 190*, 204-209. https://doi.org/10.1016/j.procs.2021.06.025

Fallahpour, S., Zadeh, M. H., & Lakvan, E. N. (2014). Use of Clustering Approach For Portfolio Management. *International SAMANM Journal of Finance and Accounting, 2*(1), 115-136.

Fama, E. (1965). The behaviour of stock market prices. *The Journal of Business, 38*, 34-105. http://www.jstor.org/stable/2350752

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*(2), 654-669. https://doi.org/10.1016/j.ejor.2017.11.054

Frénay, B., Doquire, G., & Verleysen, M. (2013). Is mutual information adequate for feature selection in regression? *Neural Networks, 48*, 1-7. http://dx.doi.org/10.1016/j.neunet.2013.07.003

Funahashi, K.-I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks, 2*(3), 183-192. https://doi.org/10.1016/0893-6080(89)90003-8

Gao, P., Zhang, R., & Yang, X. (2020). The Application of Stock Index Price Prediction with Neural Network. *Mathematical and Computational Applications, 25*(53), 1-16. https://www.mdpi.com/2297-8747/25/3/53

Gers, F., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: continual prediction with LSTM. *Neural Computation, 12*(10), 2451-71. https://doi.org/10.1162/089976600300015015

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*(5-6), 602-610. https://doi.org/10.1016/j.neunet.2005.06.042

Gu, Z., Wu, B., Hu, Z., & Yu, X. (2024). Credit risk assessment of small and micro enterprise based on machine learning. *Heliyon, 10*, 1-12. https://doi.org/10.1016/j.heliyon.2024.e27096

Gunasekaran, M., & Ramaswami, K. (2015). Portfolio Optimization Using Neuro Fuzzy System In Indian Stock Market. *Journal of Global Research in Computer Science, 3*(4), 44-47.

Hargreaves, C. A., Dixit, P., & Solanki, A. (2013). Stock Portfolio Selection using Data Mining Approach. *IOSR Journal of Engineering (IOSRJEN), 3*(11), 42-48. https://scholarbank.nus.edu.sg/handle/10635/171779

He, C., Wen, Z., Huang, K., & Ji, X. (2022). Sudden shock and stock market network structure characteristics: A comparison of past crisis events. *Technological Forecasting & Social Change, 180*(121732), 1-16. https://doi.org/10.1016/j.techfore.2022.121732

Heston, S., & Sinha, N. R. (2018). News vs. Sentiment: Predicting Stock Returns from News

Stories. *Financial Analysts Journal, 73*(3), 67-83. https://doi.org/10.2469/faj.v73.n3.3

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735-1780.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*(5), 359-366. https://doi.org/10.1016/0893-6080(89)90020-8

Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation, 9*(26). https://doi.org/10.1186/s40854-022-00441-7

Hussain, A. J., Knowles, A., Lisboa, P. J., & El-Deredy, W. (2008). Financial time series prediction using polynomial pipelined neural networks. *Expert Systems with Applications, 35*(3), 1186-1199. https://doi.org/10.1016/j.eswa.2007.08.038

Juan, D. (2022). Mean–variance portfolio optimization with deep learning based-forecasts for cointegrated stocks. *Expert Systems With Applications, 201*(117005), 1-11.

Jung, S. S., & Chang, W. (2016). Clustering stocks using partial correlation coefficients. *Physica A, 462*, 410-420. http://dx.doi.org/10.1016/j.physa.2016.06.094

Keyan, L., Jianan, Z., & Dayong, D. (2021). Improving stock price prediction using the long short-term memory model combined with online social networks. *Journal of Behavioral and Experimental Finance, 30*. https://doi.org/10.1016/j.jbef.2021.100507

Khattak, M. A., Ali, M., & Rizvi, S. A. (2021). Predicting the European stock market during COVID-19: A machine learning approach. *MethodsX, 8*. https://doi.org/10.1016/j.mex.2020.101198

Kia, A. N., Haratizadeh, S., & Shouraki, S. B. (2018). A hybrid supervised semi-supervised graph-based model to predict one-day ahead movement of global stock markets and commodity prices. *Expert Systems With Applications, 105*, 159-173. https://doi.org/10.1016/j.eswa.2018.03.037

Kılıc¸man, A., & Sivalingam, J. (2010). Portfolio Optimization of Equity Mutual Funds—Malaysian Case Study. *Advances in Fuzzy Systems, 2010*. https://doi.org/10.1155/2010/879453

Kinyua, J. D., Mutigwe, C., Cushing, D. J., & Poggi, M. (2021). An analysis of the impact of President Trump's tweets on the DJIA and S&P 500 using machine learning and sentiment analysis. *Journal of Behavioral and Experimental Finance, 29*. https://doi.org/10.1016/j.jbef.2020.100447

Kong, F. (2024). Anomaly Detection and Adaptive Learning Algorithm in Financial Risk Monitoring. *Procedia Computer Science, 247*, 988-995. https://doi.org/10.1016/j.procs.2024.10.119

Koratamaddi, P., Wadhwani, K., Gupta, M., & Sanjeevi, S. G. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. *Engineering Science and Technology, an International Journal, 24*(4), 848-859. https://doi.org/10.1016/j.jestch.2021.01.007

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems, 104*, 38-48. https://doi.org/10.1016/j.dss.2017.10.001

Kroese, D., Brereton, T., Taimre, T., & Botev, Z. (2014). Why the Monte Carlo method is so important today. *WIREs Computational Statistics, 6*, 386-392. https://doi.org/10.1002/wics.1314

Li, L. (2019, November 13). Direct Portfolio Selection Using Recurrent Reinforcement Learning. *SSRN Electronic Journal*. https://www.econbiz.de/Record/direct-portfolio-selection-using-recurrent-reinforcement-learning-lin/10012860075

Li, Q., & Zhong, W. (2022). The Application of Monte Carlo Simulation and Mean-variance in Portfolio Selection. *BCP Business & Management, 26*, 1216-1221.

Li, Y., Bu, H., Li, J., & Wu, J. (2020). The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning. *International Journal of Forecasting, 36*, 1541-1562. https://doi.org/10.1016/j.ijforecast.2020.05.001

Lohrmann, C., & Luukka , P. (2019). Classification of intraday S&P500 returns with a Random Forest. *International Journal of Forecasting, 35*, 390-407. https://doi.org/10.1016/j.ijforecast.2018.08.004

Long, J., Chen, Z., He, W., Wu, T., & Ren, J. (2020). An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing Journal, 91*. https://doi.org/10.1016/j.asoc.2020.106205

M, H., A, G. E., Menon, V. K., & P, S. K. (2018). NSE Stock Market Prediction Using Deep-Learning Models. *International Conference on Computational Intelligence and Data Science (ICCIDS 2018)*, *132*, pp. 1351-1362. http://10.1016/j.procs.2018.05.050

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems With Applications, 165*(June 2020), 113973. https://doi.org/10.1016/j.eswa.2020.113973

