Stock Price Forecasting Models during Crisis and High Volatility

Thesis

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

NAGARAJ NAIK

Reg. No.: 165061 IT16FV04

DEPARTMENT OF INFORMATION TECHNOLOGY

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALURU - 575025

December 2021

DECLARATION

By the Ph.D. Research Scholar

I hereby *declare* that the Research Thesis entitled **Stock Price Forecasting Models during Crisis and High Volatility** which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy** in **Information Technology** is a *bonafide report of the research work carried out by me*. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

NAGARAJ NAIK

Reg. No.: 165061 IT16FV04

Department of Information Technology

Place: NITK, Surathkal.

Date:

CERTIFICATE

This is to *certify* that the Research Thesis entitled **Stock Price Forecasting Models during Crisis and High Volatility** submitted by **Nagaraj Naik**, (Reg. No.: 165061 IT16FV04) as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfillment of the requirements for the award of degree of **Doctor of Philosophy**.

(Dr. Biju R Mohan)
Research Guide

Chairman - DRPC

DEDICATION AND ACKNOWLEDGMENT

It is a great opportunity to thank **Dr. Biju R Mohan** (Research Guide , Department of Information Technology, National Institute of Technology Karnataka) for being constant motivation and providing valuable suggestions throughout my research journey. I thank **Dr. Nagamma Patil** and **Dr Jaidhar C D**(Head, Department of Information Technology, National Institute of Technology Karnataka) for the support for my research work. I thank the Research Progress Assessment Committee (**RPAC**) members for their continuous support and encouragement.

I convey many thanks to all fellow doctoral students, teaching faculties and non-teaching staffs in the Department of Information Technology for encouraging to pursue hard work and their cooperation. Especially, I thank **Dr. Rathinaraj** (Ph.D, NITK) for the timely support during many odd times and am grateful to all who helped me directly/indirectly.

I extend my sincere thanks to the Department of **MeitY**, Government of India, for providing the financial support under Visvesvaraya Ph.D. Scheme for Electronics and IT to carry out the research work. Finally, I thanks to **Director** and **Deputy Director** (NITK Surathkal, Karnataka), for facilitating updated infrastructure to carry out my research.

It is always impossible without the family support to invest huge time for research. I am debted to my parents, **Mrs. Sumitra Madhukar Naik** and **Mr. Madhukar B Naik**, for the whole life. I also thank my brother and sisters. Life would be incomplete without my daughter **Diksha Nagaraj Naik** for spending their lovely time and fun talk. Finally, I should mention my source of inspiration during my research work, my wife, **Saroja Nagaraj Naik**, who consistently gave mental support all through the tough journey. Infinite thanks to her for keeping my life green and lovable.

ABSTRACT

Stock market crisis can emerge due to variations in the economic policy of a large economy. It has been observed that the crisis may originate from a large size economy, and the impact of the crisis will affect smaller economies as well. Crisis prediction is critical for the financial market, and this attracted many researchers and academicians. However the fair value of the stock price depends on stock financial parameters. There are many financial parameters such as price to earnings, company returns, company debt, etc. Identification of relevant financial parameters is a challenging task. Therefore in this work a Hybrid Feature Selection (HFS) technique is proposed to select essential financial parameters. After selecting essential financial parameters, Naive Bayes (NB) classifier is used to classify high quality stocks. Then stock price bubbles are identified using Relative Strength Index (RSI). From these bubble points, with help of moving average we selected stock crisis points. These crisis points are fed to regression techniques. The performance of the model is evaluated based on Mean Squared Error(MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). It is found from the experimental results that performance of the XGBoost model is better than the Deep Neural Network (DNN) model.

Stock price movements forecasting is an important topic for traders and stock analyst. Timely prediction of stock movements yields more profits. However, there are more than a hundred technical indicators are available, so it is essential to identify the correlation between stock price and technical indicators. Because each technical indicator signifies different aspects of stock price, so it is crucial to have an appropriate feature selection algorithm to select the correct technical indicators. Boruta feature selection is used to select the best features out of the 33 technical indicators and stock price movements are classified using the DNN, Artificial Neural Network (ANN), and Support Vector Machine (SVM). It is found from the experimental results that performance of the DNN model is better than the ANN and SVM model. The proposed DNN method improved the classification accuracy rate by 5% to 6%.

Volatility is a measure that represents the rate of change in the stock price over time, and it is calculated using standard deviations. This measure helps the investors to estimates the risk in the stock investments. The stock price is volatile due to many factors such as demand, supply, economic policy, and company earnings. Investing in a volatile market is riskier for stock traders. Most of the existing work considered Generalized Auto-regressive Conditional Hetroscedacity (GARCH) models to capture volatility, but this model fails to capture when the volatility is very high. We have considered stock price volatility estimation using the regime-switching models. The performance of the Markov-Switching GARCH (MSGARCH) and Self-Exciting Threshold AutoRegressive (SETAR) models is calculated using RMSE and Mean Absolute Percentage Error (MAPE) metrics. It is found that regime-switching models performed better than the GARCH models.

Keywords: *DNN, Hetroscedacity, Hybrid Feature Selection, Stock Crisis*

Contents

List of Figures

List of Tables

ABBREVIATIONS

Abbreviation	Meaning
US	United States
Stochastic K	Stochastic Oscillator K indicator
Stochastic K	Stochastic Oscillator D indicator
MACD	Moving Average Convergence Divergence
William R	Momentum indicator
A/D Oscillators	Accumulation/Distribution
RSI	Relative Strength Index
HFS	Hybrid Feature Selection
DNN	Deep Neural Network
XGBoost	Extreme Gradient Boosting
GARCH	Generalized AutoRegressive Conditional Heteroskedasticity
RFE	Recursive Feature Elimination
BFS	Boruta Feature Selection
RMSE	Root Mean Square Error
P/E	Price-Earnings
NB	Naive Bayes
MSE	Mean Squared Error
MAE	Mean Absolute Error
MSGARCH	Markov-Switching GARCH
EGARCH	Exponential GARCH
TGARCH	Threshold GARCH
AIC	Akaike's Information Criteria
BIC	Bayesian Information Criteria

SETAR	Self-Exciting Threshold AutoRegressive
MAPE	Mean Absolute Percentage Error

Chapter 1

INTRODUCTION

Due to the volatile nature of the stock market, predicting stock price movements is challenging. Early interpretation of stock-related critical information makes trading more profitable (Alexander, 1961). The Efficient Market Hypothesis (EMH) states that the prices of traded assets, such as stocks, already reflect all publicly available information. According to EMH (Malkiel and Fama, 1970) theory, the stocks are trading at fair values in the stock exchanges, and it is not possible to buy a stock that is undervalued. Investors are making money either by taking a chance or making riskier investments. Thus, investors will not systematically outperform the market over time. There has been an increasing number of studies (Malkiel, 2003), (Smith, 2003), (Prechter Jr and Parker, 2007), (Bollen et al., 2011) that provide evidence contrary to what is suggested by the EMH. These studies show that the stock market can be predicted to some extent. Many people within the investor community also believe that Warren Buffet's ability (Kiersz, 2015) to consistently beat the Standard & Poor's index (Loomis, 2012) hence conclude that the market can be predicted.

In stock market, there are two types of investors; 1. Institutional investor, 2. Retail investors. Institutional investors are well-trained and professional investors who manage large funds. They prefer to invest money in a long term basis. On the other hand, retail investors are individual investors who are not trained and lack professionalism. Hence individual investors are finding difficulty in stock market investments.

Short-term trading involves buying and selling stocks within 4 to 12 months time horizon. Here stocks are purchased for taking the profit in a short period. To trade in the short term requires quick decision-making for buying and selling stock. Therefore identifying when to buy and sell a stock is one of the critical challenges in the financial market.

Long-term trading involves buying a stock for more than one year to get higher returns. Here stocks are purchased for taking the long-term benefits. Such as dividends and bonuses. However, individual investors are finding it challenging to identify the long-term stock for their investments.

Most of the stock-related data can be generally classified into quantitative data and qualitative data (Suzuki and Ohkura, 2016). In the quantitative approach, investors decide stock investment based on company quantitative data such as company cash flow, balance sheet, and historical stock price. The following task is carried out by investors to make the investment decision (Barak and Modarres, 2015).

- Analyzing the company cash flow, balance sheet, and monthly earnings.
- Analyzing current stock prices with historical stock prices.
- Analyzing cyclic periods of stock prices

The qualitative approach is a subjective analysis based on unquantifiable data. In the qualitative approach, more weight is given to company products and their services. This approach highly concentrates on the quality of the company. The fund manager carries out the following task of the qualitative approach (Gummesson, 2000).

- Discussion with corporate management and observing the plan of the company.
- Analyzing management quality based on past work.
- Analyzing company products and their competitors.

Stock analysis can be performed in two ways. The first is fundamental analysis, and the second is technical analysis. A fundamental analyst studies the financial statements of the company. It includes earnings of the company, company assets, and its liability. The fundamental analysis plays an essential role in identifying the quality of stock. There are 42 financial parameters (Barak and Modarres, 2015) as list in the Fig 1.1. Hence it will be difficult for the investors to know which financial parameters have more predictive power. Technical analysis is a method of future stock price prediction based on historical stock data. It includes stock open price, close price, day high price, day low price, and volume of the stock. Technical analysis is used in short term and long term stock investments. Therefore fundamental and technical analyses are essential studies to analyze the stock prices.

Figure 1.1 Stock financial parameters list.

Financial market stock price movements purely depend on various sources of information. It is not easy to interpret information from different sources. Aggregating and processing information from various platforms is a key challenge in financial markets. For example, there were many technical indicators; it will be difficult for the day traders to know which technical indicators have more significance. Sometimes stock related news may affect the stock prices. Based on the news, the stock price may trade positively or negatively. Analyzing such news and interpreting the correlation of stock price is generally a difficult task for day traders.

The stock prices are affected by many events such as company balance sheet variation, political uncertainty, bond market rate, and global market trends. Sometimes, stocks' prices react when there is a sudden change in management or share dividend and bonus announcement. All these factors reflect in stock prices, hence it is difficult to predict the stock prices accurately. This is one of the key challenges we try to investigate.

Estimation of a stock price is carried out by using technical analysis. However, more than a hundred technical indicators exist, so it is difficult to identify the correlation between stock price and technical indicators. Because each technical indicator signifies different aspects of stock price, so it is essential to have an appropriate feature selection algorithm to select the correct technical indicators.

The following factors described in Figure 1.2 influences stock price movements.

- Inflation Rate: The inflation rate is one of the essential factors that decide the market direction in the short term. Rising inflation, to some extent, reduces the purchasing power of people's daily needs. For example, if the inflation rate is high, purchasing power will come down. Monitoring the inflation rate and identifying inflation-related stocks is a difficult task for retail investors.

- Interest Rate: The people borrow money during the low-interest rate for purchasing a house and a car. The low - interest rate also encourages the corporate sector to expand the business and get more benefits during this period. Reserve Bank of India (RBI) determines the repo rate through the auction system, and the rate reflects the liquidity conditions prevailing in the market (RBI, 2021). Hence analyzing interest rates and their relationship with technical indicators is a difficult task for retail investors.

- Company Earnings: Company earnings are one of the essential parameters which assess the company's strength. There are many financial parameters that exist to assess the strength of the company. Identifying the appropriate parameter is one of the challenging tasks for a retail investor.

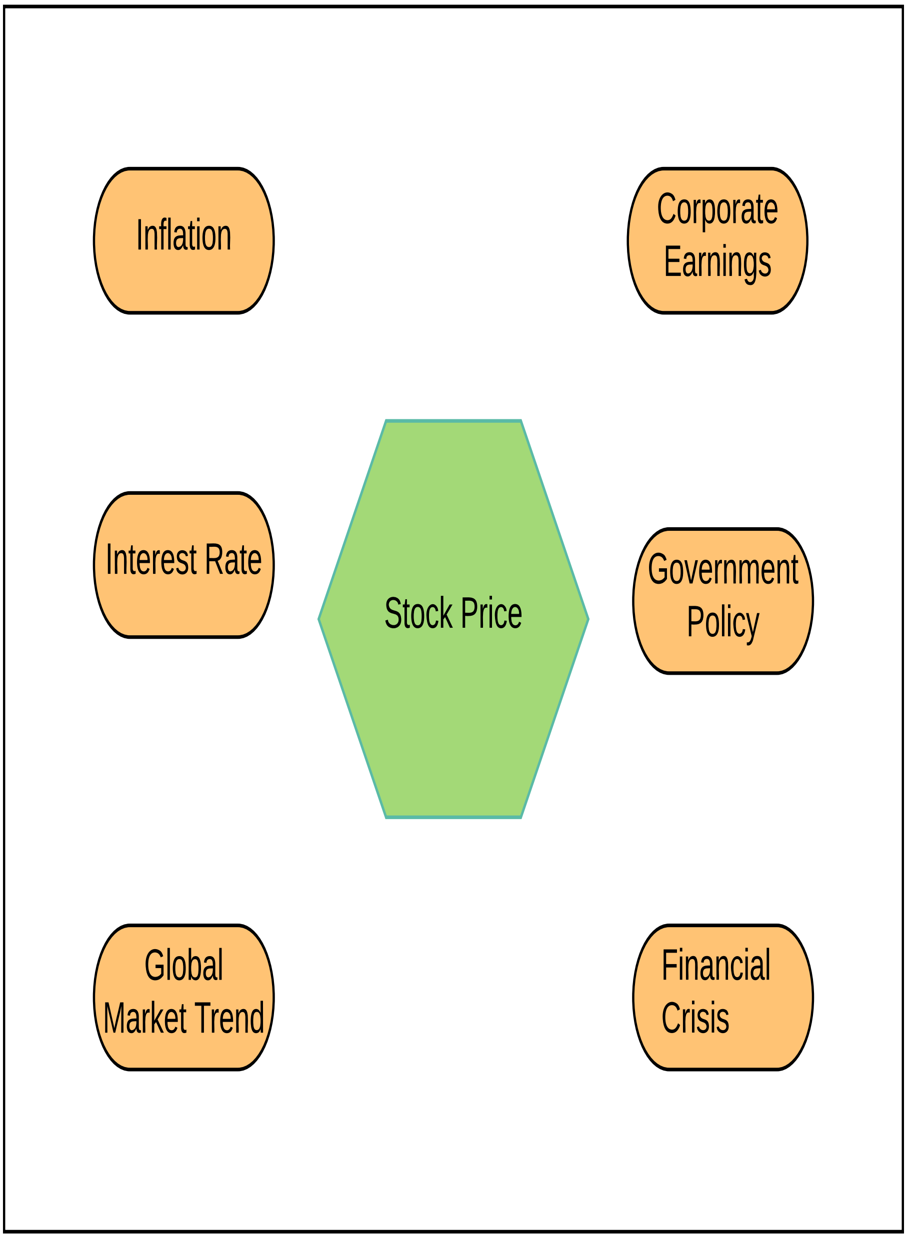


Figure 1.2 Factors influences the stock price movements.

- Global Market Trend: In 2008, the initial financial crisis was started in America's united states due to bankruptcy. This crisis affected other countries like Japan, China, USA, Germany, UK and India, and the recession started across the world. This suggests that a retail investor should carefully follow the global market trend before investing in the stock. This is very difficult for a retail investor to identify the trend of the global market and predict the Indian market trend.

- Government Policy: Government policy plays a vital role in the stock market. It gives direction to the market. Suppose if the government policy is well defined, then it will attract the investors for investing money. Therefore it is required to identify the correlation between investor mood and government policy to gain more profit.

- Financial Crisis: The stock market crash is a sudden dramatic decline of stock price due to uncertainty in the stock market. The market crash happens when there are more sellers in the market than buyers. Identifying such data is one of the challenging tasks.

1.1 Technical Analysis

Technical analysis uses technical indicators to estimates the future stock price. Traders or investors use technical analysis methods to identify potential trading opportunities based on the stock price's history (Li and Bastos, 2020).

1.1.1 Brief Description of Technical Indicators

Most of the investors use technical indicators to identify the stock price entry and exit level. In majority work, technical indicators selection is based on an expert's opinion or random selection (Kara et al., 2011), (Patel et al., 2015b). Below a few technical indicators are listed.

1. Simple Moving Average (SMA): SMA (Anbalagan and Maheswari, 2015) is a type of technical analysis used to find the average mean value of the stock price. The SMA analysis is used to know how the stock price changes over a specified number of days. The moving average is calculated based on the closing price of the stock. In Figure 1.3, we have considered moving average technical indicator to decide on stock either buy or sell. By looking at Figure 1.3, suppose if stock price trading is above the moving average technical indicator, then it indicates that stock is an uptrend. If the stock price is trading below the moving average of technical indicator, it indicates that stock is in a downtrend.

2. Exponential Moving Average (EMA): EMA (Sezer and Ozbayoglu, 2018) technical analysis, more weight is given to the recent data and less weight to the older data. In SMA, weights are given equally to all the data. The EMA gives more importance to the most recent data points, this helps the trader to make quick trading decisions based on recent data.

3. Momentum (MOM): It refers to the rate of acceleration of stock prices. It is the difference between the current closing price and the closing price of N (number of days) days ago. Based on the MOM trend, the investor will take a long or short position on a stock by looking at the acceleration in stock prices (Leivo and Pätäri, 2011)

4. Stochastic K% (STCK): George Lane (Gulisashvili, 2012) developed a stochastic indicator. It is used to analyze the stock based on the closing price over a specified number of days. Most of the analyst uses the default three days moving average. The range of stochastic indicators is between 0 to 100. The range values between 1 to 20 indicate oversold levels, and 80 to 100 technical indicator ranges indicate the overbought levels.

5. Relative Strength Index (RSI): RSI (Rodríguez-González et al., 2011) is one kind of momentum indicator. The RSI technical indicator value ranges from 0 to 100. The RSI values below 30 indicate that the stock price is oversold, and RSI values above 70 indicate the overbought levels. Figure 1.4 shows the stock price value along with technical indicators. In this technical chart, we noticed that RSI technical values lie between 1 to 100. Now the question is how to interpret the RSI indicator values? If we look at Figure 1.4, suppose if the RSI indicator value is below 30, it indicates that the stock price is oversold. Suppose, if the RSI value is above 70, then it indicates that the stock is overbought. We can notice in the chart that when RSI reaches above 70, then the stock price starts falling.

Figure 1.3 Stock Price along with Moving Average Technical Indicator.

Figure 1.4 Stock Price along with RSI Technical Indicator.

1.1.2 Stock Fundamental Analysis

A fundamental analyst studies the financial statements of the company. The fair value of the stock price depends on stock financial parameters. There are many financial parameters such as price to earnings, company returns, company debt, etc. Identification of relevant stock parameters is a challenging task. The stock financial parameters are described in Fig 1.1.

1.2 Stock Crisis

The stock price crisis is nothing but a significant drop in stock price more than 10% within a few days due to heavy selling in stocks (Roll, 1988). The reasons for the selloff in the stock market are listed below.

1. The stock is overpriced.

2. Company posts the bad earnings.

3. Global market slowdown due to trade war.

4. Geopolitical Tension.

5. Pandemic situations like COVID19.

The stock market crisis can emerge due to variations in global and local market economic policy and macroeconomic data. For example, a financial market slowdown in 2008 (Fleitas et al., 2018) is initially originated in the United States (US), and subsequently, it has affected the economies of other countries. It has been observed that the crisis may originate from a large size economy, and the impact of the crisis will affect smaller economies as well. For example, the subprime crisis originated in the US and evolved into a sovereign debt crisis in European countries. This crisis affected the Asian market, as well. Crisis prediction is critical for the financial market, and this attracted many researchers and academicians. Predicting a crisis is one of the critical issues in the proposed work.

However, (Chatzis et al., 2018) proposed stock crisis events based on deep learning classification methods. The study considered the less than one percentile of stock returns as a stock crisis point. This study ignored stock financial parameters to identify quality stock. Also this study did not consider bubble points in stock prices. When the stock price is trading above its fundamental value like earnings and its asset price, it is called the stock price bubble. Hence identification of stock price bubble is one of the objectives in this work.

1.3 Stock Price Movements Classification and Prediction

Stock price movements classification is an essential topic for traders and stock analysts. Timely predictions in stock price movements can yield more profits. Predicting stock price movement is difficult due to volatility in the financial market. The variation in economic policy, macroeconomic data, political uncertainty, and government policy are affected in the stock market's direction. The company earnings and other related financial issues also affect the stock prices. Due to this, stock prices may fluctuate, and it increases the volatility in the stock market.

The stock price classification is carried out using ANN and SVM machine learning methods (Kara et al., 2011), (Patel et al., 2015a). The DNN gains popularity in classification due to the transformation and extraction of input features and establishing a relationship between the data (Ciregan et al., 2012). Moreover, the DNN model is performed better than ANN and SVM techniques (Fan et al., 2021). Deep neural network are used extensively in medical image classification, health-care applications and extensive data analysis (Ciregan et al., 2012), (Stephen et al., 2019), (Najafabadi et al., 2015).

Most of the studies have used technical indicators for stock price movements classification (Kara et al., 2011), (Patel et al., 2015a). However, these studies considered the most popular ten technical indicators, namely moving-average, weighted moving average, momentum, stochastic K, stochastic D, MACD, larry William R,A/D oscillators, commodity channel index and RSI, etc. A very few studies considered more than ten technical indicators and studied their effects on classification accuracy, these studies show there is a possibility of enhancing classification accuracy. Therefore in this work, we have considered 33 technical indicators. In this work, hybrid feature selection method is used to identify the effectiveness of technical indicators in predictions.

1.4 Stock Price Volatility Estimation

Volatility is a measure that represents the rate of change in the stock price over time, and it is calculated using standard deviations. This measure helps the investors to estimates the risk in the stock investments. When the volatility is very high, then it is riskier to invest. So identification of volatility in the market is essential as far stock market is concerned.

Stock market volatility is estimated by using autoregressive conditional heteroscedasticity (Engle, 1982), (Robert, 1982), (Hansen and Lunde, 2005). The GARCH model has been proposed to capture the different families of conditional volatility (Bollerslev, 1986). GARCH model is useful when the variance of the stock price is not constant (Lamoureux and Lastrapes, 1990). Most of the work has used the GARCH model to estimate the volatility in the stock market (Franses and Van Dijk, 1996), (Endri et al., 2020), (Emenogu et al., 2020). The GARCH model does not capture the different variations of volatility periods because the GARCH parameters alpha and beta are restricted to less than one (Bauwens et al., 2006), (Marcucci et al., 2005), (Zhang et al., 2019), (Klaassen, 2002), (Naeem et al., 2019). However, when there is higher volatility, these parameter values can go beyond one. Hence it fails to capture the higher volatility. According to a Markov process, a solution to this problem is to allow the GARCH model's parameters to vary over time (Calvet and Fisher, 2004), (Ardia et al., 2016), (Ardia et al., 2019), (BenSaïda, 2015), (Aloui and Jammazi, 2009), (Maciel, 2020). Therefore in this work, we have considered regime-switching based on the Markov switching GARCH (MSGARCH) and Self-Exciting Threshold Autoregressive (SETAR) model instead of the plain GARCH model. To the best of our knowledge, this is the first empirical study on the Indian stock market data based on the regime-switching model.

1.5 Challenges

- The fair value of the stock price depends on stock financial parameters. There are many financial parameters such as price to earnings, company returns, company debt, etc. Identification of relevant stock parameters is a challenging task.

- Stock market crisis can emerge due to variations in global and local market economic policy and macroeconomic data. Identification of the stock crisis is one of the important challenges.

- There are more than 100 technical indicators are available. The selection of appropriate technical indicators to make a buy or sell decision based on the stock movement pattern is a challenging task.

- Nowadays, there is a possibility of a different kind of fraud or scam in the market, such as banking scams, housing scams, company accounts mismanagement, etc. Early prediction of such scams is one of the critical challenges in the stock market.

- Analyzing inflation rate, interest rate, global market trend, and corporate earnings and identifying their effects on stock price is a difficult task for retail investors.

- There are many key factors such as market liquidity, merger announcements, dividend announcements, reserve bank meeting outcome, and management change, which influence the decision of momentum of the stock price in the stock market. Selecting relevant key factors and mapping market behavior to predict stock prices' momentum is the biggest challenge.

- Identifying the order of Generalized autoregressive conditional heteroscedasticity (GARCH) is a challenging task.

1.6 Research Motivation

Nowadays, more and more valuable financial data is available for continuous analysis and investigation. The availability of high-dimensional financial data enhances the opportunities of using machine-learning algorithms to minimize the error. Identifying hidden patterns in the data may be used to predict the fluctuation of stock prices. Hence it creates interest to researchers in investigating different behavior of real-world financial data.

The stock price is highly volatile due to several factors that influence stock prices, including demand, supply, volume growth, economic policy, and corporate results. Therefore measuring the volatility of stock prices is extremely difficult.

Financial literacy among retail investors is very poor. An average retail investor does not understand the basic concept of technical analysis. During the stock crisis, retail investors are finding it difficult to make a buy or sell decision on stock investments. Therefore it creates the opportunity to develop a stock price crisis prediction and stock price prediction model to make the decision in stock investments.

1.7 Research Framework

The research framework is described in Fig 1.5. The following work is considered in this thesis.

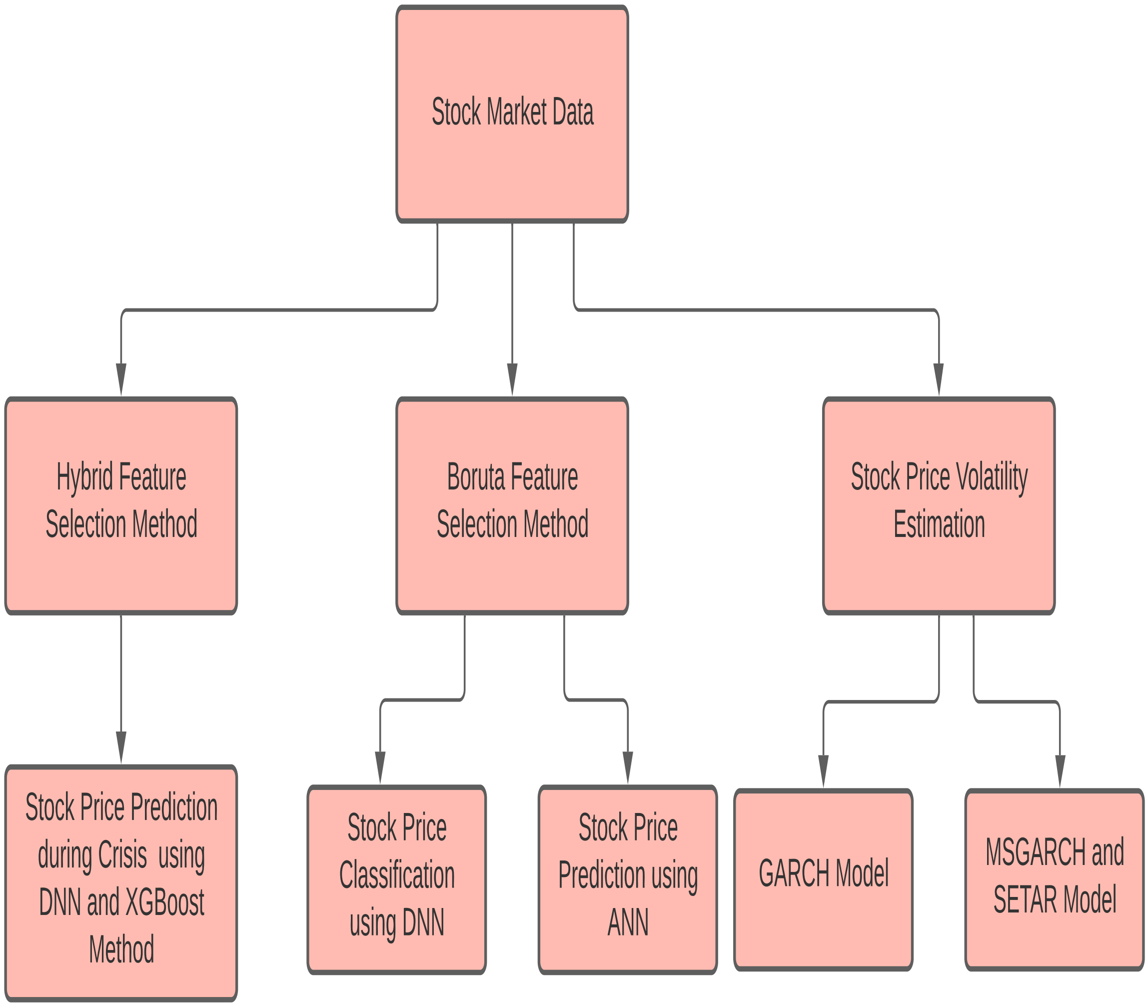


Figure 1.5 Research Framework.

1) To develop a stock price prediction during the crisis using the hybrid feature selection technique.

2) To develop an efficient approach to increase the performance of the stock price classification based on the feature selection method.

3) To develop an efficient approach to estimates volatility of the stock price using the regime-switching techniques.

Several studies have focused on technical indicators to classify stock price movements (Kara et al., 2011), (Patel et al., 2015b). These studies considered the ten most popular technical indicators, including moving average, weighted moving average, momentum, stochastic K, stochastic D, MACD, Larry William R, A/D oscillators, commodities channel index, and the relative strength index. Even though only a small number of studies examined more than ten technical indicators and examined their influence on classification accuracy, Therefore, in this work, we have considered 33 technical indicators to improve the stock price classification accuracy.

Because of the increased volatility in the stock market during a stock crisis, it is difficult to predict stock prices. To the best of our knowledge, this is the first time that a method for predicting stock prices has been used during a stock market crisis. We have developed a Hybrid Feature Selection (HFS) algorithm to remove irrelevant stock financial parameters from the stock data. The NB classifier method is considered to be the most effective method of identifying fundamentally strong stocks. Later on, the RSI approach is used to identify stock price bubbles. Stock crisis points are identified using moving average statics, which is taken into consideration. XGBoost and DNN regression methods are used to forecast the price of stocks during a stock market crisis.

During the day's trading, the stock price is highly volatile because of several factors that influence stock prices, including demand, supply, volume growth, economic policy, and business earnings. As a result, measuring the volatility of stock prices is difficult. The majority of the current research focused on GARCH models to capture volatility. However, these models were unable to capture the many variations in volatility. We have investigated regime-switching models based on the MSGARCH and SETAR models to capture the volatility of stock prices. In this study, we considered MSGARCH models used to estimate the volatility of the stock price. The experimental results are tested on Indian stock market data.

1.8 Research Contributions

Research contributions are mentioned below:

- Developed a Hybrid Feature Selection (HFS) algorithm to remove irrelevant financial parameters. In this work, Extreme Gradient Boosting (XGBoost) and

Deep Neural Network (DNN) are used as regressors to predict the stock prices during the stock crisis.

- Developed an efficient approach to increase the performance of the stock price classification based on the feature selection method.
- Developed a method to identify a trend in data using a combination of candlestick data and technical indicators to classify the stock price movements.
- Developed a method to estimate stock price volatility based on the regime-switching models.

1.9 Outline of the Thesis

The rest of the chapter is organized as follows: Chapter 2 presents a survey related work on stock crisis events, stock price movements classification, and stock price volatility estimation. Chapter 3 describes stock crisis prediction using the hybrid feature selection technique. Chapter 4 addresses the issue related to stock price classification and stock price prediction. Chapter 5 describes the volatility estimation using regime switching method. Conclusion and future work are mentioned in Chapter 6.

1.10 Summary

In this chapter, explained the basic concept of technical indicators and financial parameters are presented. A detailed explanation of stock crisis prediction, stock price movement classification, and stock price volatility estimation are presented. Further, discussed the motivation for doing this research work and finally, the detailed organization of the thesis is given in this chapter. The detailed literature survey, identified research gaps, the problem statement, and the objectives are discussed in the next chapter.

Chapter 2

Literature Review and Proposed Works

This chapter reviews the related work regarding stock price classification, stock price crisis prediction, and stock price volatility estimation. In addition to that, it discusses the outcome of the literature review, problem statements, and objectives.

2.1 Stock Price Classification and Prediction Technique

Generally, stock price movements classification is carried out using two ways, (a) Technical analysis and (b) fundamental analysis. In technical analysis, technical indicators are used to estimates the stock price movements. Technical indicators are extracted from the historical stock price. In fundamental analysis, company financial variables are considered to classify the stock price movements. Such as price to earnings, earning per share, dividend, return on equity, etc. However, most of the work considered either technical indicators or financial variables to classify or forecast the stock price trends.

The regression techniques are used to forecast the stock price, and the classification methods are used to classify up and down trends. It is found from the literature that linear regression, exponential smoothing, Autoregressive Integrated Moving Average, and Generalized Autoregressive Conditional Heteroscedasticity are the most commonly applied methods for stock price predictions (Vu et al., 2012), (Enke and Mehdiyev, 2013), (Niaki and Hoseinzade, 2013). It is also found that Artificial Neural Network(ANN) has been widely used for classifying stock price movements classification (Kara et al., 2011), (Patel et al., 2015a).

Kara et al. (2011) considered ANN and SVM methods for stock price movements classification. In this work, the authors used ten technical indicators, and these technical indicators were given to ANN and SVM to predict the stock price movements. Three layers of the neural network were considered to classify the stock price movements. Initially, the prediction model was assigned the random weight. The back-propagation method was considered to train the model. The experimental results concluded that the ANN model performed better than SVM. Patel et al. (2015a) investigated stock movements classification using ANN, SVM, RF, and NB classifier. The experimental results concluded that the NB classifier performed better than other models.

Patel et al. (2015b) investigated stock price prediction using the combination of SVR-ANN, ANN-RF, and SVR-RF methods. In this work, they considered ten technical indicators and predicted stock price for ten, fifteen, and 20 days periods. The performance of the model is evaluated using MAPE and RMSE metrics, the SVR-ANN model outperformed other combinations.

Tsai et al. (2011) used technical indicators such as moving average, moving average convergence and divergence, relative strength index, and commodity channel index to predict the stock's future price. In this work, they used historical data such as past prices and volume of stock traded to make investment decisions. In addition, they have used feature reduction techniques like PCA, GA, and decision tree.

Gunduz et al. (2017) investigated stock price movements classification using the CNN method. This work considered twenty-five technical indicators, and they used Chi-Square feature selection to remove the irrelevant features in the datasets. The performance of the models was evaluated using the F Measure metric. The CNN classifiers outperformed the logistic regression method. They used Borsa Instanbul stock exchange data for the experiments.

Barak and Modarres (2015) proposed a hybrid feature selection method which is a combination of filter and function-based clustering techniques to select the features. Financial variables like P/E, EPS, ROE, etc., were used to classify the stock price movements in this work. Classification methods like decision tree, ANN, and SVM were considered. They conducted these experiments using Tehran stock data.

Lee et al. (2019) proposed the financial volatility network model to predict the global stock indices movements. The proposed model is the combination of the Pearson correlation and VAR model. The relationship between the country indices was identified using the Pearson correlation method, and variance is computed using the VAR model. Later the volatility is estimated using the SVM, LR, and RF methods. Accuracy metric was used to evaluate the model performance. The study concludes that the SVM method performed better than LR and RF methods.

Ballings et al. (2015) proposed the ensemble method to predict the direction of stock prices. The Ensemble method is the combination of RF and AdaBoost and kernel factory techniques. The model performance was evaluated using the AUC (Area under Curve) metric. The ensemble method performed better than NN, RF, and SVM techniques. European stock exchange data were considered in the experiments.

Barak et al. (2017) proposed fusion of multiple classifiers model to predict the stock price movements. It is the combination of more than one classifier. In this work, the combination of the Bagging, Boosting, and AdaBoost model was considered to classify the stock price movements. The model performance was evaluated using the Accuracy metric. Tehran stock exchange data were considered in the experiments. The study concluded that the fusion model performance was better than the single classifier model.

Laboissiere et al. (2015) investigated the stock price prediction using the ANN method. In this work, minimum and maximum stock price data were given as input to the ANN model. The three-layer ANN was considered in the proposed work. Each input data was multiplied by a weight, and a summation of weighted input data was given to neurons. The sigmoid activation function was used in the ANN model. MAE, MAPE, and RMSE metrics were considered to evaluate the performance model. Brazilian stock exchange data were considered in the experiments.

Ince and Trafalis (2008) studied stock price prediction using technical indicators. In this work, they considered eight technical indicators to identify the trend in stock prices. SVR, MLP, and ARIMA techniques were used to predict the stock prices. The RMSE metric is used to evaluate the performance of the model. The SVR and MLP model performance was better than the ARIMA model. NASDAQ stock exchange data were considered in the experiments.

Hsu et al. (2011) proposed hybrid feature selection method.It is a combination of filter and wrapper feature selection techniques. Accuracy metric was considered to evaluate the performance of the SVM model. The study concludes that the feature selection method improves the accuracy of the model.

Preis et al. (2013) hypothesized that investors might use a Google hits ratio of pages to find the investor's sentiments to decide on stock prices. Vaisla and Bhatt (2010) showed ANN gives more reliable prediction results for high-dimensional nonlinear data.

Mitchell and Mulherin (1994) conclude news related to market sentiment has a direct impact on trading activity in the stock market. More trading volumes could be found during the positive news announcement and negative news announcements. Chan (2003) conclude that there will be a specific impact on the stock prices when the sudden change in company management. Keown and Pinkerton (1981) study concluded that huge incremental trading volume was discovered before one week of the merger announcement, and the most robust trading volume response was on the day of the merger declaration. Hu et al. (2015); Pouya et al. (2016) considered the P/E ratio and expert recommendations to select the quality stock.

Chourmouziadis and Chatzoglou (2016) found that existing trading rules were not gainful for future periods when market conditions change dynamically. The authors proposed a short-term technical trading strategy by considering the stock's daily price using fuzzy systems. Berutich et al. (2016) proposed an automatic way of buying and selling financial securities without portfolio managers' help using the fuzzy method.

Hassan et al. (2007) proposed the fusion model, a combination of the Hidden Markov Model, Artificial Neural Network (ANN), and genetic method was considered to optimize the parameters of ANN. The outcome of the proposed model performance is better than a single hidden Markov model. The mean absolute error was used to evaluate the performance of the model.

A detailed description of the common technical indicators used in many of these works (Patel et al., 2015a), (Kara et al., 2011) is given below.

Table 2.1 Related Work on Stock Price Classification and Prediction

Author	No.of Technical Indicator	Method	Target output	Performance Metric	Gap
Enke and Mehdiyev (2013)	20(S&P 500 index)	NN	Stock price	RMSE	No Feature Selection
Niaki and Hoseinzade (2013)	21(S&P 500)	ANN	Up or Down	Simulation	No Feature Selection
Cervelló-Royo et al. (2015)	10(US Dow Jones)	Template matching	Bull/Bear-flag	Simulation	No Feature Selection
Kara et al. (2011)	10(Istanbull Stock Exchange)	ANN,SVM	Up or Down	Accuracy	No Feature Selection
Patel et al. (2015a)	10(India CNX and BSE index)	ANN, RF, SVM	Up or Down	Accuracy	No Feature Selection
Patel et al. (2015b)	10(India CNX and BSE index)	SVR + ANN, RF, SVR	Stock price	RMSE	No Feature Selection
Chen and Chen (2016)	22(Taiwan TAIEX and US NASDAQ)	Template	Bull-Flag	Simulation	No Feature Selection
Qiu et al. (2016)	21(Japan Nikkei 225 index)	ANN	Stock return	MSE	No Feature Selection
Zhong and Enke (2017)	23 (S&P 500)	ANN	Up or Down	Simulation	PCA Feature Selection
Nayak et al. (2015)	10(India CNX and BSE index)	SVM-KNN	Profit and loss	RMSE	No Feature Selection
Enke and Thawornwong (2005)	21 (S&P 500 index)	Neural Network	Stock Price	RMSE	No Feature Selection

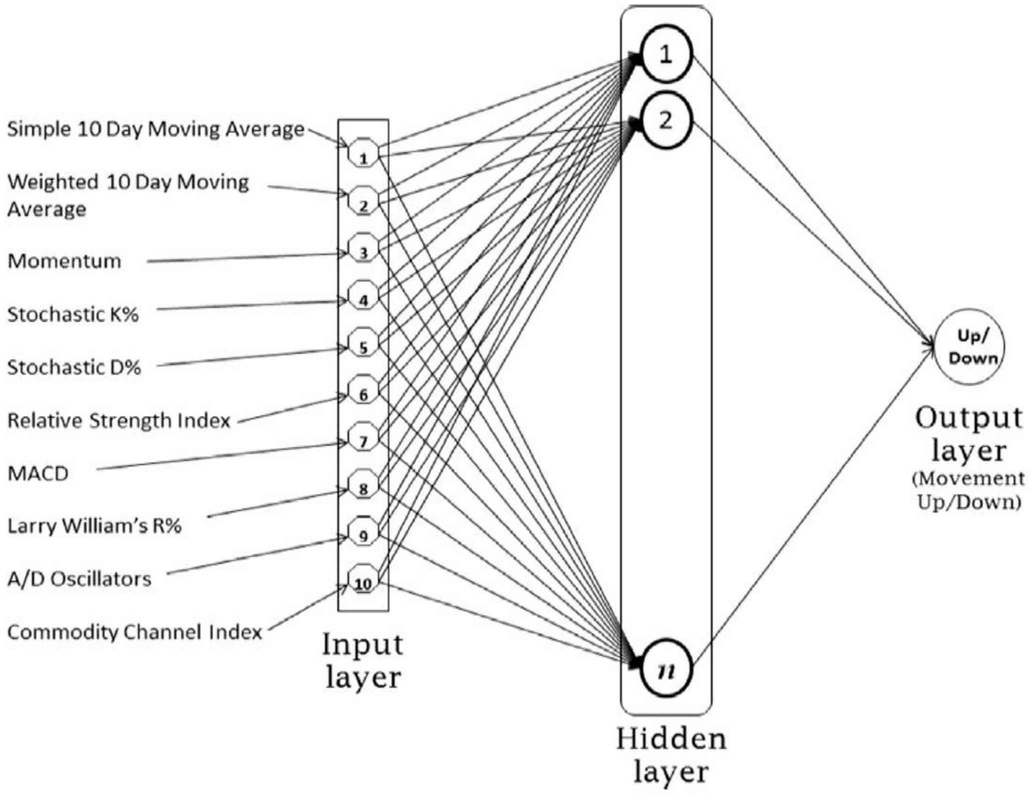


Figure 2.1 ANN Classification model (Kara et al., 2011).