Malagrino, L. S., Roman, N. T., & Monteiro, A. M. (2018). Forecasting stock market index daily direction: A Bayesian Network approach. *Expert Systems With Applications, 105*, 11-22. https://doi.org/10.1016/j.eswa.2018.03.039

Malandri, L., Xing, F. Z., Orsenigo, C., Vercellis, C., & Cambria, E. (2018). Public Mood–Driven Asset Allocation: the Importance of FinancialSentiment in Portfolio Management. *Cognitive Computation, 10*, 1167–1176. https://doi.org/10.1007/s12559-018-9609-2

Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Muhammad, K. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management, 50*, 432-451. https://doi.org/10.1016/j.ijinfomgt.2019.07.011

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91. https://www.jstor.org/stable/2975974

Meher, P., & Mishra, R. K. (2024). Risk-Adjusted Portfolio Optimization: Monte Carlo Simulation and Rebalancing. *Australasian Accounting, Business and Finance Journal, 18*(3), 85-101. https://doi.org/10.14453/aabfj.v18i3.06

Milhomem, D. A., & Dantas, M. J. (2020). Analysis of New Approaches Used in Portfolio Optimization: a Systematic Literature Review. *Production, 30*(1), 1-16. https://doi.org/10.1590/0103-6513.20190144

Mirkin, B. (1996). *Mathematical classification and clustering.* Springer Science and Business Media.

Mirnoor, S. M., & Shariati, A. (2012). Fuzzy portfolio optimization using Chen and Huang model: Evidence from Iranian mutual funds. *African Journal of Business Management, 6*(22), 6608-6616. https://doi.org/10.5897/AJBM12.303

Monsour, N., Cherif, M. S., & Abdelfattah, W. (2019). Multi-objective imprecise programming for financial portfolio selection with fuzzy returns. *Expert Systems With Applications, 138*(112810), 1-15. https://doi.org/10.1016/j.eswa.2019.07.027

Murugan, M., & Kala T, S. (2023). Large-scale data-driven financial risk management & analysis using machine learning strategies. *Measurement: Sensors, 27*, 1-9. Retrieved from https://doi.org/10.1016/j.measen.2023.100756

Nabipour, M., Nayyyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting Stock Market Trends Using MachineLearning and Deep Learning Algorthms Via Continuous and Binary Data;a Comparative Analysis. *IEEE Access, 8*, 150199-150212. 10.1109/ACCESS.2020.3015966

Nanda, S., Mahanty, B., & Tiwari, M. (2010). Clustering Indian stock market data for portfolio management. *Expert Systems with Applications, 37*, 8793-8798. https://doi.org/10.1016/j.eswa.2010.06.026

Nazareth, N., & Reddy, Y. V. (2023). Financial applications of machine learning: A literature review. *Expert Systems With Applications, 219*(119640), 1-33. https://doi.org/10.1016/j.eswa.2023.119640

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications : A survey. *Applied Soft Computing Journal, 93*, 106384. https://doi.org/10.1016/j.asoc.2020.106384

Paiva, F. D., Cardoso, R. T., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems With Applications, 115*, 635-655. https://doi.org/10.1016/j.eswa.2018.08.003

Pascoal, C., Oliveira, M. R., Pacheco, A., & Valadas, R. (2017). Theoretical evaluation of feature selection methods based on mutual information. *Neurocomputing, 226*, 168-181. http://dx.doi.org/10.1016/j.neucom.2016.11.047

Platon, V., & Constantinescu, A. (2014). Monte Carlo Method in risk analysis for investment projects. *Procedia Economics and Finance, 15*, 393-400. https://doi.org/10.1016/S2212-5671(14)00463-8

Pokhrel, N. R., Dahal, K. R., Rimal, R., Bhandari, H. N., Khatri, R. K., Rimal, B., & Hahn, W. E. (2022). Predicting NEPSE index price using deep learning models. *Machine Learning With Applications*, 1-39. https://doi.org/10.1016/j.mlwa.2022.100385

Prasad, P. C., Jaiswal, A., Shakya, S., & Singh, S. (2020). Portfolio Optimization: A Study of Nepal Stock Exchange. *Proceedings of International Conference on Sustainable Expert Systems (ICSES 2020). 176*, pp. 659-672. Springer.

Qiu, M., Song, Y., & Akagi, F. (2016). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. *Chaos, Solitons and Fractals, 85*, 1-7. http://dx.doi.org/10.1016/j.chaos.2016.01.004

Rom, B. M., & Ferguson, K. W. (1994). Post-Modern Portfolio Theory Comes of Age. *The Journal of Investing, 3*(3), 11-17. https://www.pm-research.com/content/iijinvest/3/3/11

Saud, A. S., & Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. *International Conference on Computational Intelligence and Data Science (ICCIDS 2019)*, *167*, pp. 788-798. http://10.1016/j.procs.2020.03.419

Saurabh, S., & Dey, K. (2020). Unraveling the relationship between social moods and the stock. *Journal of Behavioral and Experimental Finance, 26*. https://doi.org/10.1016/j.jbef.2020.100300

Shanmuganathan, M. (2020). Behavioural finance in an era of artificial intelligence: Longitudinal case study of robo-advisors in investment decisions. *Journal of Behavioral and Experimental Finance, 27*(100297). https://doi.org/10.1016/j.jbef.2020.100297

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379-423. Sharaf, M., El-Din Hemdan, E., El-Sayed, A., & El-Bahnasawy, N. A. (2022). A survey on recommendation systems for financial services. *Multimedia Tools and Applications, 81*, 16761-16781. https://doi.org/10.1007/s11042-022-12564-1

Sharma, D. K., Hota, H., Brown, K., & Handa, R. (2022). Integration of genetic algorithm with artificial neural network for stock market forecasting. *International Journal of System Assurance Engineering and Management, 13*, 828-841. https://doi.org/10.1007/s13198-021-01209-5

Shternshis, A., Mazzarisi, P., & Marmi, S. (2022). Measuring market efficiency: The Shannon entropy of high-frequency inancial time series. *Chaos, Solitons and Fractals, 162*, 1-16. https://doi.org/10.1016/j.chaos.2022.112403

Sun, J., Xiao, K., Liu, C., Zhou, W., & Xiong, H. (2019). Exploiting intra-day patterns for market shock prediction: A machine learning approach. *Expert Systems With Applications, 127*, 272-281. https://doi.org/10.1016/j.eswa.2019.03.006

Tan, Z., Yan, Z., & Zhu, G. (2019). Stock selection with random forest: An exploitation of excess return in the Chinese stock market. *Heliyon, 5*(8), e02310.