1. **SMA**: When the stock price candle is trading above SMA, it indicates its buy signal and is represented as +1. When the stock price candle is trading below SMA, it indicates a sell signal, and it is represented as -1 (Patel et al., 2015a).

2. **Stochastic crossover:** Determines the overbought and oversold of the stock price. The stochastic crossover is composed of two oscillators, the stochastic D% and stochastic K%. The stochastic D% is slow and stochastic K% is a fast stochastic oscillator. When fast stochastic K% crosses up through the slow stochastic D%, then it indicates a buy signal represented as +1. When slow stochastic D% crosses down through the fast stochastic K%, it indicates a sell signal represented as -1 (Patel et al., 2015a).

3. **MACD Crossover:** In this twelve days, moving average is considered the fastest, and twenty-six days are moving average are considered slower. If Twelve days are moving average crossing up through twenty-six days moving average, it indicates

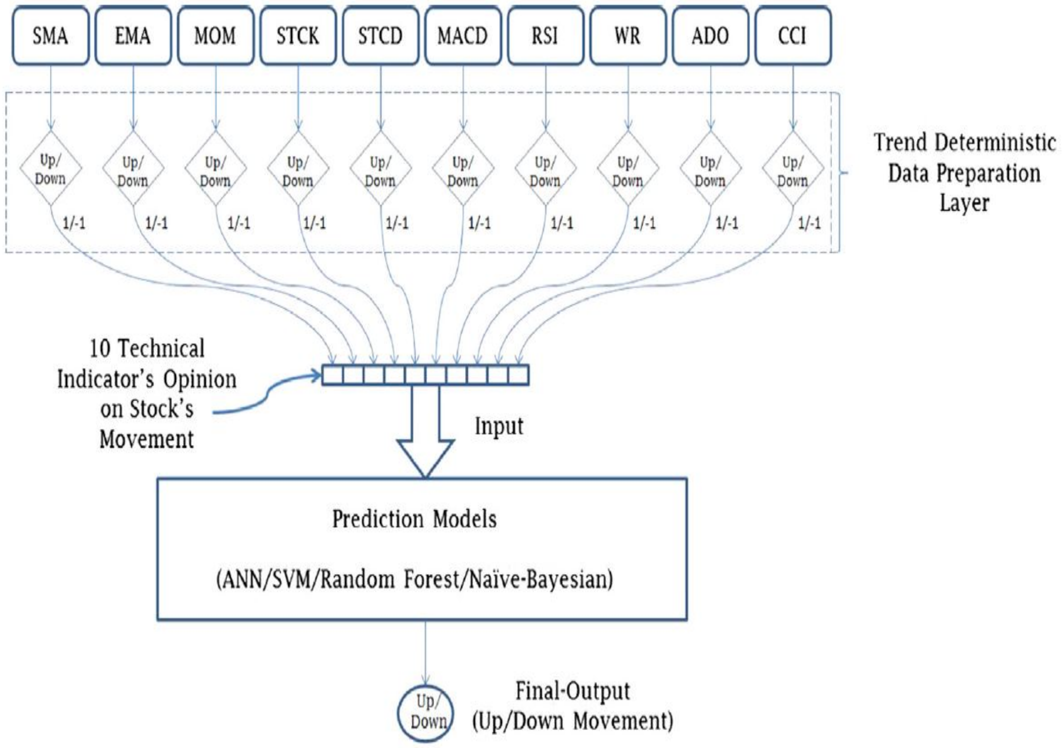


Figure 2.2 Trend Deterministic based data Classification model (Patel et al., 2015a).

a buy signal represented as +1, otherwise sell signal, represented as -1 (Patel et al., 2015a).

4. **EMA**: When the stock price candle is trading above EMA, it indicates its buy signal, and it is represented as +1. When the stock price candle is trading below EMA, it indicates a sell signal, and it is represented as -1 (Patel et al., 2015a).

5. **Larry William's (R%):** If R% technical indicator value is -80, then it indicates buy signal. It is represented as +1. If Williams, R% technical indicator value is -20, then it indicates sell signal. It is represented as -1 (Patel et al., 2015a).

Kara et al. (2011) proposed three-layer artificial neural networks and it is described in Fig 2.1. Ten technical indicators are given as input to the model. A Random forest-based prediction model is described in Fig 2.2. The continuous technical indicator features are converted into discrete data using the data preprocessing method.

In the literature Table 3.2, we have found that a limited number of technical indicators were considered to predict the stock price movements. This creates an opportunity to explore the different technical indicators and study how it is relevant for classification the stock price movements.

2.2 Stock Price Prediction

Lin et al. (2011) proposed business crisis classification based on the financial ratio features. The iterative relief method was used to select the financial ratio features. In this method, a high correlation feature greater than 0.9 was considered the best feature. Later, business crisis classification was carried out using the SVM method. The performance of the model was evaluated using the accuracy metric. The SVM model has outperformed the logit regression model.

Log Periodic Power Law (LPPL) (Jacobsson, 2009), (Sornette, 2009), (Johansen and Sornette, 1999) is a popular method to identify a bubble in the stock market. LPPL Singular method was considered to diagnose the bubble (Zhang et al., 2016b). In this work, the authors proposed the quantile regression method instead of standard least squares to diagnose bubbles. The standard least-squares method was not able to fit for outliers. This work (Filimonov and Sornette, 2013) discussed transforming a log-periodic power law and reduced nonlinear parameters in the model. The reduction of nonlinear parameters decreased the complexity of the model.

Junyu (2020) discussed financial crisis using machine learning. The work considered credit card default data to classify the financial crisis. Logistic regression, random forest, and XGboost machine learning were considered to classify the financial crisis. The accuracy metric was considered to evaluate the model performance. XGboost machine learning performed better than other models.

The overall related work is described in Table 2.2. Most of the work considered stock price prediction (Jin et al., 2019), (Göçken et al., 2019), (Zhang et al., 2018), (Li et al., 2016), (Singh et al.). There is limited work on stock crisis prediction. Chatzis et al. (2018) proposed stock crisis classification method.

Table 2.2 Related Works on Stock Crisis Prediction

Author	Techniques	Outcome	Gaps
Wosnitza and Denz (2013)	LPPL	Liquidity crisis	Physics-based study
Zhang et al. (2016b)	Quantile Regressions of Log- Periodic Power Law	Financial Crises	Quality of stocks are needs to be identified.
Li (2017)	LPPL	Chinese market bubble identification(Physics based study)	The intrinsic value of stock price needs to be estimated
Chatzis et al. (2018)	DNN	Stock Crises Classification	It is a classification task
Junyu (2020)	XGBoost	Financial Crisis Classification based on Credit card data.	Credit card default classification based on credit score

2.3 Stock Price Volatility Estimation

The study related to volatility forecasting in the financial world is comprehensive, and most of the empirical researches used ARCH and GARCH models for this purpose. The ARCH model was proposed in 1982 (Engle, 1982) and the GARCH model in 1986 (Bollerslev, 1986). GARCH model has proved to be a helpful technique when there is a conditional variance in financial data (Dieobold, 1986). The linear regression model was applied when the data shows homoscedastic, and the reason is that the variance of the error is constant (Yamaguchi, 2008). However, when the type of data is nonlinear, the error term is not constant over time. In such cases, heteroscedasticity is a crucial characteristic to develop a financial time series model.

Ariyo et al. (2014) considered Autoregressive Integrated Moving Average (ARIMA) model to predict the stock prices. Nigeria and New York stock exchange data were considered in the experiments. The ARIMA model predicts future values based on its current and past value of time series data. The Autoregressive Integrated Moving Average(ARIMA) model is widely used to find a linear relationship in the time-series application (Yürekli et al., 2005). However, most researchers found that the ARIMA model cannot identify the nonlinear pattern in data. Therefore, most of the methods considered SVM and ANN instead of the ARIMA model (Pai and Lin, 2005). Pai and Lin (2005) proposed a hybrid ARIMA model for stock price prediction. The hybrid ARIMA is the combination of the ARIMA and SVM models. The residuals are obtained using the ARIMA method, and these residuals are given as input to SVM for stock price forecasting. Blair et al. (2010) considered the ARCH model to estimate the volatility between daily returns and VIX. The study concludes that VIX volatility forecasting is outperformed compared to volatility forecast daily returns.

Atoi (2014) considered error distribution in evaluating the market volatility. In this work, the authors used GARCH, PGARCH, and EGARCH models to estimates the volatility. The authors used Normal, Student's-t, and generalized normal distributions to analyze the error distribution in the models and concluded that distribution is one of the essential parameters to improve the model performance. He has used Nigeria stock data from the year 2008 to 2013 for his experiments. The performance of the model was evaluated using the RMSE metric. In this study, the PGARCH model performs better than GARCH and EGARCH models.

The daily stock returns data were considered to evaluate the volatility of stock price (Gabriel, 2012). The data is collected from the Romania stock index from the year 2001 to 2012. In this work, the different family of the GARCH model was considered to forecast the stock returns. TGARCH and PGARCH model performs better than GARCH and IGARCH. The selection of the model is carried out using AIC and log-likelihood metrics. The future returns of the GARCH model were evaluated using RMSE, MAE, and MAPE.

Molnár (2016) proposed the range-GARCH model instead of the GARCH model. Here, the stock intraday difference from its highest and lowest price is considered range. Experiment work considered 30 stocks and six indices of Dow Jones stock market data. There were around 4423 samples of data collected from the year 1992 to 2010. The range GARCH model performs better than the standard GARCH model. The model selection was evaluated using the AIC metric. The one-day ahead future price is forecasted, and the model's performance is evaluated using the RMSE metric.

GARCH and ARCH model is considered for estimating the variance in stock prices (Chand et al., 2012). The residuals of each model were examined using a correlogram. However, the study was found that ARCH(1) could not capture the ARCH effects from its residuals. Because the residuals were generated using mean equations ARIMA(1,1,0). In this work, ARIMA(1,1,0)-GARCH(1,1) has captured the ARCH effect from its residuals, and the study concludes that GARCH performs better than the ARCH model. AIC and BIC selection criteria were considered to select the best model. The daily closing price of Muslim Commercial Bank was considered in the experiments. RMSE, MAE, MAPE metrics have been used for forecasting the stock price.

The volatility of the S&P 500 index was estimated using three GARCH models, namely GARCH, EGARCH, GJR-GARCH (Hajizadeh et al., 2012). AIC and BIC metrics were considered to select the best model. The performance of the EGARCH(3, 3) model was better than GARCH(1,1) and GJR-GARCH(1,1). To enhance the model performance, later EGARCH model residuals are given input to the ANN model. The ANN model is trained with the back-propagation method. 70% of data was considered for training, 20% for validation and 10% for testing. The experiment work considered S&P 500 data from 1998 to 2009. The performance of the model was evaluated using mean forecast error (MFE), RMSE, MAE, and MAPE.

Sharma et al. (2015) studied daily stock indices volatility forecasting using the seven GARCH models. The 21 global market indices were considered in the experiments from the year 2000 to 2013. AIC metric was considered to select the best model. The model parameters were estimated using the maximum likelihood function. The future performance of stock price was estimated using MSE and MAE metrics. The study found that the standard GARCH model performs better than TGARCH, EGARCH, AVGARCH, NGARCH, APARCH, GJR.

Abdullah et al. (2017) estimated the volatility of exchange rate in Bangladesh and the US currency using the GARCH models. The experiments considered data from the year 2008 to 2015. Normal and Student's t-distribution assumptions were considered in the GARCH models. The AR(2)–GARCH(1, 1) is performed better than EGARCH and TGARCH models.

Hu et al. (2020) proposed a hybrid model to estimate the volatility of gold price. The hybrid model is the combination of the ANN and GARCH model. The residuals are captured using the GARCH model, and it is given input to the ANN model for forecasting the price. The future performance of gold price was estimated using MSE, RMSE, and MAE metrics. The model was trained using the backpropagation method. The results shows that the hybrid model performs better than the GARCH model. Kristjanpoller and Minutolo (2015) estimated the volatility of the copper price using the hybrid deep learning method; this model is a combination of GARCH and RNN model.

In statistics, a structural break in the datasets leads to a massive difference in forecasting errors. Allaro et al. (2011) considered chow tests to identify the structural break in the datasets. In this method, time-series datasets are split into two equal parts, then the coefficient of two linear fits is compared to know whether the structural break is present or not.

Caporale and Zekokh (2019) investigated volatility of cryptocurrencies using Markov-Switching GARCH models. The GARCH models might be predicted incorrect results due to the high volatility. Therefore authors proposed regime-switching based on the MSGARCH method. The experiments considered Coindesk Bitcoin data from the year 2010 to 2018. Chen et al. (2019) proposed GARCH models for estimating volatility of wind data. Yancheng wind farm data were considered for the experiments. The wind power data were captured every 5 minutes. There were around 2016 data samples collected for work.

Fakhfekh and Jeribi (2020) investigated different types of GARCH models for volatility estimation. The work considered student-t and normal error distribution in the GARCH model. AIC and BIC information criteria were considered to evaluate the model. The performance of the EGARCH model was better than other models. CoinMarketCap data from August 2017 to December 2018 were used to carry out the experiments.

Sun and Yu (2020) used threshold GARCH method to analyze the positive and neg-

Table 2.3 Related Work on Stock Price Volatility Estimation

Authors	Methods	Datasets	Target Outcome
Song et al. (2020)	GARCH Model	S&P500	Simulation study of stock index and option data
Pan et al. (2017)	Regime-switching GARCH	WTI and Brent crude of U.S	Oil price volatility
Caporale and Zekokh (2019)	Markov Switch Garch	Bitcoin Data	Modeling volatility of cryptocurrencies
Chen et al. (2019)	GARCH Model	China Wind Power data	Volatility of wind power
Fakhfekh and Jeribi (2020)	GARCH Model with different error distribution	CoinMarketcap data	Volatility of cryptocurrencies
Sun and Yu (2020)	Threshold GARCH	S&P500	Modeling volatility of stock price
TRINH et al. (2020)	GARCH Model	Macroeconomic and Vietnam government bond data	Macroeconomic fundamentals and government bonds' volatility estimation
Emenogu et al. (2020)	Nine variants of GARCH models	Total Nigeria Plc data	Volatility of daily stock returns
Cao et al. (2020)	GARCH Model	S&P500	Volatility of VIX option
Sapuric et al. (2020)	EGARCH	Bitcoin–US	Bitcoin price volatility
Dhaene and Wu (2020)	GARCH Model	DJIA stocks	Volatility based on high-frequency intraday returns
Atoi (2014)	GARCH,PGARCH EGARCH Model	Nigeria stocks	Volatility Estimation
Molnár (2016)	GARCH,Range-GARCH Model	Dow-Jones stocks	Volatility Estimation
Hajizadeh et al. (2012)	Hybrid GARCH	S&P Data	Volatility Estimation
Sharma et al. (2015)	GARCH Family Model	Stock Indices of the Worldk	Volatility Estimation

ative news effects. The S&P 500 data were considered in the experiments. TRINH et al. (2020) investigated Vietnam government bonds price volatility using the GARCH TGARCH and EGARCH model. The study concludes that GARCH model performed better than TGARCH and EGARCH model. The experiments considered Vietnam government bonds data from 2006 to 2019.

Emenogu et al. (2020) investigated the volatility of stock price using the nine variants of GARCH models. NGARCH model performed better than other models. Nigeria Plc data from 2001 to 2017 were considered for experiments. Cao et al. (2020) investigated the volatility of VIX options data using the GARCH model. OptionMetric VIX and SPX options data were collected from 2008 to 2012 for experiments. Sapuric et al. (2020) studied the bitcoin volatility using the EGARCH model.

In most of the literature, stock price volatility is estimated using the GARCH model, and it is described in Table 2.3. Most of the work ignored the regime changes in volatility estimation. Therefore, we have applied an MSGARCH and SETAR model to estimate this work's volatility.

2.4 Outcome of Literature

The extensive literature reviews related work regarding stock price classification, stock price crisis prediction, and stock price volatility estimation and thus identified some of the following research gaps are listed below.

Research Gaps

- There are many financial parameters such as price to earnings, company returns, company debt, etc. Identification of relevant stock parameters is a challenging task.

- Most of the work considered stock price classification and prediction. However, stock price prediction during the crisis has not yet been studied in the literature.

- The stock price classification is a difficult task due to volatility in the financial market. In this work, we have estimated the volatility in the dataset and used conditional heteroscedastic models to predict the stock prices.

- Most of the work considered the most common ten technical indicators to classify the stock price movements. In this work, we have considered 33 technical indicators to classify the stock price movements.

- Most of the researchers have ignored structural breaks in the datasets. Therefore, we have considered regime-switching based on MSGARCH and SETAR models. To the best of our knowledge, this is the first empirical study on the Indian stock market data based on the regime-switching model.

- Nowadays, there is a possibility of a different kind of fraud or scam in the market, such as banking scams, housing scams, company accounts mismanagement, etc. Early prediction of such scams is one of the critical challenges in the stock market. However, we have not considered this challenge in our work.

2.5 Problem Statement

"To design and implement an effective model for the stock price prediction during the crisis, stock price classification, and stock price volatility estimation using statistical and machine learning techniques".

2.5.1 Research Objectives

1) To develop a stock price prediction during the crisis using the hybrid feature selection technique.

2) To develop an efficient approach to classify stock price movements based on the feature selection method.

3) To develop an efficient approach to estimates volatility of the stock price using the regime-switching techniques.

2.6 Summary

This chapter reviewed all the major state-of-the-art works in the area of stock price predictions. The literature work is divided into three tasks 1.Stock price movement classification and prediction, and it is discussed in Section 2.1. 2.Stock price crisis prediction is discussed in Section 2.2. 3. Stock price volatility estimation, and it is discussed in Section 2.3. The chapter concluded with an outcome of the literature review followed by a problem statement and research objectives.

Chapter 3

Stock Price Prediction during Stock Crisis

This chapter develops a model for the stock price prediction during the stock crisis using Hybrid Feature Selection (HFS) method.

The contribution of this chapter is as follows.

1. Developed a Hybrid Feature Selection (HFS) algorithm to remove irrelevant financial parameters. This is the first approach for stock price prediction during stock crisis using HFS method.

2. Naive Bayes method is used to classify the fundamentally strong stock.

3. Stock price bubble identification using the RSI method. Usually, RSI is used to identify whether a stock is overbought or oversold, but in this work RSI is used for bubble identification.