https://doi.org/10.1016/j.heliyon.2019.e02310
Tobisova, A., Senova, A., & Rozenberg, R. (2022). Model for Sustainable Financial Planning and Investment Financing Using Monte Carlo Method. *Sustainability, 14*. https://doi.org/10.3390/su14148785
Tsai, C.-F., & Hsiao, Y.-C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems, 50*(1), 258-268. https://doi.org/10.1016/j.dss.2010.08.028
Tsai, C.-F., Lin, Y.-C., Yen, D. C., & Chen, Y.-M. (2011). Predicting stock returns by classifier ensembles. *Applied Soft Computing, 11*(2), 2452-2459. https://doi.org/10.1016/j.asoc.2010.10.001
Ünal, H. T., & Başçiftçi, F. (2022). Evolutionary design of neural network architectures: a review of three decades of research. *Artificial Intelligence Review, 55*, 1723-1802. Retrieved from https://doi.org/10.1007/s10462-021-10049-5
Vo, N. N., He, X., Liu, S., & Xu, G. (2019). Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems, 124*(July), 113097. https://doi.org/10.1016/j.dss.2019.113097
Wang, W., Li, W., Zhang, N., & Liu, K. (2020). Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems With Applications, 143*, 113042. https://doi.org/10.1016/j.eswa.2019.113042
Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems With Applications, 79*, 153-163. http://dx.doi.org/10.1016/j.eswa.2017.02.041
Xing, Z. F., Cambria, E., & Zhang, Y. (2019). Sentiment-aware volatility forecasting. *Knowledge-Based Systems, 176*, 68-76. https://doi.org/10.1016/j.knosys.2019.03.029
Yehia, N. (2024, April 07). *Understanding Long Short-Term Memory (LSTM) Networks*. Retrieved from ML ARCHIVE: https://mlarchive.com/deep-learning/understanding-long-short-term-memory-networks/
Zaremba, A., Kizys, R., Tzouvanas, P., Aharon, D. Y., & Demir, E. (2021). The quest for multidimensional financial immunity to the COVID-19 pandemic: Evidence from international stock markets. *Journal of International Financial Markets, Institutions & Money, 71*. https://doi.org/10.1016/j.intfin.2021.101284
Zeng, G. (2015). A Unified Definition of Mutual Information with Applications in Machine Learning. *Mathematical Problems in Engineering, 2015*, 1-29. https://dx.doi.org/10.1155/2015/201874
Zhang, L., Chen, C., Bu, J., & He, X. (2012). A Unified Feature and Instance Selection Framework Using Optimum Experimental Design. *IEEE Transactions on Image Processing, 21*(5), 2379-2388. https://ieeexplore.ieee.org/document/6129509
Zhang, Y., Chu, G., & Shen, D. (2021). The role of investor attention in predicting stock prices: The long short-term memory networks perspective. *Finance Research Letters, 38*. https://doi.org/10.1016/j.frl.2020.101484

Annexure

List of companies from Nifty 100 index

		Dataset 1:	Dataset 2:	Dataset 3:	Dataset 4:	Dataset 5:	Dataset 6:	Dataset 7:
	Events:	**Demonetization**	**Implementation of GST**	**General Elections 2019**	**COVID-19**	**Ukraine's invasion by Russia**	**Hamas attack on Israel**	**General Elections 2024**
Sr. No.	**Constituents of Nifty100 as on:**	**30-Sep-16**	**26-May-17**	**29-Mar-19**	**27-Sep-19**	**30-Sep-21**	**13-Jul-23**	**28-Mar-24**
1	A B B India Ltd.	01-04-2016	01-04-2016	01-04-2016	xxx	xxx	31-03-2023	31-03-2023
2	A C C Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	xxx
3	Adani Energy Solutions Ltd.	xxx	xxx	xxx	xxx	19-03-2020	19-03-2020	19-03-2020
4	Adani Enterprises Ltd.	xxx	xxx	xxx	xxx	31-03-2021	31-03-2021	31-03-2021
5	Adani Green Energy Ltd.	xxx	xxx	xxx	xxx	25-09-2020	25-09-2020	25-09-2020
6	Adani Ports & Special Economic Zone Ltd.	10-09-2008	10-09-2008	10-09-2008	10-09-2008	10-09-2008	10-09-2008	10-09-2008
7	Adani Power Ltd.	xxx	xxx	xxx	xxx	xxx	xxx	28-03-2024
8	Adani Total Gas Ltd.	xxx	xxx	xxx	xxx	xxx	30-09-2022	30-09-2022
9	Adani Wilmar Ltd.	xxx	xxx	xxx	xxx	xxx	31-03-2023	xxx
10	Ambuja Cements Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
11	Apollo Hospitals Enterprise Ltd.	28-09-2012	xxx	xxx	xxx	31-03-2021	31-03-2021	31-03-2021
12	Ashok Leyland Ltd.	28-09-2015	28-09-2015	28-09-2015	28-09-2015	xxx	xxx	xxx

13	Asian Paints Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
14	Aurobindo Pharma Ltd.	19-09-2014	19-09-2014	19-09-2014	19-09-2014	19-09-2014	xxx	xxx
15	Avenue Supermarts Ltd.	xxx	xxx	29-09-2017	29-09-2017	29-09-2017	29-09-2017	29-09-2017
16	Axis Bank Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
17	Bajaj Auto Ltd.	22-10-2009	22-10-2009	22-10-2009	22-10-2009	22-10-2009	22-10-2009	22-10-2009
18	Bajaj Finance Ltd.	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016
19	Bajaj Finserv Ltd.	01-04-2013	01-04-2013	01-04-2013	01-04-2013	01-04-2013	01-04-2013	01-04-2013
20	Bajaj Holdings & Invst. Ltd.	xxx	xxx	29-03-2019	29-03-2019	29-03-2019	29-03-2019	29-03-2019
21	Bandhan Bank Ltd.	xxx	xxx	28-09-2018	28-09-2018	28-09-2018	xxx	xxx
22	Bank Of Baroda	01-12-2005	01-12-2005	01-12-2005	01-12-2005	30-09-2021	30-09-2021	30-09-2021
23	Berger Paints India Ltd.	xxx	xxx	xxx	27-09-2019	27-09-2019	27-09-2019	27-09-2019
24	Bharat Electronics Ltd.	01-04-2016	01-04-2016	xxx	xxx	xxx	30-09-2022	30-09-2022
25	Bharat Forge Ltd.	01-12-2005		xxx	xxx	xxx	xxx	xxx
26	Bharat Heavy Electricals Ltd.	01-12-2005	01-12-2005	01-12-2005	xxx	xxx	xxx	xxx
27	Bharat Petroleum Corpn. Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
28	Bharti Airtel Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
29	Biocon Ltd.	xxx	xxx	28-09-2018	28-09-2018	28-09-2018	xxx	xxx

30	Bosch Ltd.	10-10-2011	10-10-2011	10-10-2011	10-10-2011	10-10-2011	10-10-2011	10-10-2011
31	Britannia Industries Ltd.	29-05-2015	29-05-2015	29-05-2015	29-05-2015	29-05-2015	29-05-2015	29-05-2015
32	Cairn India Ltd. [Merged]	12-12-2007	xxx	xxx	xxx	xxx	xxx	xxx
33	Canara Bank	xxx	xxx	xxx	xxx	xxx	31-03-2023	31-03-2023
34	Castrol India Ltd.	01-04-2016	xxx	xxx	xxx	xxx	xxx	xxx
35	Cholamandalam Investment & Finance Co. Ltd.	xxx	xxx	xxx	xxx	30-09-2021	30-09-2021	30-09-2021
36	Cipla Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
37	Coal India Ltd.	21-04-2011	21-04-2011	21-04-2011	21-04-2011	21-04-2011	21-04-2011	21-04-2011
38	Colgate-Palmolive (India) Ltd.	22-10-2009	22-10-2009	22-10-2009	22-10-2009	22-10-2009	22-10-2009	22-10-2009
39	Container Corpn. Of India Ltd.	02-03-2006	02-03-2006	02-03-2006	02-03-2006	xxx	xxx	xxx
40	Cummins India Ltd.	01-12-2005	01-12-2005	xxx	xxx	xxx	xxx	xxx
41	D L F Ltd.	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016
42	Dabur India Ltd.	10-10-2011	10-10-2011	10-10-2011	10-10-2011	10-10-2011	10-10-2011	10-10-2011
43	Divi'S Laboratories Ltd.	27-04-2012	27-04-2012	29-03-2019	29-03-2019	29-03-2019	29-03-2019	29-03-2019
44	Dr. Reddy'S Laboratories Ltd.	29-12-2008	29-12-2008	29-12-2008	29-12-2008	29-12-2008	29-12-2008	29-12-2008
45	Eicher Motors Ltd.	02-02-2015	02-02-2015	02-02-2015	02-02-2015	02-02-2015	02-02-2015	02-02-2015
46	Emami Ltd.	01-04-2016	01-04-2016	xxx	xxx	xxx	xxx	xxx

47	F S N E-Commerce Ventures Ltd.	xxx	xxx	xxx	xxx	xxx	31-03-2022	xxx
48	G A I L (India) Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
49	General Insurance Corpn. Of India	xxx	xxx	02-04-2018	02-04-2018	xxx	xxx	xxx
50	Gland Pharma Ltd.	xxx	xxx	xxx	xxx	30-06-2021	xxx	xxx
51	Glaxosmithkline Consumer Healthcare Ltd. [Merged]	27-04-2012	27-04-2012	xxx	xxx	xxx	xxx	xxx
52	Glaxosmithkline Pharmaceuticals Ltd.	27-03-2009	27-03-2009	xxx	xxx	xxx	xxx	xxx
53	Glenmark Pharmaceuticals Ltd.	12-01-2009	12-01-2009	xxx	xxx	xxx	xxx	xxx
54	Godrej Consumer Products Ltd.	28-09-2012	28-09-2012	28-09-2012	28-09-2012	28-09-2012	28-09-2012	28-09-2012
55	Grasim Industries Ltd.	01-10-2010		02-04-2018	02-04-2018	02-04-2018	02-04-2018	02-04-2018
56	H C L Technologies Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
57	H D F C Asset Mgmt. Co. Ltd.	xxx	xxx	29-03-2019	29-03-2019	29-03-2019	29-03-2019	
58	H D F C Bank Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
59	H D F C Life Insurance Co. Ltd.	xxx	xxx	28-09-2018	28-09-2018	28-09-2018	28-09-2018	28-09-2018
60	Havells India Ltd.	xxx	15-11-2016	15-11-2016	15-11-2016	15-11-2016	15-11-2016	15-11-2016
61	Hero Motocorp Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
62	Hindalco Industries Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
63	Hindustan Aeronautics Ltd.	xxx	xxx	xxx	xxx	xxx	30-09-2022	30-09-2022

64	Hindustan Petroleum Corpn. Ltd.	17-06-2009	17-06-2009	17-06-2009	17-06-2009	17-06-2009	xxx	xxx
65	Hindustan Unilever Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
66	Hindustan Zinc Ltd.	01-04-2016	01-04-2016	01-04-2016	01-04-2016	xxx	xxx	xxx
67	Housing Development Finance Corpn. Ltd. [Merged]	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	xxx	xxx
68	I C I C I Bank Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
69	I C I C I Lombard General Insurance Co. Ltd.	xxx	xxx	28-09-2018	28-09-2018	28-09-2018	28-09-2018	28-09-2018
70	I C I C I Prudential Life Insurance Co. Ltd.	xxx	31-03-2017	31-03-2017	31-03-2017	31-03-2017	31-03-2017	31-03-2017
71	I T C Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
72	Indian Oil Corpn. Ltd.	19-10-2015	19-10-2015	19-10-2015	19-10-2015	19-10-2015	19-10-2015	19-10-2015
73	Indian Railway Catering & Tourism Corpn. Ltd.	xxx	xxx	xxx	xxx	xxx	30-09-2022	30-09-2022
74	Indian Railway Finance Corpn. Ltd.	xxx	xxx	xxx	xxx	xxx	xxx	28-03-2024
75	Indraprastha Gas Ltd.	xxx	xxx	xxx	xxx	26-06-2020	xxx	xxx
76	Indus Towers Ltd.	28-03-2014	28-03-2014	28-03-2014	28-03-2014	28-03-2014	28-03-2014	xxx
77	Indusind Bank Ltd.	25-03-2011	25-03-2011	25-03-2011	25-03-2011	25-03-2011	25-03-2011	25-03-2011
78	Info Edge (India) Ltd.	xxx	xxx	xxx	xxx	26-06-2020	26-06-2020	26-06-2020
79	Infosys Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005

80	Interglobe Aviation Ltd.	30-09-2016	30-09-2016	30-09-2016	30-09-2016	30-09-2016	30-09-2016	30-09-2016
81	J S W Steel Ltd.	23-06-2008	23-06-2008	23-06-2008	23-06-2008	23-06-2008	23-06-2008	23-06-2008
82	Jindal Steel & Power Ltd.	xxx	xxx	xxx	xxx	30-09-2021	13-07-2023	13-07-2023
83	Jio Financial Services Ltd.	xxx	xxx	xxx	xxx	xxx	xxx	28-03-2024
84	Jubilant Foodworks Ltd.	xxx	xxx	xxx	xxx	31-03-2021	xxx	xxx
85	Kotak Mahindra Bank Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
86	L & T Finance Ltd.	xxx	xxx	02-04-2018	02-04-2018	xxx	xxx	xxx
87	L I C Housing Finance Ltd.	01-12-2005	01-12-2005	xxx	xxx	xxx	xxx	xxx
88	Larsen & Toubro Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
89	Life Insurance Corpn. Of India	xxx	xxx	xxx	xxx	xxx	08-08-2022	08-08-2022
90	Ltimindtree Ltd.	xxx	xxx	xxx	xxx	25-09-2020	25-09-2020	25-09-2020
91	Lupin Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	xxx	xxx
92	M R F Ltd.	xxx	xxx	29-09-2017	xxx	xxx	xxx	xxx
93	Mahindra & Mahindra Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
94	Marico Ltd.	28-09-2015	28-09-2015	28-09-2015	28-09-2015	28-09-2015	28-09-2015	28-09-2015
95	Maruti Suzuki India Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
96	Muthoot Finance Ltd.	xxx	xxx	xxx	xxx	26-06-2020	26-06-2020	xxx