4. In this work, moving average is used to identify the stock crisis.

5. Prediction of the stock price during the crisis using the XGBoost and DNN method. To the best of our knowledge, this is the first time the stock price is predicted during the stock market crisis. This will help traders to trade even in stock market crisis.

3.1 Methodology

Zhang et al. (2016a) proposed the LPPL method to identify the bubble in stock prices. The bubble is nothing but the exponential growth of stock prices. This study has not been considered a fundamental analysis to recognize the quality of stock. Benjamn

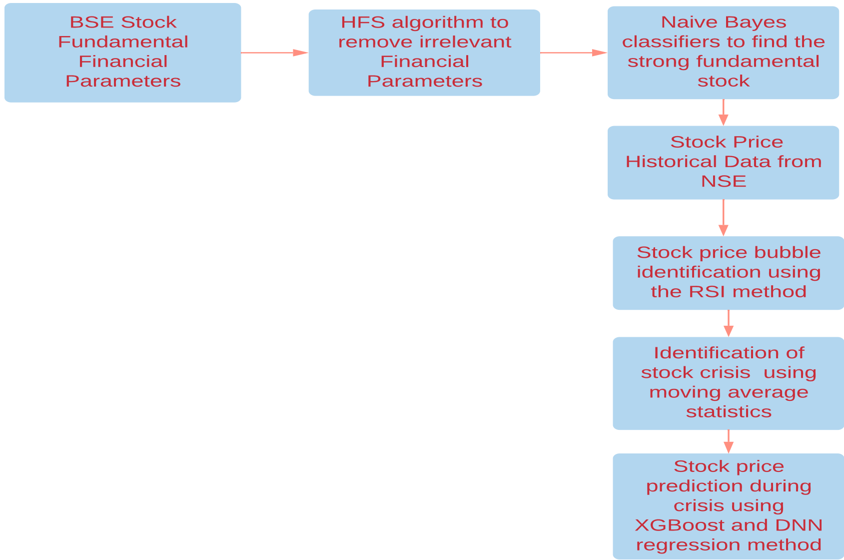


Figure 3.1 Overall proposed work.

Grahman (Graham, 1965) earned money in the stock market, but he has lost money during the stock crisis in 1929. Later he has published a book " The interpretation of financial statements" (Graham and McGolrick, 1975) on security analysis based on stock fundamentals. According to Benjamn Grahman, the stock price's fair value is based on company earnings, dividend, and asset value. Therefore we have considered these financial parameters to recognize the quality of stock.

The overall flow of the proposed work is depicted in Fig 3.1. We developed a Hybrid Feature Selection (HFS) algorithm for forecasting the future stock price during the stock crisis using the XGBoost and DNN regression method.

3.1.1 Hybrid Feature Selection(HFS) algorithm to remove irrelevant financial parameters

The fair value of the stock price depends on stock financial parameters. There are many financial parameters such as price to earnings, company returns, company debt, etc. Identification of relevant stock parameters is a challenging task. Therefore we developed a Hybrid Feature Selection (HFS) technique to select an essential financial parameters feature. HFS technique combines two individual algorithms, namely Recursive Feature Elimination (RFE) and Boruta Feature Selection (BFS). Also, we performed an intersection operation to the outcome of this combination. The proposed work is depicted in Fig 3.2. The step-by-step proposed HFS technique is described in Algorithm 1.

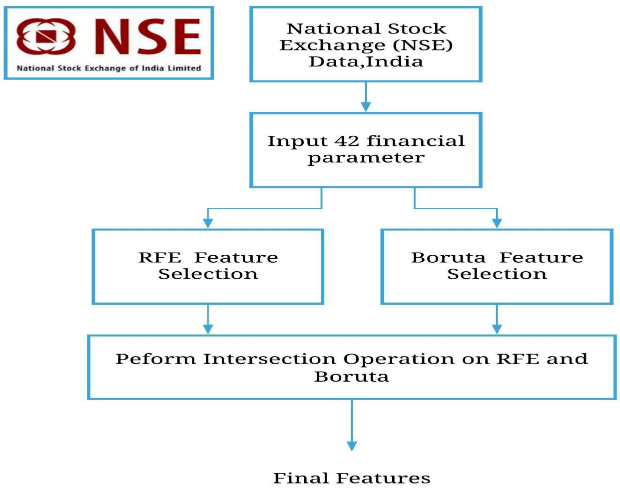


Figure 3.2 HFS feature selection.

Fundamentals of NIFTY50 stock financial parameters are retrieved from Bombay Stock Exchange (BSE), India (BSE, 2021). Stock financial parameters list are depicted in Fig 3.3. We have considered 42 various stock financial parameters of NIFTY50 stock.

The next task is identifying relevant stock financial parameters using the Recursive Feature Elimination (RFE) method. The 42 stock financial parameters are given as the input to the RFE to selects the best feature. The way RFE works is, it is a backward-compatible way of making feature selection. It starts with initially all the features, builds the model using a random forest regression method. Here price to earning (P/E) financial parameters is considered as target variable for regression. The next step is to remove the feature based on the Root mean square error (RMSE) score and build the model again. The RMSE score greater than nine is considered irrelevant feature (Lin et al., 2011).

Boruta feature selection (BFS) method is used to remove the irrelevant feature. The 42 stock financial parameters are given as the input to the BFS to selects the best feature.

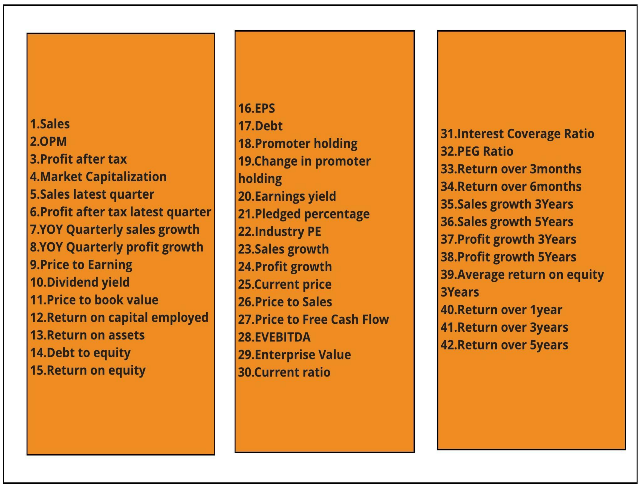


Figure 3.3 Stock financial parameters list.

Algorithm 1: Hybrid Feature Selection algorithm

1: Input the 42 stock financial parameters of the stock under consideration.
2: Select the best feature by performing a feature selection algorithm, namely Information gain(IG), RFE, and Boruta.
3: RFE Algorithm as follows:
4: For each financial parameter feature T_i, $i = 1..\mathrm{T}$
5: Financial parameters evaluation using random forest.
6: Keep important financial parameter feature T_i and remove weak financial parameters.
7: end RFE.
8: Boruta algorithm as follows:
9: Duplicates the dataset and shuffles the values in each column. These values are called shadow features.
10: Creates a shadow or duplicates feature of the financial parameters.
11: Train the random forest method to find important financial parameter features.
12: Each iteration, compare the original feature with the shadow feature Z score.
13: Remove the feature with the least Z score.
14: end Boruta .
15: Perform intersection operation RFE $\cap$ Boruta.
16: Output best feature.

The way boruta works is that it creates the shadow financial parameters, duplicates the dataset, and shuffles each column's values. Here Price to Earning (P/E) financial parameters is considered as target variable for regression. Next step is train the model using random forest regression to find important financial parameter.

The final selected features of the HFS method are given as input to the Naive Bayes to classify the quality stock.

3.1.2 Naive Bayes classifiers method to find the strong fundamental stock

Naïve Bayes is widely used in text classification and sentiment analysis to identify positive and negative sentiments (Srikanth et al., 2021), (Najjar and Al-augby, 2021), (Aditya et al., 2021). In this chapter, the NB classification method is considered to find the best strong fundamental stock based on stock financial parameters. Here price to earning (P/E) financial parameters is considered as target variable, i.e., probability of p (quality of stock). The Next is to calculate the individual probability of each stock financial parameter with the target variable. The probability of Financial Parameters (FP) and Quality Stock (QS) is defined in Equation 3.1. Here higher probability of stock is considered as fundamentally strong stock. From the NIFTY50 stock, the top high probability of fundamentally strong stock is considered in the experiments based on the Naive Bayes classifier.

$$P(FP \mid QS) = \frac{P(QS \mid FP)P(FP)}{P(QS)} \tag{3.1}$$

3.1.3 Stock price bubble identification using the RSI method

Relative Strength Index (RSI) statistics are used to find the bubble in stock price. The RSI technical indicator value ranges from 0 to 100. The RSI values below 30 indicate that the stock price is oversold, and RSI values above 70 indicate the overbought levels. When the RSI indicator value reaches above 70, there is a high chance that stock price is falling. Due to overprice in stock. We have considered the first 22 fundamentally strong stocks for computing the RSI. To compute the RSI value, we required the historical stock price data collected from the National Stock Exchange (NSE) portal. We have considered historical stock data from 2007 to March 2021. Next is to calculate the RSI value based on stock price using the below equations.

$$RSI(\#Days) = 100 - (100/(1 + Avg(Gain)/Avg(Loss))) \tag{3.2}$$

Most of the existing work RSI computed based on 14 days (Kara et al., 2011), (Patel et al., 2015a). However, in our approach, we have considered 200 days in RSI to find the stock price bubble. The reason is 14 days is used for intraday trading and not for the long term. The overprice in stock is nothing but a stock price bubble. The bubbles are captured using RSI statistics. The next step is the identification of stock crisis points based on the bubble of the stock price.

3.1.4 Identification of stock crisis using moving average statistics

After identification of bubble in stock price, the next step is stock price crisis identification. The identification of stock crisis points is carried out by using the moving average technique. We have considered two moving averages, 50 days and 200 days. The moving average is computed based on the stock price. The first moving average of 50 days indicates the stock price's short movements, and the second moving average of 200 days indicates long movements of the stock price. The short movements of stock price trades below its long price movements indicate the downtrend in stock price. Such data points are considered stock crisis points. Using moving average statistics, we have identified the stock crisis point. The next step is a prediction of future stock crisis points using XGBoost and the DNN regression method.

3.1.5 Stock Price Prediction during Crisis using XGBoost Regression Technique

From entire datasets, we have obtained the subset of stock bubble point using RSI method and stock crisis point using the moving average statistics. These subsets of datasets are given input to the XGBoost regression method. The feature of subsets datasets are described in Fig 3.4.

The linear regression model can be applied when the data is homoscedastic, i.e., when the error variance is constant (Aalen, 1989). However, stock price data is nonlinear, and the error term is not constant over time. XGBoost (Chen et al., 2015) and DNN

Figure 3.4 Stock crisis input data.

(Ronaghi et al., 2020) are more popular methods when the type of data is nonlinear. Therefore, we have considered XGBoost and DNN methods.

XGBoost machine learning is one of the methods to solve complex data-driven real-world problems. The stock price crisis data are given as input to the XGBoost method. The closing price of the stock is considered as the target variable for regression. The residuals differences between the observed and predicted values are calculated. Next XGBoost fits a regression tree to the residuals. This fitting is called as XGBoost tree. Each tree starts a single leaf, and all residuals are placed into the leaf. The next step is to calculate the similarity score for residuals to split the tree. The similarity score and output value are defined in Equations 3.3 and 3.4. The goal is to find an output value for the leaf is to minimize the residuals. For that, we square the output value from the new tree and scale it with λ. If $\lambda > 0$, then we will shrink the output value. Because we are optimizing the output value from the first tree, we can replace the previous prediction. Here lambada is setting to zero.

$$Similarityscore = \frac{sumofresidual, square}{numberofresidual + \lambda} \tag{3.3}$$

$$Outputvalue = \frac{sumofresidual}{numberofresidual + \lambda} \tag{3.4}$$

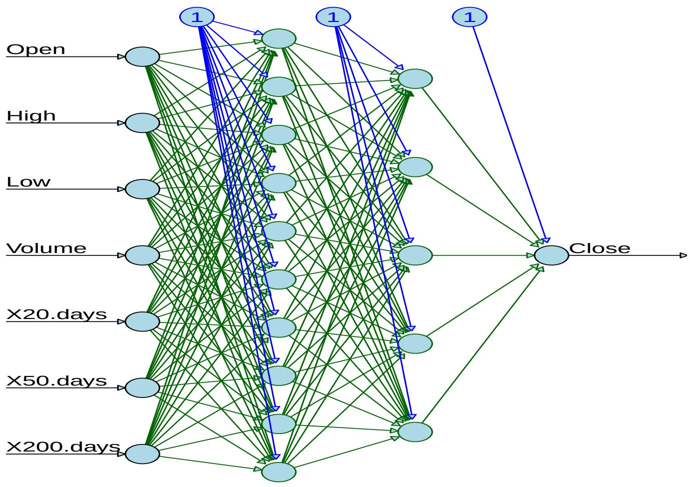


Figure 3.5 DNN method.

3.1.6 Stock Price Prediction during Crisis using DNN Regression Technique

A Neural network is most popular to deal with nonlinear data (Glezakos et al., 2009), (Demertzis et al., 2018), (Iliadis et al., 2007), (Bougoudis et al., 2014). The single neural network structure is defined in Equation 3.5. It has one layer and one activation function. The single neural network takes inputs and calculates the weighted sum of the inputs and adds a bias. This calculation is represented in the form of a transfer function. This calculated weighted sum is passed an input to an activation function to generate the output. Here W is weight, $B, B1, B2$ are bias, h is the hidden layer, h_1 is the hidden layer 1, h_2 is the hidden layer 2, and δ is the activation function.

$$h = \delta(Wh_1 + B) \tag{3.5}$$

In the proposed work, we have considered a deep neural network with 2 hidden layers, and it is defined as follows.

$$h_1 = \delta 1(Wh_1 + B1),$$

$$h_2 = \delta 2(Wh_2 + B2)$$

From entire datasets, we have obtained the subset of stock bubble point using RSI method and stock crisis point using the moving average statistics. These subsets of

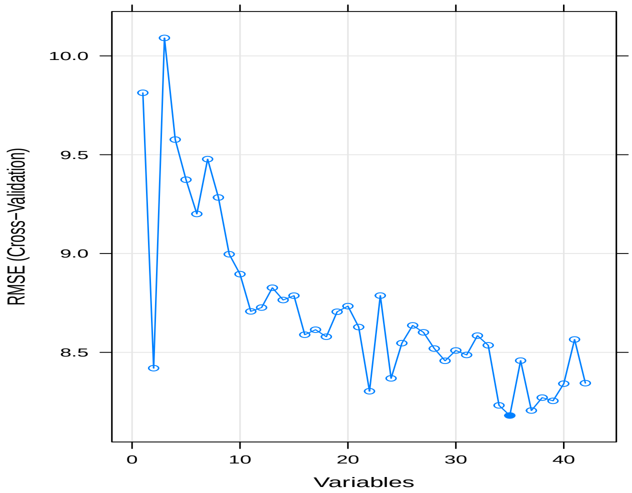


Figure 3.6 Feature RMSE score based on RFE.

datasets are given input to the DNN regression method. The feature of subsets datasets are described in Fig 3.4. The proposed DNN is depicted in Fig 3.5. Here closing price of a stock is considered as the target variable for the DNN regression method. We have normalized independent input variables by subtracting mean from each value, dividing them by standard deviation. The rectified linear unit activation function is considered in the hidden layer.

3.2 Experiment and Result Discussion

The data analysis is carried out using the R Studio environment. The stock financial parameters are retrieved from Bombay Stock Exchange (BSE), India (BSE, 2021). We have considered 42 various stock financial parameters of NIFTY50 stock and parameters list are depicted in Fig 3.3 in Section 3.1.1. In this work, we have considered HFS method to select the relevant features in stock financial parameters. HFS technique combines two individual algorithms, namely Recursive Feature Elimination (RFE) and Boruta Feature Selection (BFS). Also, we performed an intersection operation to the outcome of this combination. In RFE method the RMSE score is computed to select the best feature. The RMSE score greater than nine is considered irrelevant, and each feature RMSE score is depicted in Fig 3.6. In this work we have considered 42 features

Figure 3.7 Final Selected Feature based on RFE.

out of that 35 features are selected by RFE and it is depicted in Fig 3.7.

Later we have considered the BFS method to select the best feature of financial parameters. Using the BFS method, essential features are selected, and it is depicted in Fig 3.8. Based on the BFS method, 17 features are obtained as relevant, and the final feature list is depicted in Fig 3.9. We have performed an intersection operation on the outcome of the RFE and BFS feature selection method. Finally, we have obtained 13 features, depicted in Fig 3.10.

Next task is identification of strong fundamental stock based on stock financial parameters. For that we have considered NB classification method to find the best strong fundamental stock based on stock financial parameters. Here price to earning (P/E) financial parameters is considered as target variable. We have found the probability score for each stock using the NB method, and it is described in Table 3.1. The probability score is more significant than 0.5, considered a fundamentally strong stock, and less than 0.5 is considered as fundamentally weak stock.

The bubbles are captured using RSI statistics, and it is described in Table 3.2 and 3.3. The Table 3.2 and 3.3 describes the starting point of bubble in stock price. In

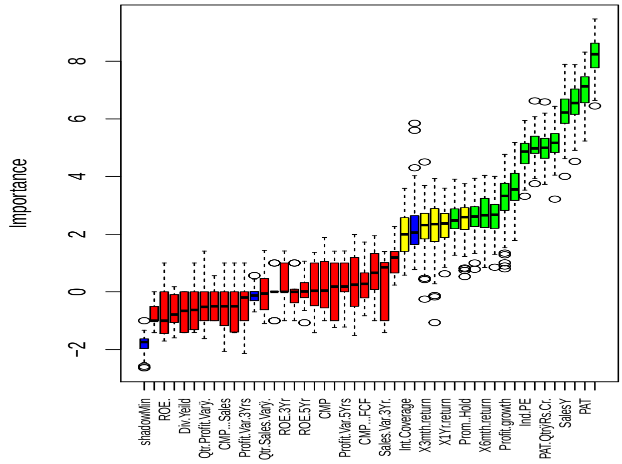


Figure 3.8 Important feature based on BFS.

Table 3.1 Probability Score using NB method.

Stock Name	Probability Score(Yes)	Probability Score(No)
Kotak Bank	0.6	0.4
ICICI Bank	0.6	0.4
Axis Bank	0.6	0.4
Infosys	0.65	0.35
Yes Bank	0.6	0.4

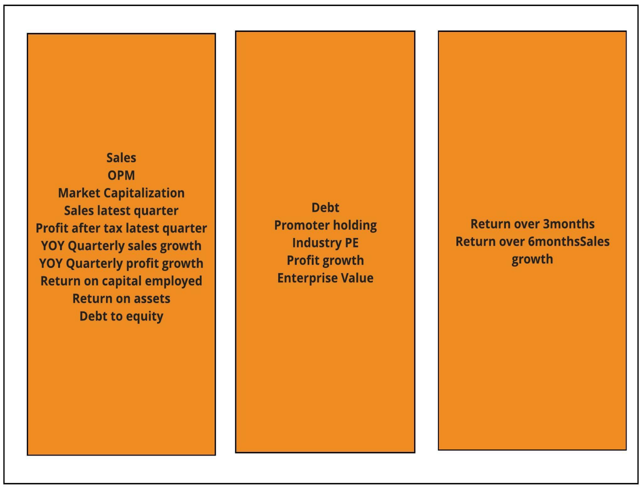


Figure 3.9 Final Selected Feature based on BFS.

Figure 3.10 Final Selected Feature based on HFS.

Table 3.2 the RSI value for all stock is greater than 70. This indicates the stock price is in overbought position. This point of stock price in our experiments is considered as bubble in stock price. We have captured the stock price data which has RSI value greater than 70 from entire datasets.

After identification of bubble in stock price, the next step is stock price crisis identification. We have considered two moving averages, 50 days and 200 days. The moving average is computed based on the stock price. If the stock price below 50 days and 200 days moving average then it is considered as starting point for stock crisis and it is depicted in Fig 5.3. In Fig 5.3 shows that the blue line indicates 50 days moving average, and the pink line indicates 200 days moving average. The stock crisis point is captured using moving average statistics, and it is described in Table 3.4 and 3.5.