97	N H P C Ltd.	01-04-2016	01-04-2016	01-04-2016	01-04-2016	xxx	xxx	xxx
98	N M D C Ltd.	01-04-2013	01-04-2013	01-04-2013	01-04-2013	01-04-2013	xxx	xxx
99	N T P C Ltd.	24-09-2007	24-09-2007	24-09-2007	24-09-2007	24-09-2007	24-09-2007	24-09-2007
100	Nestle India Ltd.	xxx	xxx	xxx	27-09-2019	27-09-2019	27-09-2019	27-09-2019
101	New India Assurance Co. Ltd.	xxx	xxx	28-09-2018	28-09-2018	xxx	xxx	xxx
102	Oil & Natural Gas Corpn. Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
103	Oil India Ltd.	27-09-2013	27-09-2013	xxx	xxx	xxx	xxx	xxx
104	Oracle Financial Services Software Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	xxx	xxx	xxx
105	P I Industries Ltd.	xxx	xxx	xxx	xxx	30-09-2021	30-09-2021	xxx
106	Page Industries Ltd.	xxx	xxx	29-03-2019	29-03-2019	xxx	31-03-2023	xxx
107	Petronet L N G Ltd.	xxx	31-03-2017	31-03-2017	31-03-2017	xxx	xxx	xxx
108	Pidilite Industries Ltd.	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016
109	Piramal Enterprises Ltd.	30-09-2016	30-09-2016	30-09-2016	30-09-2016	30-09-2016	xxx	xxx
110	Power Finance Corpn. Ltd.	12-12-2007	12-12-2007	xxx	27-09-2019	xxx	xxx	28-03-2024
111	Power Grid Corpn. Of India Ltd.	14-03-2008	14-03-2008	14-03-2008	14-03-2008	14-03-2008	14-03-2008	14-03-2008
112	Procter & Gamble Hygiene & Health Care Ltd.	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	xxx
113	Punjab National Bank	01-12-2005	01-12-2005	xxx	27-09-2019	27-09-2019	xxx	29-09-2023

114	R E C Ltd.		31-03-2017	xxx	xxx	xxx	xxx	28-03-2024
115	Reliance Industries Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
116	S B I Cards & Payment Services Ltd.	xxx	xxx	xxx	xxx	31-07-2020	31-07-2020	31-07-2020
117	S B I Life Insurance Co. Ltd.	xxx	xxx	02-04-2018	02-04-2018	02-04-2018	02-04-2018	02-04-2018
118	S R F Ltd.	xxx	xxx	xxx	xxx	xxx	31-03-2022	31-03-2022
119	Sammaan Capital Ltd.	27-03-2015	27-03-2015	27-03-2015	27-03-2015	xxx	xxx	xxx
120	Samvardhana Motherson Intl. Ltd.	19-09-2014	19-09-2014	19-09-2014	19-09-2014	xxx	30-09-2022	30-09-2022
121	Shree Cement Ltd.	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016
122	Shriram Finance Ltd.	01-10-2010	01-10-2010	01-10-2010	01-10-2010			29-09-2023
123	Siemens Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
124	State Bank Of India	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
125	Steel Authority Of India Ltd.	01-12-2005	01-12-2005	01-12-2005		30-09-2021		
126	Sun Pharmaceutical Inds. Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
127	Sun T V Network Ltd.	xxx	26-05-2017	xxx	xxx	xxx	xxx	xxx
128	T V S Motor Co. Ltd.	xxx	xxx	xxx	xxx	xxx	xxx	29-09-2023
129	Tata Consultancy Services Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
130	Tata Consumer Products Ltd.	xxx	xxx	xxx	xxx	25-09-2020	25-09-2020	25-09-2020

131	Tata Motors Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
132	Tata Power Co. Ltd.	01-12-2005	01-12-2005	xxx	xxx	xxx	08-08-2022	08-08-2022
133	Tata Steel Ltd.	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005	01-12-2005
134	Tech Mahindra Ltd.	04-04-2007	04-04-2007	04-04-2007	04-04-2007	04-04-2007	04-04-2007	04-04-2007
135	Titan Company Ltd.	25-03-2011	25-03-2011	25-03-2011	25-03-2011	25-03-2011	25-03-2011	25-03-2011
136	Torrent Pharmaceuticals Ltd.	01-04-2016	01-04-2016	xxx	xxx	26-06-2020	26-06-2020	26-06-2020
137	Trent Ltd.	xxx	xxx	xxx	xxx	xxx	xxx	29-09-2023
138	U P L Ltd.	22-10-2009	22-10-2009	22-10-2009	22-10-2009	22-10-2009	22-10-2009	xxx
139	Ultratech Cement Ltd.	04-04-2007	04-04-2007	04-04-2007	04-04-2007	04-04-2007	04-04-2007	04-04-2007
140	United Breweries Ltd.	01-04-2013	01-04-2013	29-03-2019	29-03-2019	xxx	xxx	xxx
141	United Spirits Ltd.	27-03-2015	27-03-2015	27-03-2015	27-03-2015	27-03-2015	27-03-2015	27-03-2015
142	Varun Beverages Ltd.	xxx	xxx	xxx	xxx	xxx	31-03-2023	31-03-2023
143	Vedanta Ltd.	27-03-2009	27-03-2009	27-03-2009	27-03-2009	31-03-2021	31-03-2021	31-03-2021
144	Vodafone Idea Ltd.	10-10-2011	10-10-2011	10-10-2011	10-10-2011	xxx	xxx	xxx
145	Wipro Ltd.	27-09-2013	27-09-2013	27-09-2013	27-09-2013	27-09-2013	27-09-2013	27-09-2013
146	Yes Bank Ltd.	01-10-2010	01-10-2010	01-10-2010	01-10-2010	31-03-2021	xxx	xxx
147	Zee Entertainment Enterprises Ltd.	25-03-2011	25-03-2011	25-03-2011	25-03-2011	xxx	xxx	xxx

148	Zomato Ltd.	xxx	xxx	xxx	xxx	xxx	31-03-2022	31-03-2022
149	Zydus Lifesciences Ltd.	01-04-2016	01-04-2016	01-04-2016	01-04-2016	01-04-2016	xxx	29-09-2023

Outliers on dataset 1: **Demonetization**

Sr. No.	Company Name
1	Bajaj Finserv Ltd.
2	Bharat Petroleum Corpn. Ltd.
3	Marico Ltd.
4	Shriram Finance Ltd.
5	Vodafone Idea Ltd.

Cluster Formation on dataset 1: **Demonetization**

Company Name	Company Name	Company Name
Cluster 1		
Asian Paints Ltd.	Britannia Industries Ltd.	Castrol India Ltd.
Colgate-Palmolive (India) Ltd.	Dabur India Ltd.	Eicher Motors Ltd.
Emami Ltd.	Pidilite Industries Ltd.	Hindustan Unilever Ltd.
Procter & Gamble Hygiene & Health Care Ltd.	Godrej Consumer Products Ltd.	United Breweries Ltd.
Cluster 2		
A B B India Ltd.	Grasim Industries Ltd.	Oil & Natural Gas Corpn. Ltd.
A C C Ltd.	Hero Motocorp Ltd.	Ambuja Cements Ltd.
Power Grid Corpn. Of India Ltd.	Housing Development Finance Corpn. Ltd. [Merged]	Samvardhana Motherson Intl. Ltd.
Aurobindo Pharma Ltd.	Indian Oil Corpn. Ltd.	Reliance Industries Ltd.
Bajaj Auto Ltd.	Kotak Mahindra Bank Ltd.	Shree Cement Ltd.
Bharat Electronics Ltd.	Mahindra & Mahindra Ltd.	Tata Power Co. Ltd.
Bosch Ltd.	Maruti Suzuki India Ltd.	Titan Company Ltd.
Cairn India Ltd. [Merged]	N H P C Ltd.	Torrent Pharmaceuticals Ltd.
D L F Ltd.	N M D C Ltd.	U P L Ltd.
Divi'S Laboratories Ltd.	N T P C Ltd.	Ultratech Cement Ltd.
G A I L (India) Ltd.		
Cluster 3		
Hindustan Petroleum Corpn. Ltd.	Power Finance Corpn. Ltd.	Zydus Lifesciences Ltd.
Cluster 4		
Hindalco Industries Ltd.	Piramal Enterprises Ltd.	Tata Steel Ltd.
Hindustan Zinc Ltd.	Vedanta Ltd.	
Cluster 5		
Adani Ports & Special Economic Zone Ltd.	Glaxosmithkline Consumer Healthcare Ltd. [Merged]	Glaxosmithkline Pharmaceuticals Ltd.
Apollo Hospitals Enterprise Ltd.	Glenmark Pharmaceuticals Ltd.	Oracle Financial Services Software Ltd.

Container Corpn. Of India Ltd.	Tata Consultancy Services Ltd.	Infosys Ltd.
Ashok Leyland Ltd.	Dr. Reddy'S Laboratories Ltd.	Oil India Ltd.
Bharat Forge Ltd.	Lupin Ltd.	Larsen & Toubro Ltd.
Bharti Airtel Ltd.	H C L Technologies Ltd.	Siemens Ltd.
Cipla Ltd.	I T C Ltd.	Wipro Ltd.
Coal India Ltd.	Indus Towers Ltd.	Tech Mahindra Ltd.
Cummins India Ltd.		

Outliers on dataset 2: **GST dataset**

Sr. No.	**Company Name**
1	Bajaj Finserv Ltd.
2	Divi’s Laboratories Ltd.
3	Glaxosmithkline Pharmaceuticals Ltd.
4	Shriram Finance Ltd.

Cluster Formation on dataset 2**: GST dataset**

Company Name	**Company Name**	**Company Name**
Cluster 1		
Tata Consultancy Services Ltd.	Zee Entertainment Enterprises Ltd.	Oracle Financial Services Software Ltd.
A C C Ltd.	G A I L (India) Ltd.	Pidilite Industries Ltd.
Ambuja Cements Ltd.	H C L Technologies Ltd.	Siemens Ltd.
Bajaj Auto Ltd.	Hero Motocorp Ltd.	A B B India Ltd.
Bharti Airtel Ltd.	Indus Towers Ltd.	Tata Power Co. Ltd.
Bosch Ltd.	Interglobe Aviation Ltd.	United Breweries Ltd.
Cipla Ltd.	Larsen & Toubro Ltd.	United Spirits Ltd.
Cummins India Ltd.	Mahindra & Mahindra Ltd.	Wipro Ltd.
Dabur India Ltd.	N T P C Ltd.	Emami Ltd.
Cluster 2		
Shree Cement Ltd.	D L F Ltd.	N M D C Ltd.
Tata Steel Ltd.	Havells India Ltd.	Petronet L N G Ltd.
Titan Company Ltd.	Hindustan Zinc Ltd.	Reliance Industries Ltd.
Power Grid Corpn. Of India Ltd.	Housing Development Finance Corpn. Ltd. [Merged]	Samvardhana Motherson Intl. Ltd.
Ultratech Cement Ltd.	Kotak Mahindra Bank Ltd.	U P L Ltd.
Zydus Lifesciences Ltd.	N H P C Ltd.	
Cluster 3		
Container Corpn. Of India Ltd.	Glaxosmithkline Consumer Healthcare Ltd. [Merged]	Glenmark Pharmaceuticals Ltd.

Aurobindo Pharma Ltd.	Ashok Leyland Ltd.	Oil India Ltd.
Coal India Ltd.	I T C Ltd.	Tech Mahindra Ltd.
Oil & Natural Gas Corpn. Ltd.	Infosys Ltd.	Torrent Pharmaceuticals Ltd.
Dr. Reddy'S Laboratories Ltd.	Lupin Ltd.	
Cluster 4		
Maruti Suzuki India Ltd.	Adani Ports & Special Economic Zone Ltd.	Sun T V Network Ltd.
Piramal Enterprises Ltd.	Hindalco Industries Ltd.	Vedanta Ltd.
Cluster 5		
J S W Steel Ltd.	Bharat Electronics Ltd.	
Cluster 6		
Godrej Consumer Products Ltd.	Asian Paints Ltd.	Procter & Gamble Hygiene & Health Care Ltd.
Marico Ltd.	Britannia Industries Ltd.	Hindustan Unilever Ltd.
Eicher Motors Ltd.	Colgate-Palmolive (India) Ltd.	
Cluster 7		
Power Finance Corpn. Ltd.	Bharat Petroleum Corpn. Ltd.	Indian Oil Corpn. Ltd.
Hindustan Petroleum Corpn. Ltd.		

Outliers on dataset 3: **General Elections 2019**

Sr. No.	Company Name
1	Bajaj Finserv Ltd.
2	Grasim Industries Ltd.
3	Biocon Ltd.
4	Shriram Finance Ltd.
5	Kotak Mahindra Bank Ltd.
6	Hindustan Unilever Ltd.
7	Coal India Ltd.

Cluster Formation on dataset 3: **General Elections 2019**

Company Name	Company Name	Company Name
Cluster 1		
A B B India Ltd.	Hindalco Industries Ltd.	N M D C Ltd.
A C C Ltd.	Hindustan Zinc Ltd.	N T P C Ltd.
Adani Ports & Special Economic Zone Ltd.	Housing Development Finance Corpn. Ltd. [Merged]	Oracle Financial Services Software Ltd.
Ambuja Cements Ltd.	Indian Oil Corpn. Ltd.	Oil & Natural Gas Corpn. Ltd
Bajaj Auto Ltd.	Indus Towers Ltd.	Petronet L N G Ltd.
Container Corpn. Of India Ltd.	Godrej Consumer Products Ltd.	Power Grid Corpn. Of India Ltd.