From the datasets, we have captured stock price bubble point using RSI method and stock crisis point using the moving average methods. These subsets of datasets are given input to the DNN and XGBoost regression method. The partial datasets for Axis Bank data are described in Table 3.6.

The XGBoost and DNN regression method was used to predict stock price during stock crisis. In this work, we considered a few NIFTY 50 stocks in the experiments from January 2007 to April 2021. To obtain the best results, we have fine-tuned the parameters of the XGBoost regression method. We have varied the size of the decision tree from 100 to 600, and the learning rate increase from 0.001 to 0.3. For the DNN method, we have varied the learning rate from 0.001 to 0.3. To validate the model performance, we have considered the ten cross-fold validation. It is the most popular statistical method to validate the results. In this method, datasets are divided into training sets and test sets, and a test set was used to evaluate the model's performance. In our experiments, we have divided datasets into ten folds. 90% of data is considered for training, and 10% is for testing. For each cross fold, we have validated the results, and at the end, we have considered the average results of 10 cross folds.

Table 3.2 Stock price bubbles are captured using RSI statistics

Stock Name	RSI Value	Bubble Point Date	Bubble Price
Kotak Bank	76	16/Dec/2019	1698
	81	04/Oct/2010	255
	83	10/Dec/2007	348
ICICI Bank	81	23/Dec/2019	543
	79	01/Dec/2014	332
	81	08/Nov/2010	231
	79	29/Oct/2007	243
Axis Bank	75	03/June/2019	815
	72	06/Sep/2016	634
	81	02/Mar/2015	645
	82	04/Feb/2013	298
	80	04/Oct/2010	315
	86	14/Jan/2008	240
Yes Bank	78	31/Jul/2017	369
	81	04/Feb/2013	105
	75	10/Dec/2007	53
Bandhan Bank	83	06/Aug/2018	766
Indusland Bank	77	25/Jun/2018	2006
	79	27/May/2013	522
	87	10/Dec/2007	133
Tcs	79	15/Jan/2007	338
Infosys	83	20/Jan/2014	472
	79	03/Jan/2011	431
	83	02/Jan/2007	297
Hcl Tech	81	17/Feb/2020	616
	71	09/Mar/2015	525
	79	27/Mar/2006	82
Wipro	80	25/Feb/2019	291
	73	02/Mar/2015	253
	71	19/Feb/2007	154
Tech Mahindra	78	17/Feb/2020	844
	79	07/Feb/2015	738
	95	22/Jan/2007	465
Mind Tree	81	16/Feb/2020	1044
	80	04/Jan/2010	185
	79	19/Mar/2007	228
Hero Moto	75	04/Sep/2017	4005
	76	01/Dec/2014	3226
Eicher	77	04/Sep/2017	33021
	76	08/Oct/2007	493

Table 3.3 Stock price bubbles are captured using RSI statistics

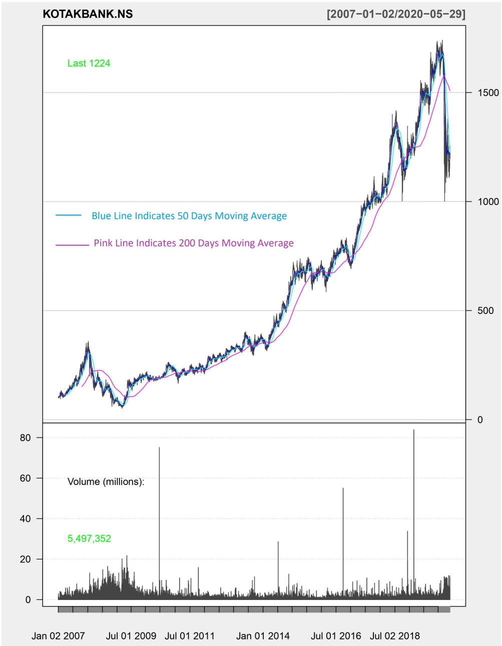
Stock Name	RSI Value	Bubble Point Date	Bubble Price
TVS	89	01/Jan/2018	775
	82	27/Jan/2015	315
	88	13/Sep/2010	79
	85	03/Apr/2006	89
Sun Pharma	75	03/Sep/2018	666
	81	06/Apr/2015	1170
	80	28/Apr/2008	149
Dr.Reddy	76	17/Aug/2015	4949
	83	08/May/2006	849
Cipla	71	17/Sep/2018	680
	71	25/Aug/2008	240
Torrent Pharma	74	21/Jan/2019	1893
	72	28/May/2007	132
AuroBindo	76	10/Sep/2018	822
	82	08/Nov/2010	129
	73	25/Jan/2007	83
Biocon	77	26/Feb/2018	323
	86	01/Nov/2010	72
	81	12/Nov/2007	55
Cadila	74	31/Jul/2017	549
	71	19/Oct/2015	447
	73	11/Jul/2011	196
	82	02/May/2006	106

Table 3.4 Stock price crisis point is captured using moving average statistics

Stock Name	**Crash Point Date**	**Crash Price**
Kotak Bank	02/Mar/2020	1618
	27/Dec/2010	226
	21/Jan/2008	265
ICICI Bank	02/Mar/2020	543
	09/Feb/2015	307
	10/Jan/2011	199
	04/Feb/2008	211.50
Axis Bank	15/Jul/2019	758
	26/Sep/2016	549
	20/Apr/2015	536
	18/Mar/2013	265
	22/Nov/2010	283
	03/Mar/2008	190
Yes Bank	23/Oct/2017	337
	17/Jun/2013	96
	21/Jan/2008	45
Bandhan Bank	24/Sep/2018	590
Indusland Bank	03/Sep/2018	1906
	22/Jul/2013	458
	28/Jan/2008	101
Tcs	28/May/2007	311
Infosys	18/Mar/2014	440
	14/Feb/2011	385
	05/Mar/2007	268
Hcl Tech	02/Mar/2020	568
	27/Apr/2015	452
	02/May/2006	73
Wipro	01/Jul/2019	279
	13/Apr/2015	229
	26/Feb/2007	133
Tech Mahindra	02/Mar/2020	774
	13/Mar/2015	671
	25/Jun/2007	365
Mind Tree	09/Mar/2020	855
	25/Jan/2010	155
	09/Jul/2007	197
Hero Moto	09/Oct/2017	3759
	19/Jan/2015	2929
Eicher	27/Nov/2017	30367
	17/Dec/2007	427

Table 3.5 Stock price crisis point is captured using moving average statistics

Stock Name	**Crash Point Date**	**Crash Price**
TVS	29/Jan/2018	697
	23/Mar/2015	265
	13/Dec/2010	72
	22/May/2006	68
Sun Pharma	22/Oct/2018	592
	27/Apr/2015	957
	06/Oct/2008	140
Dr.Reddy	02/Nov/2015	3974
	29/May/2006	682
Cipla	22/Oct/2018	628
	06/Oct/2008	224
Torrent Pharma	22/Apr/2019	1793
	23/Jul/2007	114
AuroBindo	17/Dec/2018	741
	24/Jan/2011	121
	23/Jul/2007	71
Biocon	16/Jul/2018	303
	17/Jan/2011	64
	14/Jan/2008	44
Cadila	14/Aug/2017	492
	30/Nov/2015	396
	01/Aug/2011	175
	22/May/2006	82



(a) Kotak-Bank

(b) ICICI-Bank

(c) Axis-Bank

(d) Yes-Bank

Figure 3.11 Stock Crisis Point

Table 3.6 Partial Datasets for Axis Bank.

Open	High	Low	Volume	20 Days	50 Days	200 Days	Close
813.55	814.65	803.1	6427567	768.46	762.77	669.38	812.65
807.55	827.75	805.5	9515354	772.51	764.29	670.62	822.8
824	825.95	805.8	14728937	776.12	765.22	671.68	807.7
811.1	813.25	801.3	6884094	779.77	766.19	672.73	804
807.4	816.75	807.4	5923875	783.72	767.33	673.82	814.15
817.15	821.45	807	5471529	787.86	768.71	674.79	814.8
814.05	815.75	809.1	4707478	791.97	769.84	675.78	813.55
811.25	823.4	810.75	9056584	796.88	771.06	676.86	820.15
820.1	823.1	797.45	8919370	800.35	771.47	677.77	801.2
800.1	801	775.1	10151212	801.78	771.71	678.54	777.7
781.95	784.85	770.1	8347632	801.45	771.88	679.29	775.75
780.8	789	763.35	6411258	801.34	772.06	680.02	770.7
772	775.45	758.85	9190732	800.95	772.25	680.69	771.4
774.75	778.25	762.8	12610688	800.68	772.42	681.39	771.05
771	775.25	761	4866859	799.17	772.59	682.01	762.85
761.25	783	756.55	10044436	797.55	772.95	682.66	781.65
782.9	792	781.7	8715757	796.56	773.51	683.31	788.6

The loss function is described in Fig3.12. After 2000 iterations, there is a significant drop in the error.

The model's performance is evaluated based on MSE, MAE, and RMSE score and it is defined in Equations 3.6, 3.7 and 3.8. Here y_i represents the observed value. x_i represents the predicted value and n is the total number of elements in datasets. The proposed HFS based XGBoost performs better than the DNN method, described in Table 3.7. Table 3.7 using HFS based XGBoost method shows that the lowest RMSE value for Kotak Bank, ICICI Bank, Axis Bank, and Infosys are 11.9457,4.43981, 4.237013, and 28.82983.

$$mse = (\frac{1}{n})\sum_{i=1}^{n}\left|(y_i - x_i)^2\right| \tag{3.6}$$

$$mae = (\frac{1}{n})\sum_{i=1}^{n}|y_i - x_i| \tag{3.7}$$

$$rmse = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y_i - x_i)^2} \tag{3.8}$$

We have considered the Friedman test (O'Gorman, 2001) to validate the post results of the DNN and XGBoost method, and it is defined in Equation 3.9. K is a number of the prediction model, N is the total number of elements, R_j is the sum of the ranks for the j prediction model.

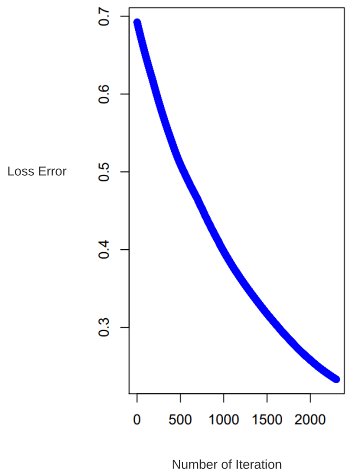
$$\frac{12}{NK(K+1)}\sum_{j=1}^{K}R_j^2 - 3N(K+1) \tag{3.9}$$

To check the results of DNN and XGBoost prediction model are significant or not, we have defined the null hypothesis, and alternate hypotheses are given below.

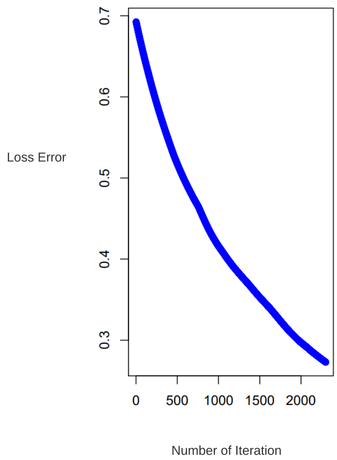
H_0: The result of the DNN and XGBoost prediction model are the same.

H_1:The results of the DNN and XGBoost prediction models are different.

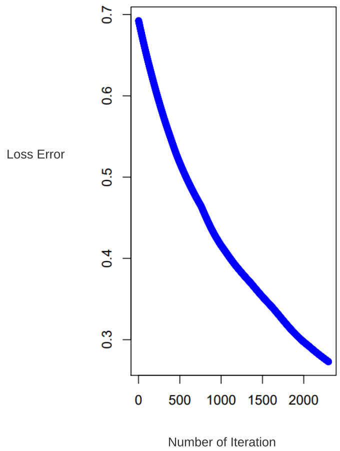
For Kotak Bank stock, we validated the result using the Friedman test. We found chi-squared value is 28.6243, df = 1 and p-value 0.0364. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the results of the DNN and XGBoost prediction model are the same for Kotak Bank stock. For Axis Bank stock, we found chi-squared value is 22.278, df = 1, and p-value 0.0431. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the DNN and



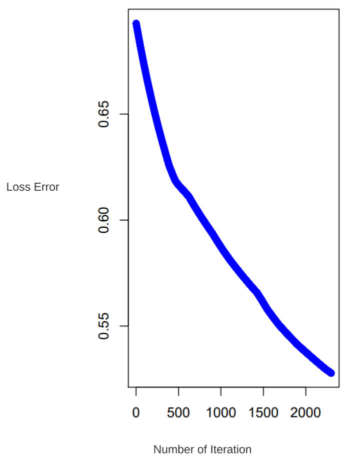
(a) Kotak Bank



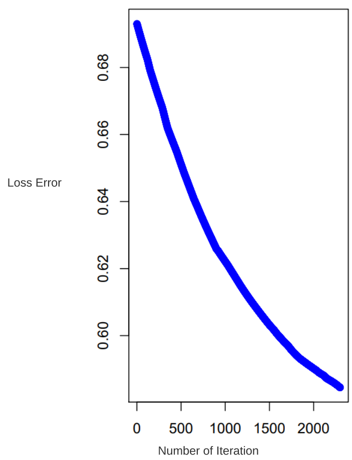
(b) ICICI Bnak



(c) Axis Bank



(d) Yes Bank



(e) Infosys

Figure 3.12 Loss function error rate

Table 3.7 Results.

Stock Name	Prediction Model	MSE	MAE	RMSE
Kotak Bank	HFS based XGBoost	142.6998	7.918674	11.9457
ICICI Bank	HFS based XGBoost	19.71191	2.486814	4.43981
Axis Bank	HFS based XGBoost	17.95228	2.440138	4.237013
Infosys	HFS based XGBoost	831.1591	21.00927	28.82983
Yes Bank	HFS based XGBoost	47.69003	5.190128	6.905797
Kotak Bank	HFS based DNN	38518.46	147.2966	196.2612
ICICI Bank	HFS based DNN	310.6973	13.5881	17.62661
Axis Bank	HFS based DNN	564.9369	14.07273	23.7684
Infosys	HFS based DNN	28.28089	4.512494	5.317978
Yes Bank	HFS based DNN	192.8784	7.667211	13.88807

XGBoost prediction model results are the same for Axis Bank stock. For Infosys stock, we found chi-squared value is 4.2118, df = 1, and p-value 0.04014. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the outcome DNN and XGBoost model are the same for Infosys stock. For ICICI Bank stock, we found chi-squared value is 0.0074349, df = 1, and p-value 0.01765. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the DNN and XGBoost prediction model results are the same for ICICI Bank stock. For YES Bank stock, we found chi-squared value is 0.0027855, df = 1 and p-value 0.0324. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the DNN and XGBoost prediction model results are the same for YES Bank stock. Friedman statistical test are described in Table 3.8. Based on the Friedman statistical test, we concluded that the DNN and XGBoost prediction model results are significant.

Table 3.8 Friedman test.

Stock	Chi-squared value	p-value
Kotak Bank	28.6243	0.0364
Axis Bank	22.278	0.0431
Infosys	4.2118	0.04014
ICICI Bank	0.0074349	0.01765
Yes Bank	0.0027855	0.0324

To the best of our knowledge, this is the first time the stock price is predicted during stock crisis. Hence we have not compared the proposed model results with existing works.

The stock prices are affected due to many events such as company balance sheet variation, political uncertainty, bond market rate, and global market trends. Sometimes, stocks' prices react when there is a sudden change in management or share dividend and bonus announcement. Financial market stock price movements purely depend on various sources of information. It is not easy to interpret information from different sources. Aggregating and processing information from various platforms is a crucial challenge for future work.

3.3 Summary

Stock price prediction during the stock crisis is difficult due to more volatility in the stock market. To the best of our knowledge, this is the first approach to address stock price prediction during the stock crisis. We have developed a Hybrid Feature Selection (HFS) algorithm to remove irrelevant stock financial parameters. The NB classifier method is considered to find the fundamentally strong stock. Later, stock price bubbles are identified by using the RSI method. Moving average statics are considered to identify the stock crisis points. Stock price is predicted during stock crisis using the XGBoost and DNN regression method. The performance of the model is evaluated based on MSE, MAE, and RMSE. The effectiveness of XGBoost and DNN models is quantified by using the Friedman test. HFS based XGBoost performs better than HFS based DNN method. Therefore in future work, different fundamentals stock and technical parameters can be applied to improve the model performance. We have explored a limited number of technical parameters of stock prices. In future, the researchers can explore other technical indicators to predict the crisis point. There is more scope to improve and fine-tune the XGBoost method with different optimizers. Parameters optimization for XGBoost and DNN methods using evolutionary algorithms can be also be done as future work.

Chapter 4

Stock Price Movements Classification and Prediction based on Technical Indicators

This chapter proposes a model for the stock price movement classification using the feature selection method. The objectives of this chapter are defined below.

1. Boruta feature selection is used to select the best features out of the 33 technical indicators and classified the stock price movements using the Deep Neural Network (DNN), Artificial Neural Network (ANN), and Support Vector Machine (SVM).

2. Proposed a novel hybrid technique for stock price movements classification based on candlestick data and technical indicators

Several studies have focused on technical indicators to classify stock price movements (Kara et al., 2011). These studies considered the ten most popular technical indicators, including moving average, weighted moving average, momentum, stochastic K, stochastic D, MACD, Larry William R, A/D oscillators, commodities channel index, and the relative strength index. Although few research has explored the impact of more than ten technical indicators on classification accuracy, therefore we have considered 33 technical indicators to increase classification accuracy.

4.1 Feature Selection based Stock Price Movements Classification Model

Technical analysis is one of the popular methods to measures the future price of the stock. It involves statistical analysis to identify the trend of the stock price based on

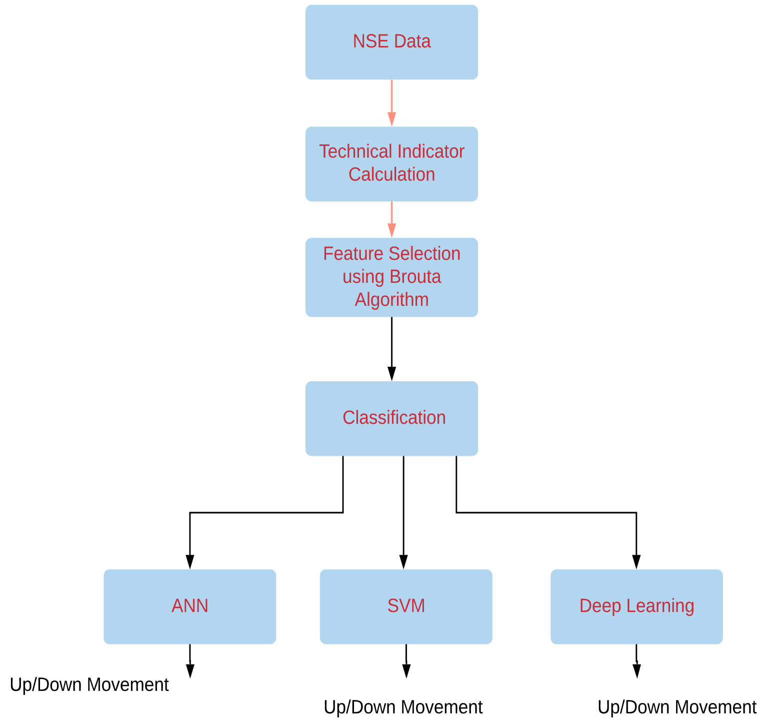


Figure 4.1 Overall proposed work.

market data, such as open-stock price, close price, day high price, day low price, and volume of the stock. Technical analysis is different from fundamental analysis. The technical analysis measures a stock by using the historical stock price data like the closing of stock price and volume. However, the fundamental analysis does not consider the historical stock price data to estimates the stock prices. In fundamental analysis, stock evaluation is based on company financial data such as P/E, EPS, ROE, ROCE, etc. The fundamental analysis is used to find the quality of stock.

In the literature, we have found that a limited number of technical indicators were considered to classify the stock price movements. In this work, we have considered 33 technical indicators to classify the stock price movements. However, the identification of relevant stock technical indicators is a challenging task. To the best of our knowledge, Boruta feature selection method (Kursa et al., 2010) is not considered for technical indicators selection. Therefore, we have used the Boruta feature selection method to identify the relevant stock technical indicators in this work.