Bosch Ltd.	L & T Finance Ltd.	Siemens Ltd.
Cipla Ltd.	Larsen & Toubro Ltd.	Sun Pharmaceutical Inds. Ltd.
J S W Steel Ltd.	Lupin Ltd.	Tata Steel Ltd.
D L F Ltd.	M R F Ltd.	Ultratech Cement Ltd.
G A I L (India) Ltd.	Mahindra & Mahindra Ltd.	United Spirits Ltd.
Bharat Petroleum Corpn. Ltd.	Marico Ltd.	Zydus Lifesciences Ltd.
H C L Technologies Ltd.	N H P C Ltd.	
Cluster 2		
Asian Paints Ltd.	Dr. Reddy'S Laboratories Ltd.	Reliance Industries Ltd.
Aurobindo Pharma Ltd.	Havells India Ltd.	Shree Cement Ltd.
Avenue Supermarts Ltd.	Procter & Gamble Hygiene & Health Care Ltd.	Tata Consultancy Services Ltd.
Bajaj Holdings & Invst. Ltd.	Infosys Ltd.	Tech Mahindra Ltd.
Britannia Industries Ltd.	Page Industries Ltd.	Titan Company Ltd.
Colgate-Palmolive (India) Ltd.	Pidilite Industries Ltd.	U P L Ltd.
Dabur India Ltd.	Piramal Enterprises Ltd.	United Breweries Ltd.
Divi'S Laboratories Ltd.	I T C Ltd.	Wipro Ltd.
Cluster 3		
Ashok Leyland Ltd.	Maruti Suzuki India Ltd.	Tata Motors Ltd.
Eicher Motors Ltd.	Sammaan Capital Ltd.	Vedanta Ltd.
Hindustan Petroleum Corpn. Ltd.	Samvardhana Motherson Intl. Ltd.	Zee Entertainment Enterprises Ltd.
Hero Motocorp Ltd.	Steel Authority Of India Ltd.	

Outliers on dataset 4: **COVID19**

Sr. No.	Company Name
1	Bajaj Finserv Ltd.
2	D L F Ltd.
3	Larsen & Toubro Ltd.
4	Shriram Finance Ltd.
5	Kotak Mahindra Bank Ltd.
6	Nestle India Ltd.

Cluster Formation on dataset 4: **COVID19**

Company Name	Company Name	Company Name
Cluster 1		
A C C Ltd.	Dr. Reddy'S Laboratories Ltd.	Pidilite Industries Ltd.
Asian Paints Ltd.	Eicher Motors Ltd.	Ambuja Cements Ltd.
Adani Ports & Special Economic Zone Ltd.	Housing Development Finance Corpn. Ltd. [Merged]	Procter & Gamble Hygiene & Health Care Ltd.

Ashok Leyland Ltd.	Grasim Industries Ltd.	Reliance Industries Ltd.
Power Grid Corpn. Of India Ltd.	Hindustan Petroleum Corpn. Ltd.	Samvardhana Motherson Intl. Ltd.
Avenue Supermarts Ltd.	Hindustan Unilever Ltd.	Shree Cement Ltd.
Bajaj Auto Ltd.	Godrej Consumer Products Ltd.	Tata Consultancy Services Ltd.
Bajaj Holdings & Invst. Ltd.	Infosys Ltd.	Sun Pharmaceutical Inds. Ltd.
Berger Paints India Ltd.	Interglobe Aviation Ltd.	Siemens Ltd.
Bharat Petroleum Corpn. Ltd.	Maruti Suzuki India Ltd.	Tech Mahindra Ltd.
Britannia Industries Ltd.	N H P C Ltd.	Titan Company Ltd.
Colgate-Palmolive (India) Ltd.	N M D C Ltd.	Ultratech Cement Ltd.
Dabur India Ltd.	Page Industries Ltd.	United Spirits Ltd.
Divi'S Laboratories Ltd.	Petronet L N G Ltd.	
Cluster 2		
Oracle Financial Services Software Ltd.	Container Corpn. Of India Ltd.	Zee Entertainment Enterprises Ltd.
Piramal Enterprises Ltd.	Bosch Ltd.	L & T Finance Ltd.
Sammaan Capital Ltd.	Cipla Ltd.	Lupin Ltd.
Tata Steel Ltd.	Coal India Ltd.	Mahindra & Mahindra Ltd.
U P L Ltd.	J S W Steel Ltd.	Marico Ltd.
United Breweries Ltd.	G A I L (India) Ltd.	N T P C Ltd.
Vedanta Ltd.	H C L Technologies Ltd.	Oil & Natural Gas Corpn. Ltd.
Hindustan Zinc Ltd.	Havells India Ltd.	Wipro Ltd.
I T C Ltd.	Hero Motocorp Ltd.	Biocon Ltd.
Indian Oil Corpn. Ltd.	Hindalco Industries Ltd.	Zydus Lifesciences Ltd.
Indus Towers Ltd.		

Outliers on dataset 5: **Ukraine-Russia Dataset**

Sr. No.	Company Name
1	Bajaj Finserv Ltd.
2	Adani Enterprises Ltd.
3	Adani Green Energy Ltd.
4	Biocon Ltd.
5	Kotak Mahindra Bank Ltd.
6	Cholamandalam Investment & Finance Co. Ltd.

Cluster Formation on dataset 5: **Ukraine-Russia Dataset**

Company Name	Company Name	Company Name
Cluster 1		

A C C Ltd.	Tata Consultancy Services Ltd.	Hindustan Petroleum Corpn. Ltd.
Asian Paints Ltd.	Shree Cement Ltd.	Indian Oil Corpn. Ltd.
Bosch Ltd.	Sun Pharmaceutical Inds. Ltd.	Infosys Ltd.
Cipla Ltd.	Reliance Industries Ltd.	Marico Ltd.
Divi'S Laboratories Ltd.	Tata Consumer Products Ltd.	N T P C Ltd.
H C L Technologies Ltd.	Torrent Pharmaceuticals Ltd.	Pidilite Industries Ltd.
Ultratech Cement Ltd.		
Cluster 2		
Ltimindtree Ltd.	Adani Ports & Special Economic Zone Ltd.	Apollo Hospitals Enterprise Ltd.
Steel Authority Of India Ltd.	Hindalco Industries Ltd.	J S W Steel Ltd.
Tata Steel Ltd.	D L F Ltd.	Jindal Steel & Power Ltd.
Vedanta Ltd.	Gland Pharma Ltd.	
Cluster 3		
Aurobindo Pharma Ltd.	Dr. Reddy'S Laboratories Ltd.	I T C Ltd.
Bajaj Auto Ltd.	Eicher Motors Ltd.	Indraprastha Gas Ltd.
Berger Paints India Ltd.	Hero Motocorp Ltd.	Lupin Ltd.
Bharat Petroleum Corpn. Ltd.	Hindustan Unilever Ltd.	Maruti Suzuki India Ltd.
Britannia Industries Ltd.	Housing Development Finance Corpn. Ltd. [Merged]	Zydus Lifesciences Ltd.
Colgate-Palmolive (India) Ltd.	Dabur India Ltd.	
Cluster 4		
Havells India Ltd.	Ambuja Cements Ltd.	Titan Company Ltd.
Oil & Natural Gas Corpn. Ltd.	Avenue Supermarts Ltd.	U P L Ltd.
Siemens Ltd.	Bajaj Holdings & Invst. Ltd.	United Spirits Ltd.
Tech Mahindra Ltd.	Grasim Industries Ltd.	Wipro Ltd.
Cluster 5		
Procter & Gamble Hygiene & Health Care Ltd.	Nestle India Ltd.	
Cluster 6		
G A I L (India) Ltd.	Coal India Ltd.	Mahindra & Mahindra Ltd.
Godrej Consumer Products Ltd.	S B I Cards & Payment Services Ltd.	Power Grid Corpn. Of India Ltd.
Indus Towers Ltd.	P I Industries Ltd.	N M D C Ltd.
Jubilant Foodworks Ltd.	Muthoot Finance Ltd.	