Figure 4.2 Technical Indicators used.

The flow of the proposed model is described in Fig 4.1. The data are retrieved from National Stock Exchange (NSE), India. We have considered a few Nifty50 stocks like ICICI Bank, SBI Bank, Yes Bank, and Kotak Bank. In this study, we have considered 33 different technical indicators for stock price classification, and it is depicted in Fig 4.2.

The few technical indicator descriptions are listed below

- Simple Moving Average (SMA): It is a mathematical moving average computed by adding price history and dividing the result by the number of periods (Anbalagan and Maheswari, 2015).

- Exponential Moving Average: : It is a moving average (MA) in which the most recent data points are given a higher weight than the older ones (Sezer and Ozbayoglu, 2018).

  EMA=(closing price previous day's EMA)× smoothing constant as a decimal + previous day's EMA

- Momentum Indicator : Momentum indicators are used to assess the strength or weakness of a stock's price over a period of time. Momentum is a measure of the rate at which stock prices are rising or falling (Leivo and Pätäri, 2011).

  Momentum Oscillator = (Price today / Price n periods ago) x 100

- Stochastic oscillator: STOCHASTIC Oscillators (also known as stochastic oscillators) are momentum indicators that measure the relationship between a security's closing price and a range of prices over a specified period (Gulisashvili, 2012).
  %K=(C-L14/ H14 -L14)×100
  where: C = The most recent closing price
  L14 = The lowest price traded of the 14 previous trading sessions
  H14 = The highest price traded during the same14-day period
  %K = The current value of the stochastic indicator

- Moving Average Convergence Divergence: Movement of the moving average convergence divergence (MACD) is a trend-following momentum indicator that depicts the relationship between the price of an asset and two moving averages of its movement (Patel et al., 2015a).
  MACD=12Period EMA - 26Period EMA

- Relative Strength Index:Technical analysts use the relative strength index (RSI) to determine whether a stock or other asset is overbought or oversold. The RSI evaluates the magnitude of recent price fluctuations and is used to determine whether a stock or other asset has reached an overbought or oversold situation (Rodríguez-González et al., 2011).

  RSI = (100 – 100 / (1 + RS)
  RS = Gain on Average/Loss on Average

- Williams R: Williams Percent Range (or Williams percent R) is a form of momentum indicator that ranges between 0 and 100. This is used to identify whether a stock is overbought or oversold (Kara et al., 2011).

Wiliams %R=Highest High-Close/ Highest High-Lowest Low

where Highest High=Highest price in the lookback period, typically 14 days.

Close=Most recent closing price.

Lowest Low=Lowest price in the lookbackperiod, typically 14 days.

Algorithm 2: Boruta Feature Selection

1: Input 33 technical indicators F.

2: Create duplicate/shadow copies of technical indicators D.

3: Do the random shuffle of original technical indicators F and duplicate copies of technical indicator D.

4: Apply random forest algorithm to find important technical indicators.

5: Calculate $Z = \frac{(F-\mu)}{\sigma}$.

6: Select technical indicators which has highest Z score value.

These technical indicators are computed based on formulas (Kara et al., 2011) which are described in Table 4.1.

The proposed task is carried out by using two approaches. First Boruta Feature Selection (BFS) method is used to find the important features of a technical stock indicator. Initially, the BFS method creates duplicate/shadow copies of input features in the dataset. Then we create a model which includes the shadow features and original features and evaluates the importance of each feature using the random forest method. Here the closing price is the target variable. A random forest algorithm has been applied to find important technical indicators based on higher mean values (Z). In this algorithm, we have considered the Z score threshold value as 0.80. If any technical indicator feature has a threshold value is greater than 0.80, then it is considered an important feature (Kursa et al., 2010). The step-by-step proposed Boruta feature selection algorithm is stated in Algorithm 2. The ICICI Bank and SBI Bank stock are considered for experiments. These stocks are trading with significant volumes in the NSE market.

Table 4.1 Technical Indicators and its formulas (Kara et al., 2011)

Technical Indicator Name	**Calculation**	**Number of days**
Simple Moving Average (SMA)	(C_t+C_{t-1}+....C_{t-n+1})/n	5,10,14,30,50,100,200
Exponential Moving Average	(C_t-$SMA(n)_{t-1}$) * (2/n+1) +$SMA(n)_{t-1}$	5,10,14,30,50,100,200
Momentum Indicator	C_t-C_n-9	5,10,14
Stochastic oscillator K	100*((C_t - L_t(n)) / (H_t(n) - L_t(n))	14
Stochastic oscillator D	(100*((C_t - L_t(n)) / (H_t(n) - L_t(n)))/3	14
Moving Average Convergence Divergence	$SMA(n) - SMA(n)$	26,13,19,45,25,15
Relative Strength Index	100 - (100 / (1 + Avg(Gain)/Avg(Loss)))	14,28
Williams R	((H_t-C_t)/(H_n-L_n))x 100	14,28,50,100
Accumulation distribution index	H_t-C_{t-1}/H_t-L_t ((C_t / L) / (H / C)) / (H / L)	14
Commodity Channel Index	(H+L+C/3) - SMA)/ (0.015 * Mean deviation)	14,50,100

Using the BFS method, we have obtained 23 best technical indicators in ICICI Bank stock and 22 best technical indicators in SBI Bank stock. These technical indicators are given as input to the classification model like DNN, ANN, and SVM models for stock price classification prediction.

4.1.1 Stock Price Classification using DNN

DNN is extensively used in medical image classification, extensive data analysis, and electronic health record analysis. Nowadays, these methods gain the highest accuracy compares to the other machine learning algorithm. However, (Kara et al., 2011) was proposed as a framework for stock classification using ANN and SVM. Therefore we have considered the DNN method for stock price classification.

In the proposed work, we have considered DNN model (Candel et al., 2016). First, we have used the Boruta feature selection method to find the relevant technical indicators and removed irrelevant technical indicators. Later, the relevant technical indicators are given as input to the DNN model. The DNN model is used to classify the stock price up and down movement, and it is described in Fig 4.3.

It has five layers of interconnected neuron units through which data is transformed. The input layer neurons represent the technical indicators, and it is denoted by t_i. W_i denotes the weights of the neurons. Each input data is multiplied by a weight, and a summation of weighted input data was given to neurons. Stochastic gradient descent with back-propagation has been used to adjust the weight. Bias input is given to each layer except the output layer of the model. The objective function $L(W,Bias|j)$ aims is to reduce the classification error in the data.

The weighted combination of input summation is denoted in Equation 4.1.

$$\alpha = \left(\sum_{i=1}^{n} W_i t_i + Bias \right) \tag{4.1}$$

The activation function Tanh and rectified linear units are used. The model supports the regularization function to avoid overfitting, as shown in Equation 4.2.

$$L(W,Bias|j) = L(W,Bias|j) + \lambda 1 R1(W,Bias|j) + \lambda 2 R2(W,Bias|j) \tag{4.2}$$

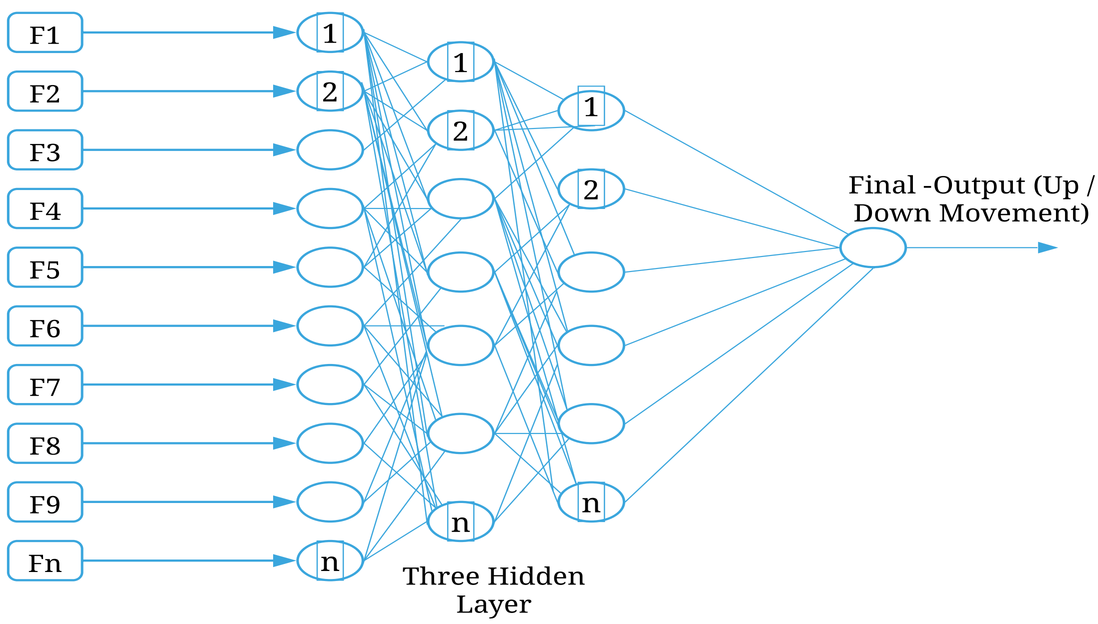


Figure 4.3 Proposed five layer Deep Learning model

4.1.2 Stock Price Classification using ANN

We have used the Boruta feature selection method to find the relevant technical indicators and removed irrelevant technical indicators. The selected relevant technical indicators are given as input to the ANN model. In this work, the ANN is used to classify the stock price movements. ANN has three layers, and each layer is connected to the other. The neurons represent the technical indicators. The sigmoid function activation function is used in the ANN model. The back-propagation method was considered to train the model. The threshold value of 0.5 has been set. A gradient descent momentum parameter is considered to determine the weights and to reduce the global minimum.

4.1.3 Stock Price Classification using SVM

The SVM prediction model is used to predict the stock price movements. SVM is based on the VC learning theory, and one of its major components was developed by Vapnik (Vapnik and Chervonenkis, 1974), (Boser et al., 1992). SVMs are also showing strong performances in real-world applications. SVM hyperplane is constructed based on input vectors. To separate the input vector, two hyperplanes are constructed. Stock market data are non-linear separable datasets, and SVM can be more effective when datasets are non-linear.

Polynomial and radial basis kernel functions are shown in Equation 4.3 and Equation 4.4.

$$PolynomialFunction : K(f_i, f_j) = (f_i.f_j + 1)^d \tag{4.3}$$

$$RadialBasisFunction : K(f_i, f_j) = exp(\gamma||f_i - f_j||^2) \tag{4.4}$$

$K(f_i, f_j)$ are input vectors and $f_i - f_j^2$ represent the squared euclidean distance.

The degree of a polynomial function is represented by d, and γ represents the constant of radial basis function. We have varied the SVM parameter's degree value from 1 to 4, and gamma is 0.1 to 5.

4.2 Stock Price Classification based on Candlestick Data and Technical Indicators.

Kara et al. (2011) proposed model for stock price classification using a three-layer artificial neural network, and it described in Fig 4.4. The work considered technical indicators as input variables for ANN. Hu et al. (2019) proposed stock price classification using a candlestick pattern. However, the combination of candlestick data and technical indicators based on stock price prediction are not found in the literature. Therefore we have proposed a method to identify a trend in data using a combination of candlestick data and technical indicators to classify the stock price movements.

An overall framework for a stock price classification model is described in Fig 4.5. First, we have to extract the candlestick data, and second is technical indicators data. These combined data are given as input to the proposed algorithm to find the trend in data.

Technical indicators computations are described in Algorithm 3. In this algorithm, we have considered ten technical indicators and computed technical indicators using

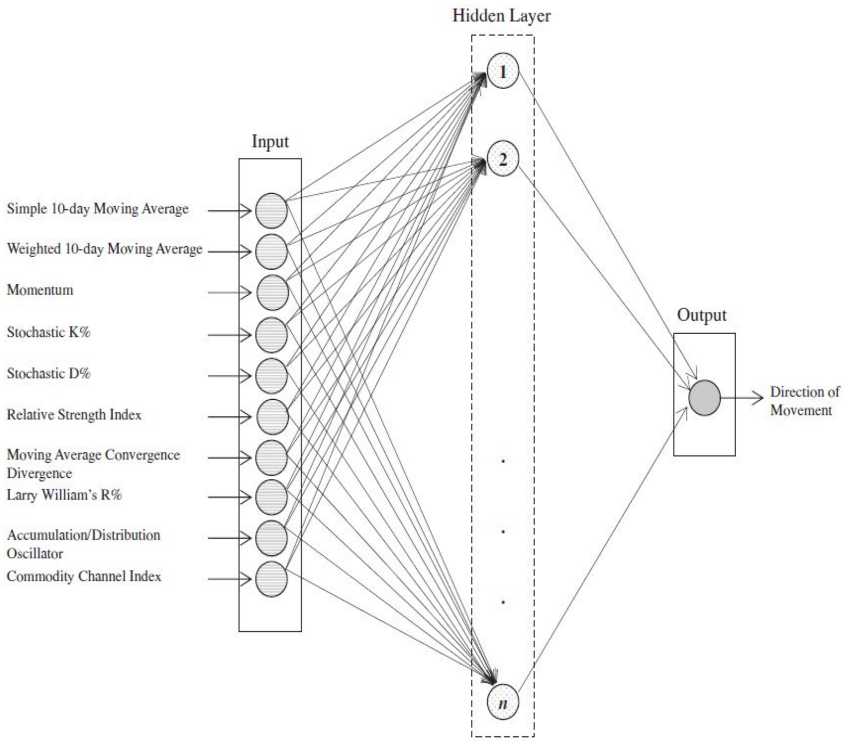


Figure 4.4 ANN Classification model Kara et al. (2011).

Algorithm 3: Technical Indicators Computation

1: Input NSE Data .

2: Calculate the technical indicators using (Kara et al., 2011) methods.

3: Output technical indicators.

(Kara et al., 2011) methods.

Later, candlestick pattern data calculations are described in Algorithm 4. The candlestick data are retrieved from the NSE portal. In this algorithm, we have computed the mean value of candlestick data.

We have proposed a method to identify a trend in data using a combination of candlestick data and technical indicators, and it is described in Algorithm 5. In this algorithm, line number 1, if SMA (Simple Moving Average) technical values are trading above the past ten days average value of the stock price and candlestick pattern data is greater than the mean value of stock price, then it is considered as up movement,

Algorithm 4: Candlestick Pattern Data

1: Input NSE Data .

2: Extract day low, high, open and close price.

3: Compute mean value of candlestick data.

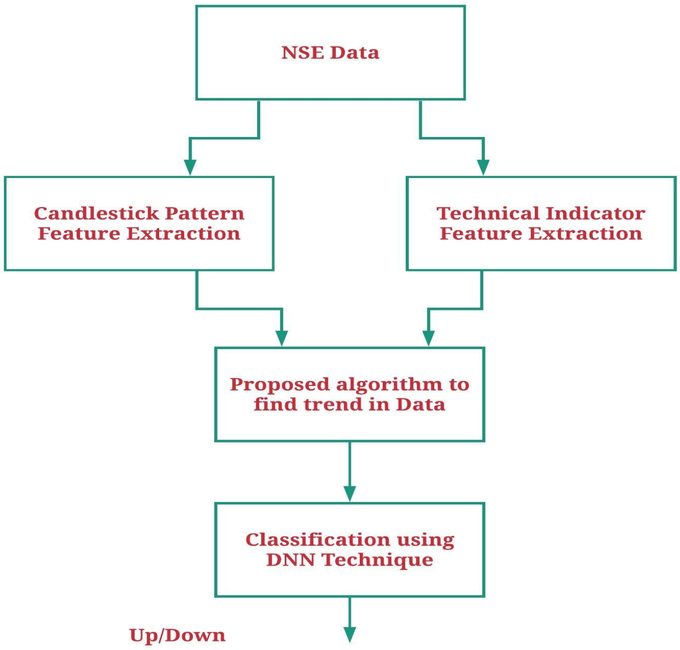


Figure 4.5 Overall Framework of Proposed model.

otherwise down movement. In line number 2, if EMA (Exponential Moving Average) technical values are trading above the past ten days average value of the stock price and candlestick pattern data is greater than the mean value of the stock price, then it is considered as up movement, otherwise down movement. In line number 3, if MOM (Momentum) technical values are trading above the past ten days average value of the stock price and candlestick pattern data is greater than the mean value of the stock price, then it is considered as up movement, otherwise down movement. In line number 4, if STCK (Stochastic crossover K) technical values are trading above STCD 14 day average value of the stock price and candlestick pattern data is greater than the mean value of the stock price, then it is considered as up movement, otherwise down movement. In line number 5, if RSI (Relative Strength Index) technical values are trading above

Algorithm 5: Proposed Algorithm to find trend in data

Input : Candlestick pattern Data and continues technical indicator values.

for *each technical indicator i=1 to 10* **do**

- **if** *(SMA >10 Days price AND Candlestick Pattern >Mean Value)* **then**
  T1=Up else T1=Down
- **if** *(EMA >10 Days price AND Candlestick Pattern >Mean Value)* **then**
  T2=Up else T2=Down
- **if** *(MOM >10 Days price AND Candlestick Pattern >Mean Value)* **then**
  T3=Up else T3=Down
- **if** *(STCK Fast Moving 14 DAYS >STCD Slow moving 14 days price AND Candlestick Pattern >Mean Value)* **then**
  T4=Up else T4=Down
- **if** *(RSI >30 price AND Candlestick Pattern >Mean Value)* **then**
  T5=Up else T5=Down
- **if** *(MACD Fast Moving 9 >MACD Slow Moving 26 days price AND Candlestick Pattern >Mean Value)* **then**
  T6=Up else T6=Down
- **if** *(R <-80 price AND Candlestick Pattern >Mean Value)* **then**
  T7=Up else T7=Down
- **if** *(A/D <-100 AND Candlestick Pattern >Mean Value)* **then**
  T8=Up else T8=Down
- **if** *(CCI >100 AND Candlestick Pattern >Mean Value)* **then**
  T9=Up else T9=Down
- **if** *(WMA >20 Days price AND Candlestick Pattern >Mean Value)* **then**
  T10=Up else T10=Down

End for.

30 and candlestick pattern data is greater than the mean value of the stock price, then it is considered up movement, otherwise down movement. In line number 6, if MACD (Moving Average Convergence Divergence) technical values are trading above MACD 26 day average value of the stock price and candlestick pattern data is greater than the mean value of the stock price, then it is considered as up movement, otherwise down movement. In line number 7, if R (Williams R) technical values are trading below -80 and candlestick pattern data is greater than the mean value of the stock price, then it is considered as up movement, otherwise down movement. In line number 8, if A/D (Accumulation Distribution Index) technical values are trading below -100 and candlestick pattern data is greater than the mean value of the stock price, then it is considered as up movement, otherwise down movement. In line number 9, if CCI (Commodity Channel Index) technical values are trading above 100 and candlestick pattern data is greater than the mean value of the stock price, then it is considered as up movement, otherwise down movement. In line number 10, if WMA(Weighted Moving Average) technical values are trading above the past 20 days average value of the stock price and candlestick pattern data is greater than the mean value of the stock price, then it is considered as up movement, otherwise down movement. Using proposed algorithm we have computed each technical indicators values either up and down movements. These ten technical indicators up and down movements data are given input to the DNN model to classify the stock price movements.

4.2.1 Stock Price Classification using DNN

DNN is one of the popular methods for learning lots of different types of patterns. It is composed of more than one hidden layer in a neural network. The neural network is designed based on human brains. Human brains consist of billions of small cells known as neurons.

The proposed work, DNN, is considered to classify stock price's up and down movements, as shown in Fig 4.6. It is five layers of architecture with three hidden layers, and each layer transfers the data between the neurons. It takes the input are ten technical indicator values, and it is denoted by x_i. Each of these technical indicators value is multiplied by weight W_i, and it is denoted as $W_i x_i$. Initially, some random values were assigned. Each of these weights assigned to input value and weighted sum performed, and it is defined in Equation 4.5. The bias is typical for each neuron. There is one bias per neuron.