Outliers on dataset 6: **Hamas-Israel dataset**

Sr. No.	Company Name

1	Bajaj Finserv Ltd.
2	Bharat Petroleum Corpn. Ltd.
3	Marico Ltd.
4	Shriram Finance Ltd.
5	Vodafone Idea Ltd.

Cluster Formation on dataset 6: **Hamas Israel**

Company Name	**Company Name**	**Company Name**
Cluster 1		
Divi's Laboratories Ltd.	A C C Ltd.	Vedanta Ltd.
Eicher Motors Ltd.	Apollo Hospitals Enterprise Ltd.	Wipro Ltd.
G A I L (India) Ltd.	Asian Paints Ltd.	Infosys Ltd.
Godrej Consumer Products Ltd.	Power Grid Corpn. Of India Ltd.	S B I Cards & Payment Services Ltd.
Grasim Industries Ltd.	Berger Paints India Ltd.	Reliance Industries Ltd.
Havells India Ltd.	Bharat Petroleum Corpn. Ltd.	A B B India Ltd.
Hero Motocorp Ltd.	Bosch Ltd.	Tata Steel Ltd.
Hindalco Industries Ltd.	Coal India Ltd.	Tech Mahindra Ltd.
Hindustan Unilever Ltd.	Dabur India Ltd.	Titan Company Ltd.
Indian Railway Catering & Tourism Corpn. Ltd.	Tata Consultancy Services Ltd.	Torrent Pharmaceuticals Ltd.
Indus Towers Ltd.	Shree Cement Ltd.	United Spirits Ltd.
Info Edge (India) Ltd.	Avenue Supermarts Ltd.	Bajaj Auto Ltd.
Kotak Mahindra Bank Ltd.	Tata Consumer Products Ltd.	Bharti Airtel Ltd.
Marico Ltd.	Tata Power Co. Ltd.	J S W Steel Ltd.
Maruti Suzuki India Ltd.	U P L Ltd.	Ltimindtree Ltd.
P I Industries Ltd.	S R F Ltd.	Sun Pharmaceutical Inds. Ltd.
Pidilite Industries Ltd.		
Cluster 2		
Adani Total Gas Ltd.	Adani Energy Solutions Ltd.	Bajaj Finserv Ltd.
Cluster 3		
F S N E-Commerce Ventures Ltd.		
Cluster 4		
Ambuja Cements Ltd.	Adani Ports & Special Economic Zone Ltd.	Samvardhana Motherson Intl. Ltd.
Varun Beverages Ltd.	Adani Enterprises Ltd.	
Cluster 5		
Siemens Ltd.	H C L Technologies Ltd.	Indian Oil Corpn. Ltd.

Ultratech Cement Ltd.	I T C Ltd.	Dr. Reddy'S Laboratories Ltd.
Tata Motors Ltd.	Bajaj Holdings & Invst. Ltd.	Larsen & Toubro Ltd.
Cipla Ltd.	Mahindra & Mahindra Ltd.	Oil & Natural Gas Corpn. Ltd.
D L F Ltd.	N T P C Ltd.	
Cluster 6		
Nestle India Ltd.	Britannia Industries Ltd.	Colgate-Palmolive (India) Ltd.
Procter & Gamble Hygiene & Health Care Ltd.		
Cluster 7		
Bharat Electronics Ltd.	Adani Wilmar Ltd.	

Outliers on dataset 7: **General Elections 2024**

Sr. No.	Company Name
1	Interglobe Aviation Ltd.
2	Nestle India Ltd.
3	R E C Ltd.
4	Shriram Finance Ltd.
5	Sun Pharmaceutical Inds. Ltd.

Cluster Formation on dataset 7: **General Elections 2024**

Company Name	Company Name	Company Name
Cluster 1		
Oil & Natural Gas Corpn. Ltd.	Adani Power Ltd.	Hero Motocorp Ltd.
T V S Motor Co. Ltd.	Bajaj Auto Ltd.	Hindalco Industries Ltd.
Tata Motors Ltd.	Bharat Electronics Ltd.	Hindustan Aeronautics Ltd.
Tata Power Co. Ltd.	Bharat Petroleum Corpn. Ltd.	Indian Oil Corpn. Ltd.
Tata Steel Ltd.	Cipla Ltd.	Mahindra & Mahindra Ltd.
Vedanta Ltd.	Coal India Ltd.	N T P C Ltd.
Zydus Lifesciences Ltd.	G A I L (India) Ltd.	
Cluster 2		
Indian Railway Catering & Tourism Corpn. Ltd.	Samvardhana Motherson Intl. Ltd.	Apollo Hospitals Enterprise Ltd.
Info Edge (India) Ltd.	Adani Enterprises Ltd.	Torrent Pharmaceuticals Ltd.
Larsen & Toubro Ltd.	Ambuja Cements Ltd.	United Spirits Ltd.
Pidilite Industries Ltd.	Havells India Ltd.	Varun Beverages Ltd.
Titan Company Ltd.	Avenue Supermarts Ltd.	Colgate-Palmolive (India) Ltd.
Siemens Ltd.	Bharti Airtel Ltd.	A B B India Ltd.

Tata Consumer Products Ltd.	Bosch Ltd.	Grasim Industries Ltd.
Godrej Consumer Products Ltd.		
Cluster 3		
Cholamandalam Investment & Finance Co. Ltd.	S B I Cards & Payment Services Ltd.	Power Grid Corpn. Of India Ltd.
Hindustan Unilever Ltd.	Bajaj Holdings & Invst. Ltd.	Reliance Industries Ltd.
I T C Ltd.	Berger Paints India Ltd.	H C L Technologies Ltd.
Infosys Ltd.	Britannia Industries Ltd.	S R F Ltd.
J S W Steel Ltd.	Asian Paints Ltd.	Shree Cement Ltd.
Kotak Mahindra Bank Ltd.	Dabur India Ltd.	Tata Consultancy Services Ltd.
Ltimindtree Ltd.	Divi'S Laboratories Ltd.	Tech Mahindra Ltd.
Marico Ltd.	Dr. Reddy'S Laboratories Ltd.	Ultratech Cement Ltd.
Maruti Suzuki India Ltd.	Eicher Motors Ltd.	Wipro Ltd.
Cluster 4		
Adani Total Gas Ltd.	Adani Energy Solutions Ltd.	Bajaj Finserv Ltd.
Cluster 5		
D L F Ltd.	Trent Ltd.	Zomato Ltd.
Adani Ports & Special Economic Zone Ltd.		