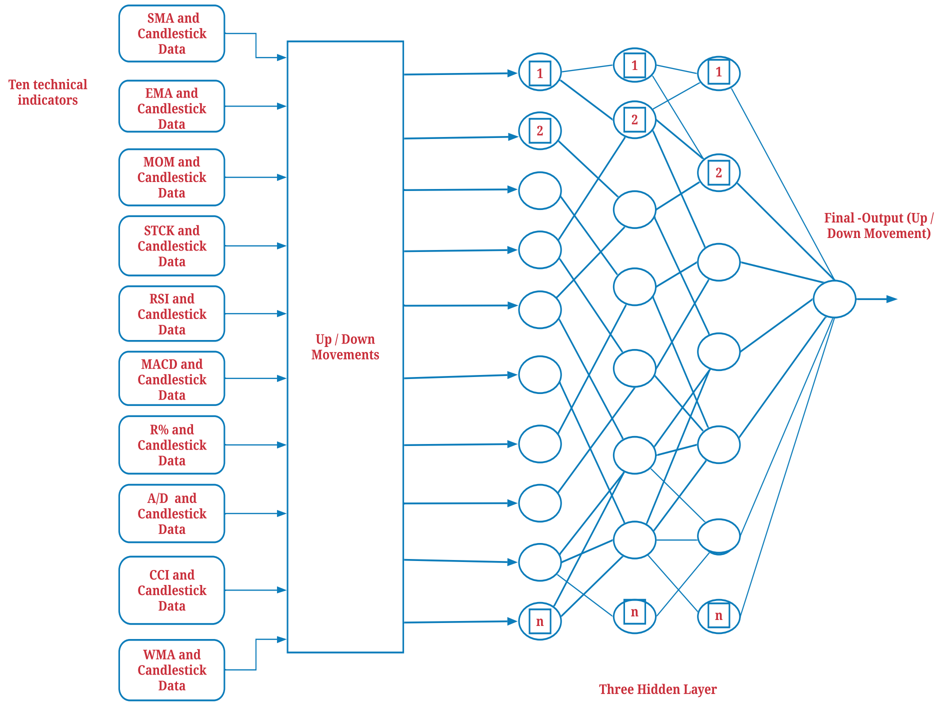


Figure 4.6 Framework of proposed five layer Deep Neural Network.

An activation function takes the weighted sum of input and bias. The Relu activation function is considered in the experiment. Relu is the most popular function used to deal with nonlinear data. It gives an output of x if x is positive and 0 otherwise and it represented as $f(x) = max(0,x)$. There is no upper limit in Relu. In DNN, we considered the backpropagation method to improve the performance of the network. The cost function helps to reduce the error rate in the model. The cost value is the difference between the predicted output and the actual output. The least-cost value was obtained by making adjustments to the weights and biases iteratively throughout the training process. The stochastic gradient descent (SGD) mechanism was considered to change the weight and bias to minimize the cost functions.

$$\alpha = \left(\sum_{i=1}^{n} W_i x_i + B \right) \tag{4.5}$$

4.3 Feature Selection based Stock Price Prediction model

The overall Framework of the proposed model is described in Fig 4.7. In this work, we have considered the ANN regression method for stock price prediction. The data are retrieved from NSE. We have considered 33 different combinations of technical indicators and evaluated them based on formulas which are described in Table 4.1. Each technical indicator's mean and standard deviation are described in Table 4.2 and 4.3. The mean and standard deviation are describes the variability in the datasets.

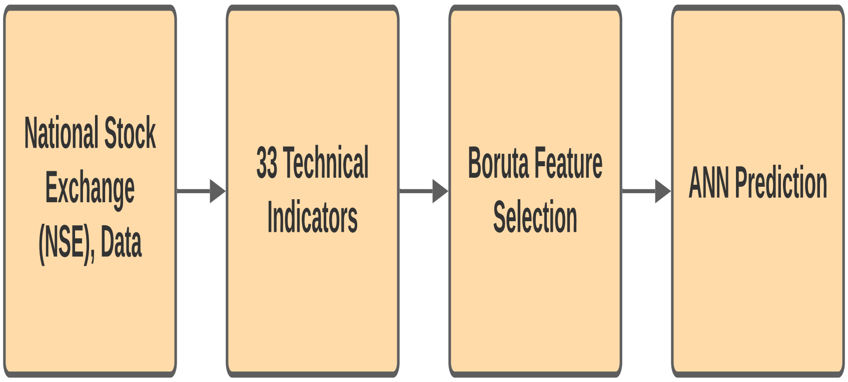


Figure 4.7 Overall Framework of Proposed Prediction model

We have used the Boruta feature selection method to select the feature of technical indicators, and it is described in section 4.3.

4.3.1 ANN Regression model for Stock Price Prediction

The ANN can provide a more accurate prediction model for the larger amount of data, nonlinear data, non-stationary and has been a popular approach for stock market pre-

Table 4.2 Stock ICICI Bank.

Technical indicator feature	Mean	Standard Deviation
SMA 5 Days	216.549	62.827
SMA 10 Days	216.261	62.669
SMA 14 Days	216.031	62.55
SMA 30 Days	214.945	62.784
SMA 50 Days	213.291	64.339
SMA 100 Days	209.286	67.928
SMA 200 Days	201.18	75.655
EMA 5 Days	216.624	62.848
EMA 10 Days	216.433	62.725
EMA 14 Days	216.28	62.632
EMA 30 Days	215.344	63.25
EMA 50 Days	214.008	64.736
EMA 100 Days	210.856	68.175
EMA 200 Days	203.719	76.419
MOM 5 Days	100.448	5.387
MOM 10 Days	100.874	7.31
MOM 14 Days	101.235	8.754
STCK 14 Days	52.285	20.702
STCD 14 Days	52.276	27.17
MACD(26,13,19) Days	51.653	11.499
MACDSIGNAL	51.19	9.063
MACD(45,25,15)Days	0.687	4.144
MACDSIGNAL	0.673	3.854
RSI 14 Days	1.086	4.427
RSI 28 Days	1.048	4.126
R 14 Days	-47.145	30.338
R 28 Days	-45.845	29.86
R 50 Days	-43.801	29.845
R 100 Days	-40.368	28.87
A/D 14 Days	62.82	43.84
CCI 14 Days	5.333	109.066
CCI 50 Days	16.038	111.717
CCI 100 Days	19.023	108.486

Table 4.3 Stock SBI Bank.

Technical indicator feature	Mean	Standard Deviation
SMA 5 Days	233.104	48.851
SMA 10 Days	232.942	48.712
SMA 14 Days	232.812	48.617
SMA 30 Days	231.995	49.351
SMA 50 Days	230.397	52.283
SMA 100 Days	226.354	58.626
SMA 200 Days	217.983	70.292
EMA 5 Days	233.147	48.859
EMA 10 Days	233.039	48.732
EMA 14 Days	232.952	48.641
EMA 30 Days	232.092	50.087
EMA 50 Days	230.708	52.959
EMA 100 Days	227.321	59.213
EMA 200 Days	219.478	71.34
MOM 5 Days	100.299	5.253
MOM 10 Days	100.587	7.287
MOM 14 Days	100.839	8.75
STCK 14 Days	50.03	22.151
STCD 14 Days	50.012	28.2
MACD(26,13,19) Days	50.537	13.078l
MACDSIGNAL	50.331	10.165
MACD(45,25,15) Days	0.424	5.167
MACDSIGNAL	0.412	4.877
RSI 14 Days	0.714	5.764
RSI 28 Days	0.699	5.447
R 14 Days	-50.128	31.247
R 28 Days	-50.758	31.386
R 50 Days	-49.651	31.272
R 100 Days	-47.956	30.974
A/D 14 Days	3.616	111.983
CCI 14 Days	6.386	111.802
CCI 50 Days	7.235	113.422
CCI 100 Days	-92.427	170.555

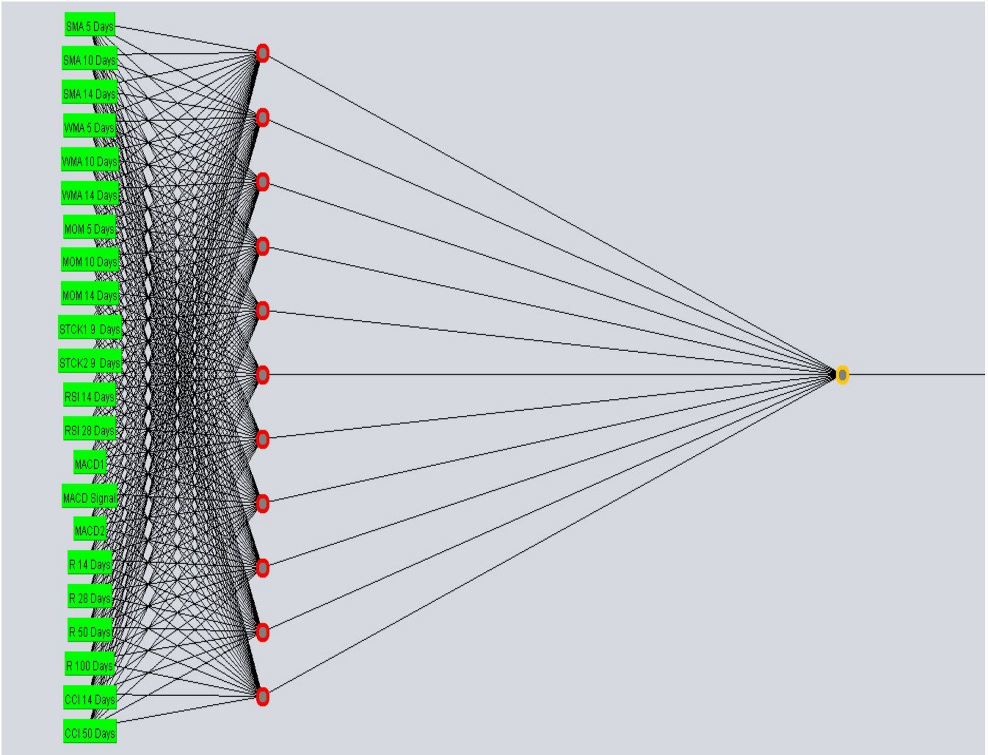


Figure 4.8 The Proposed ANN Regression Prediction model

diction. Patel et al. (2015b) investigated stock price prediction using the combination of SVR-ANN, ANN-RF, and SVR-RF method. However, they have considered only ten technical indicators. Therefor in this work, we have considered 33 technical indicators to predict the stock price. To identify the relevant stock technical indicators we have considered Boruta feature selection method. This is the first approach for stock price prediction based on the Boruta method to the best of our knowledge. However Boruta method (Kursa et al., 2010) is not considered for technical indicators selection. Therefore, we have used the Boruta feature selection method to identify the relevant stock technical indicators in this work.

We have considered Boruta feature selection method to select the relevant stock technical indicators for stock price prediction. Boruta feature selection algorithm are discussed in section 4.1. The relevant technical indicators are given as input to the ANN regression model , and it is described in Figure 4.8. In this work, the ANN is used to predict the stock prices. ANN has three layers, and each layer is connected to the other. The neurons represent the technical indicators. The back-propagation method was considered to train the model. The sigmoid function activation function is used in the ANN regression model. A gradient descent momentum parameter is considered to determine the weights and to reduce the global minimum.

4.4 Results and Analysis

4.4.1 Feature Selection based Stock Price Movements Classification Model

The ICICI Bank, SBI Bank, Yes Bank and Kotak Bank stock are considered for experiments. These stocks are trading with significant volumes in the NSE market. Using the BFS method, we have obtained 23 best features in ICICI Bank stock, 22 best features in SBI Bank stock, 24 best features in Yes Bank stock and 23 best features in Kotak Bank stock. The obtained results using BFS are described in Table 4.4 and 4.5.

For each stock on the closing basis, we have computed a stock class tag up and down by comparing the stock's current price and its previous price. We have used the Accuracy and F-Measure metric to evaluate the performance of the DNN, SVM, RF and ANN model. The Accuracy and F-measure are given in Equation 4.6 and Equation 4.7. NSE datasets consist of 2400 rows. We have used tenfold cross-validation in the experiment. The Experiment is carried out in the R Studio platform.

$$Accuracy = \frac{TruePositive + TureNegative}{TruePostive + TureNegative + FalsePositive + FalseNegative} \tag{4.6}$$

$$F-Measure = 2 \times \frac{precision \times recall}{precision + recall} \tag{4.7}$$

The performance of the proposed classification model is better than existing work, and it is described in Table 4.6. We have fine-tuned the parameters to get the optimized results.

Stock market predictions are a difficult task for stock fund managers and financial analysts due to unstable stock data, which is noisy and nonlinear. The study focused on stock price movement classification on a daily basis. We conclude that boruta feature selection is a useful method for the identification of relevant technical indicators. The

Table 4.4 Stock ICICI Bank after feature selection.

Technical indicator feature	meanImp	medianImp	minImp	maxImp	Z-Score	Decision
SMA 5 days	5.905820	5.805600	3.5773017	8.330396	1.00000000	Confirmed
SMA 10 days	4.412296	4.450179	2.5664467	6.381897	0.94949495	Confirmed
SMA 14 days	3.904888	3.865653	1.7642646	5.652071	0.86868687	Confirmed
WMA 5 days	5.108720	5.116738	1.9129550	7.403865	0.94949495	Confirmed
WMA 10 days	4.715484	4.789195	2.7125061	6.435136	0.95959596	Confirmed
WMA 14 days	4.357254	4.259337	1.9938488	6.654934	0.91919192	Confirmed
MOM 5 days	18.692034	18.659231	16.9670413	20.344669	1.00000000	Confirmed
MOM 10 days	10.406941	10.374657	9.1582195	11.926020	1.00000000	Confirmed
MOM 14 days	9.074784	9.058466	7.0379555	10.879017	1.00000000	Confirmed
STCK1 9 days	21.269620	21.339074	18.4123667	23.683564	1.00000000	Confirmed
STCK2 9 days	25.729514	25.897507	22.7613795	28.723707	1.00000000	Confirmed
RSI 14 days	15.487287	15.438833	13.8918381	16.732855	1.00000000	Confirmed
RSI 28 days	8.608997	8.602433	7.0550301	10.168651	1.00000000	Confirmed
MACD1 (26,13,9) days	16.152372	16.184757	14.2102377	18.254305	1.00000000	Confirmed
MACD SIGNAL	9.164158	9.182269	7.2546778	10.961601	1.00000000	Confirmed
MACD (15,25,15)	6.046035	5.977311	3.7503675	8.219024	1.00000000	Confirmed
R 14 days	41.157543	41.103670	38.5168811	44.780878	1.00000000	Confirmed
R 28 days	24.057435	24.017315	22.3425494	26.501515	1.00000000	Confirmed
R 50 days	19.071730	19.049960	17.2172337	21.637499	1.00000000	Confirmed
R 100 days	12.560288	12.563706	10.8720538	14.235407	1.00000000	Confirmed
CCI 14 days	17.287549	17.246896	15.4191912	19.279900	1.00000000	Confirmed
CCI 50 days	9.203503	9.272982	7.1159672	11.020916	1.00000000	Confirmed
CCI 100 days	6.259018	6.366987	3.7899522	8.369810	0.98989899	Confirmed

Table 4.5 Stock SBI Bank after feature selection.

Technical indicator feature	meanImp	medianImp	minImp	maxImp	Z-Score	Decision
SMA 5 days	5.4606723	5.5251714	3.1273954	7.653046	1.00000000	Confirmed
SMA 10 days	3.6382503	3.6526033	0.8677842	6.344588	0.81818182	Confirmed
WMA 5 days	4.8639987	4.9798041	2.1810807	7.662506	0.96969697	Confirmed
WMA 10 days	4.2772306	4.2978051	2.0651554	8.159167	0.91919192	Confirmed
WMA 14 days	3.7222162	3.9236859	0.8468174	6.263128	0.87878788	Confirmed
MOM 5 days	19.2684238	19.1960889	17.7018349	21.289981	1.00000000	Confirmed
MOM 10 days	8.5963907	8.6058484	6.7912856	9.896962	1.00000000	Confirmed
MOM 14 days	9.1535623	9.1827829	7.7671871	10.944682	1.00000000	Confirmed
STCK1 9 days	22.8154629	22.8720799	20.0837955	25.518795	1.00000000	Confirmed
STCK2 9 days	21.6583364	21.6522237	19.0834711	23.759105	1.00000000	Confirmed
RSI 14 days	15.1007526	15.0648255	13.6517412	17.075092	1.00000000	Confirmed
RSI 28 days	9.0990375	9.0610509	6.6856226	10.547252	1.00000000	Confirmed
MACD1 (26,13,9) days	14.6987380	14.6169198	12.7518900	16.634019	1.00000000	Confirmed
MACD SIGNAL	8.3677429	8.4367255	6.3049964	10.389434	1.00000000	Confirmed
MACD (15,25,15)	6.1898880	6.2728086	4.1248029	7.845896	1.00000000	Confirmed
R 14 days	37.9879805	37.9176749	34.9232519	40.963526	1.00000000	Confirmed
R 28 days	25.7820969	25.7832651	23.5718441	27.630123	1.00000000	Confirmed
R 50 days	18.4300230	18.4008101	16.4257141	20.670375	1.00000000	Confirmed
R 100 days	14.0062758	13.9214321	11.6389235	16.356005	1.00000000	Confirmed
A/D	17.2844799	17.2619585	15.0965778	19.177774	1.00000000	Confirmed
CCI 14 days	7.1505888	7.1446862	5.2038171	9.449850	1.00000000	Confirmed
CCI 50 days	8.7992939	8.8263404	6.7923020	10.430373	1.00000000	Confirmed

Table 4.6 Result Comparison-Stock Price Movements Classification

Stock	Existing Work (Patel et al., 2015a)					
	ANN		SVM		RF	
	Accuracy	F-Measure	Accuracy	F-Measure	Accuracy	F-Measure
ICICI Bank	73.12%	0.7470	68.55%	0.6935	77.12%	0.7877
SBI Bank	74.12%	0.7248	70.35%	0.7080	78.85%	0.7987
Yes Bank	72.12%	0.7414	71.35%	0.7130	77.15%	0.7638
Kotak Bank	73.12%	0.7532	72.35%	0.7210	76.35%	0.7637

Stock	**Proposed Feature Selection based Stock Price Movements Classification Model**					
	ANN		SVM		**DNN**	
	Accuracy	F-Measure	Accuracy	F-Measure	**Accuracy**	**F-Measure**
ICICI Bank	79.42%	0.796	76.61%	0.769	**83.10%**	**0.815**
SBI Bank	79.60%	0.793	77.01%	0.776	**84.50%**	**0.824**
Yes Bank	78.32%	0.781	76.63%	0.753	**83.67%**	**0.833**
Kotak Bank	78.60%	0.733	77.51%	0.786	**83.90%**	**0.844**

study also demonstrated that DNN model performance is better than machine learning techniques. The contribution of this study can be summarized as follows. First is the technical indicators selection and identification of the relevant technical indicators by using the Boruta feature selection technique. The second is we have improvised the results of stock price movements classification using the DNN method.

4.4.2 Stock Price Classification based on Candlestick data and Technical Indicators

In this work, stock price data collected from NSE, India, and website URL is (`http://www.nseindia.com/stock`). The stocks considered herein are Infosys, Reliance, Hdfc, and Hdfc bank stocks. The datasets range from the year 2008 to 2021. We have studied two commonly used metrics referenced in the literature, namely, Accuracy and F-Measures, and they are described in Equations 4.6 and 4.7. These metrics are considered to evaluate the performance of the DNN model. We use ten-fold cross-validation in the experiment.

Table 4.7 shows that the highest accuracies for Reliance stock, Infosys, Hdfc Bank, and Hdfc are 0.8361, 0.8178, 0.8211, and 0.8349, respectively. The performance of the proposed DNN model is better than existing work and is described in Table 4.7. The reason for the better performance is the data preparation uses a combination of candlestick data and technical indicators. In addition, we have built a model on a five-layer deep neural network; the existing model was built on a three-layer artificial neural network.

In this study, we make predictions of the stock price's up and down movements in the NSE. The first contribution of this study is the identification of a trend in data by using a combination of candlestick data and technical indicators. Existing research has not considered these combinations. The second contribution of this study is using DNN to classify and accurately predict a stock price is up and down movement. The proposed deep neural network prediction model outperforms the ANN model. This study focuses on Intraday predictions.

Table 4.7 Results Comparisons-Stock Price Classification based on Candlestick Data.

Stock Names	Prediction Models	Accuracy	F-Measure
Reliance	ANN (Kara et al., 2011)	0.7275	0.7392
Reliance	**Proposed Model**	**0.8361**	**0.8646**
Infosys	ANN (Kara et al., 2011)	0.7130	0.7364
Infosys	**Proposed Model**	**0.8178**	**0.8327**
Hdfc Bank	ANN (Kara et al., 2011)	0.7465	0.7315
Hdfc Bank	**Proposed Model**	**0.8211**	**0.8410**
Hdfc	ANN (Kara et al., 2011)	0.7531	0.7432
Hdfc	**Proposed Mode**	**0.8349**	**0.8558**

4.4.3 Feature Selection based Stock Price Prediction model

There are two phases, the first phase of the experiment considered feature selection technique. The second phase of the experiment is the prediction model. In the first phase of the experiment, we have used the Boruta method to select the best feature based on the Z Score. ICICI Bank and State bank of India stocks after feature selection are described in Table 4.4 and Table 4.5. These selected features of the technical stock indicator are given as input to the ANN regression model. The MAE and RMSE are used to evaluate the performance of the prediction model, and it is described in Equation 4.8 and Equation 4.9. The proposed ANN Regression prediction model performance is compared with existing work, and it is shown in Table 4.8. All tests are trained with a ten-fold cross-validation-based model. From this study, we can conclude that technical indicators are important to predict stock prices.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{4.8}$$

Table 4.8 Result Comparison with Existing work.

Stock Name	Prediction Model	MAE	RMSE
ICICI Bank	ANN (Patel et al., 2015a)	27.0583	36.0444
ICICI Bank	**Proposed Feature Selection based ANN Model**	**15.1221**	**19.9444**
State Bank of India	ANN (Patel et al., 2015a)	27.7392	36.4834
State Bank of India	**Proposed Feature Selection based ANN Model**	**17.4341**	**23.1585**

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \tag{4.9}$$

The study focused on stock prediction for short-term trading. The stock data is collected from the National Stock Exchange (NSE). This study has two objectives, first is the selection and identification of the technical indicators of the relevant technical indicators by using Boruta feature selection techniques. The second objective is an accurate stock price prediction model. The performance of the model was measured using RMSE and MAE metrics. The ICICI Bank MAE is 15.12, and State Bank of India MAE is 14.4. The experimental work shows that the ANN prediction model outperforms existing work by decreasing the error rate in the prediction. The future work can be identified the microeconomics, macroeconomics factor, and fundamental analysis to find the quality of stocks.

4.5 Summary

Stock price movements classification and prediction are challenging for traders and stock analysts. Predicting and classify of stock prices on a daily basis is a difficult task due to more ups and downs in the financial market. Therefore, there is a need for a more robust predictive model to classify and predict stock prices. Most of the existing work is based on machine learning techniques and considered very few technical indicators to predict and classify the stock price movements. This chapter addresses the three tasks (a)Feature selection and stock price classification (b)Stock price classification using a combination of candlestick data and technical indicators(c) Stock price prediction.

The first is the selection and identification of the technical indicators of the relevant technical indicators by using the Boruta feature selection technique. To classify stock price movements, we have considered the DNN model. The performance of the DNN model is better than the ANN, RF and SVM model. The experimental results significantly improve the classification accuracy rate by 5% to 6%.

The second is stock price movements classification using a combination of candlestick data and technical indicators. We proposed the method to identify the trend in data using the combination of candlestick data and technical indicator values. The outcome of this method is given as inputs to a deep neural network (DNN) to classify a stock is up and down movements. National Stock Exchange (NSE), Indian datasets are considered for an experiment from the years 2008 to 2021. Experimental results compared with Artificial Neural Network (ANN) model. The proposed DNN model outperforms state-of-the-art methods by 8-11% in predicting up and down movements of a given stock.

The third is an accurate prediction of the stock price using the ANN method. To predict stock prices, we have used ANN (Artificial Neural Network) regression prediction model, and model performance is evaluated using metrics is Mean absolute error (MAE), and Root means square error (RMSE). The experimental results are better than the existing method by decreasing the error rate in the prediction to 12%.

Chapter 5

Stock Price Volatility Estimation using Regime Switching Technique

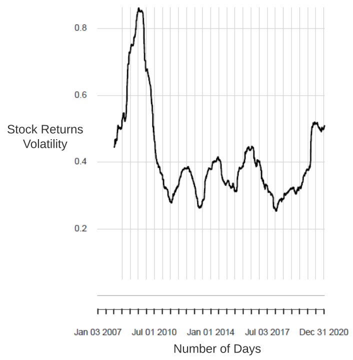
In this chapter, we have considered regime-switching based on the Markov switching GARCH (MSGARCH) and Self-Exciting Threshold Autoregressive (SETAR) model instead of the plain GARCH model. This is the first empirical study on the Indian stock market data based on the regime-switching model to the best of our knowledge.

5.1 Volatility

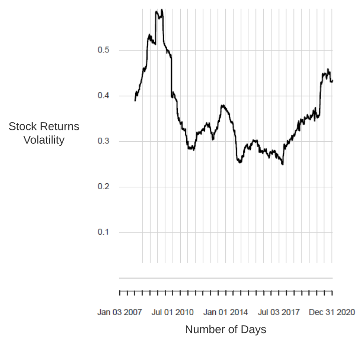
Volatility in finance is a statistic to measure the rate of change in the stock price over time, and it is calculated using standard deviations. The volatility statistics help the investor to estimates the risk in the stock or stock index. When the volatility is very high, then it is riskier to invest. So identification of volatility in the market is essential as far stock market is concerned. This chapter has considered MSGARCH and SETAR models to estimate the volatility in stock and stock indices.

5.2 Methodology

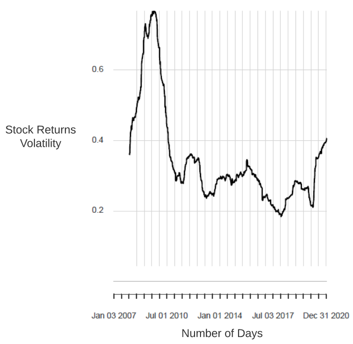
We have considered a few NIFTY 50 stocks like Tata Steel, Bharti Airtel, HCL Tech, NTPC, and stock indices like NIFTY 50, Bank Nifty, Nifty 100, and Nifty 500. The datasets were considered from January 2007 to April 2021 (NSE, 2021). It includes the bull and bear market data. There are 3259 rows in the datasets. We have considered the closing price of the stock to estimates the volatility. The annualized stock indices and stock price volatility is estimated using the standard deviation as depicted in Figure 5.1 and 5.2. The stock returns are computed using the stock closing price. Let Y_t is the closing price of stock at time t, and it is defined in Equation 5.1. Stock price returns are depicted in Figure 5.3,5.4,5.5 and 5.6.



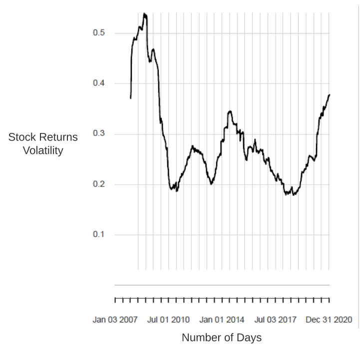
(a) Tata Steel stock



(b) Bharti Airtel stock



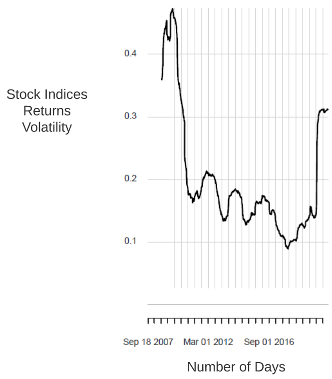
(c) HCL TECH stock



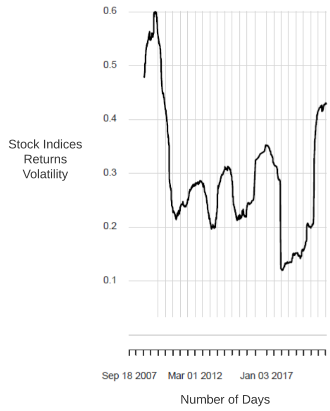
(d) NTPC stock

Figure 5.1 Annualized Volatility of stock

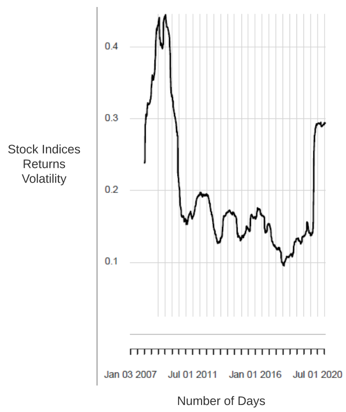
$$StockReturn = [\frac{Y_t - Y_{t-1}}{Y_{t-1}}] \tag{5.1}$$



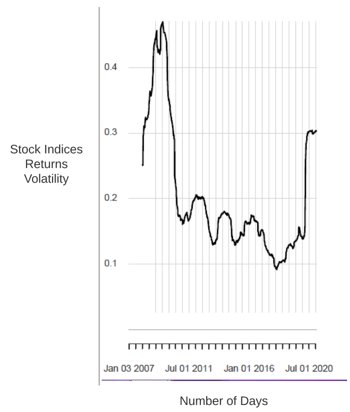
(a) Nifty 50 Index



(b) Bank Stock Index



(c) Nifty 500 Index



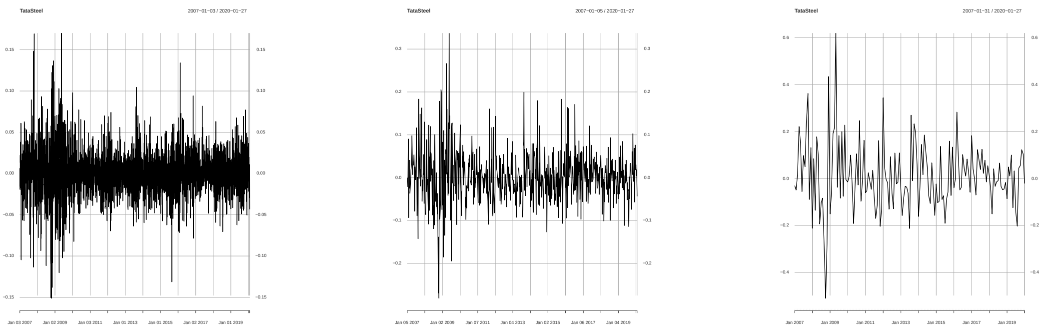
(d) Nifty 100 Index

Figure 5.2 Annualized Volatility of Indices

5.2.1 Structural Break

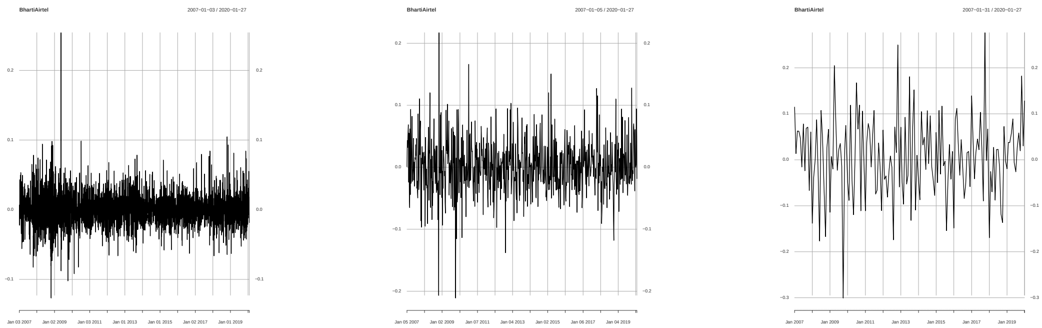
Structural breaks in the data leads to errors in forecasting, in this work we have considered the structural breaks in the data while forecasting the volatility. Chow test is used to test the structural break in the stock datasets. In this method, time-series datasets are split into two equal parts, then the coefficient of two linear fits are compared to know whether the structural break is present or not. Let us consider the linear regression equations in 5.2 and 5.3 .

$$Y = C + \beta X + \varepsilon \tag{5.2}$$



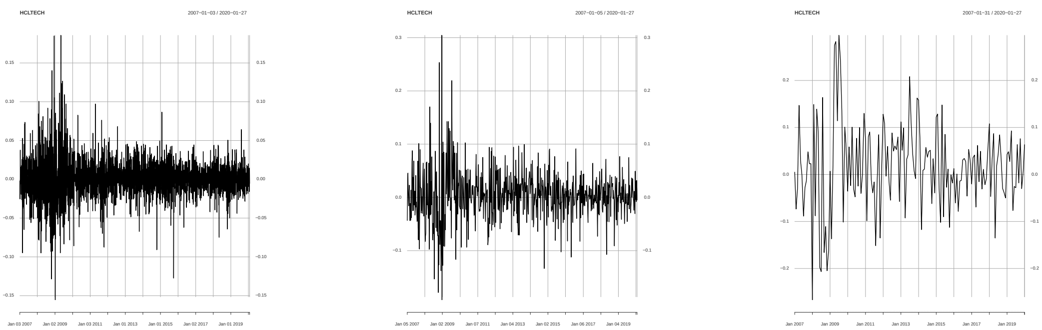
(a) Daily Returns (b) Weekly Returns (c) Monthly Returns

Figure 5.3 Tata Steel stock returns



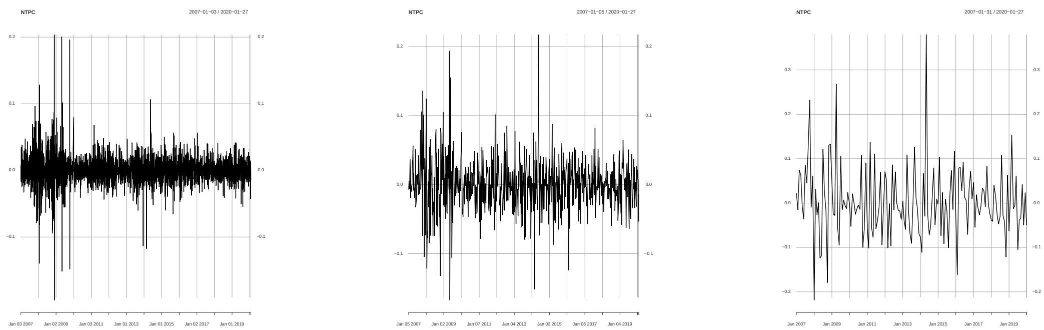
(a) Daily Returns (b) Weekly Returns (c) Monthly Returns

Figure 5.4 Bharti Airtel stock returns



(a) Daily Returns (b) Weekly Returns (c) Monthly Returns

Figure 5.5 Hcl Tech Bank stock returns



(a) Daily Returns (b) Weekly Returns (c) Monthly Returns

Figure 5.6 NTPC stock returns

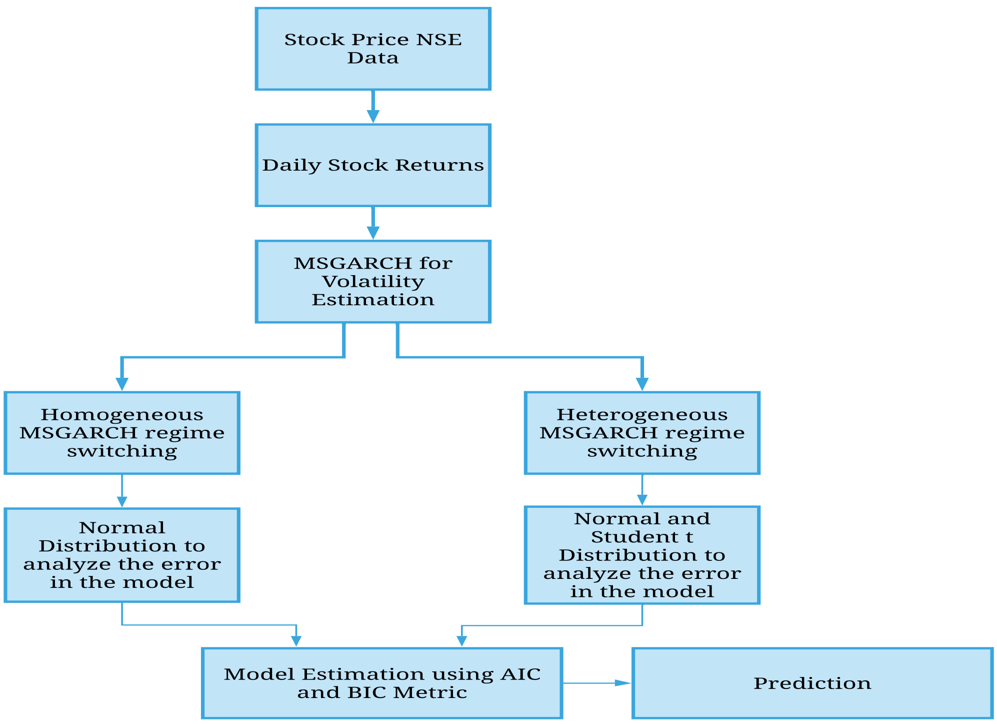


Figure 5.7 Overall proposed work.

$$Y1 = C_1 + \beta_1 X_1 + \varepsilon_1 \tag{5.3}$$

Here Y, $Y1$ are the dependent variables, X,$X1$ are independent variables. C,$C1$ are the constants and β,β_1 are slope of the line. ε and ε_1 are the error term in regression model. We have used 5.2 and 5.3 equations to fit the datasets. To check two linear re-

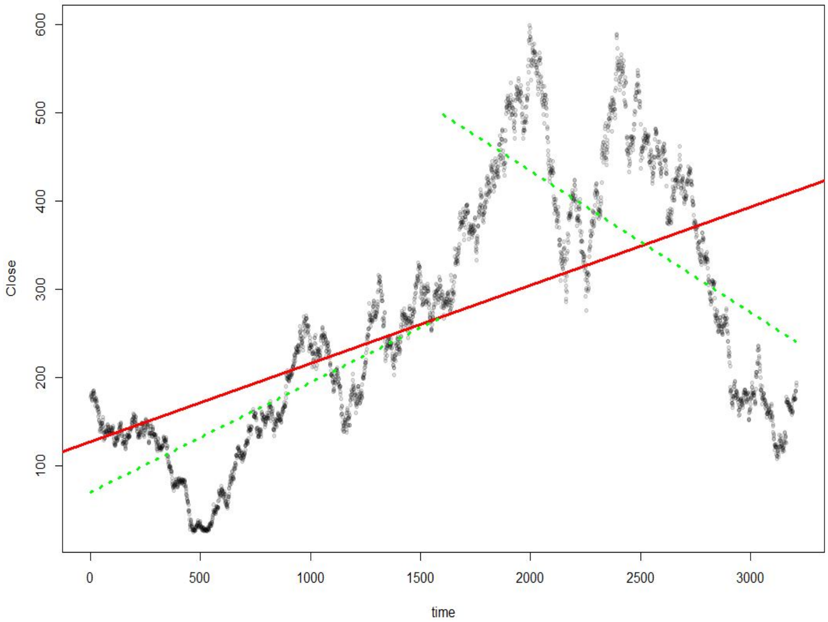


Figure 5.8 Chow Test on Tata Steel Stock

gression fits are similar or not, we have defined null hypothesis and alternate hypothesis in equation 5.4 and 5.5.

$$H_0 : \beta = \beta_1 ; C = C1 \tag{5.4}$$

$$H_1 : \beta \neq \beta_1 ; C \neq C1 \tag{5.5}$$

The p-value is greater than 0.05, indicating a structural break in the stock datasets. The experimental results show that there is a structural break in the datasets, and it is described in Figure 5.8 and Table 5.1.

Table 5.1 F Score and p value

Stock	F Score	p value
Tata Steel	2.9238	0.05386
Bharthi Airtel	0.14442	0.8655
HCL Tech	12.335	4.5587
NTPC	30.457	9.459

5.2.2 Non Regime Switching

Stock price data are dynamic. Autoregressive–Moving-Average(ARMA) model is useful for linear time series data (Chon and Cohen, 1997),(Noda et al., 1996),(Aarstol, 1999). Hence ARMA models are not able to capture the dynamic volatility in time series data. Therefore most of the work considered the GARCH model for dynamic volatility estimation (Wang and Wang, 2021),(Ioannidis et al., 2021), (Xing et al., 2021), (Escobar-Anel et al., 2021), (Huang et al., 2021). The GARCH model is useful when the variance of time series data is not constant. The GARCH model is defined in equation 5.6 and 5.7. ε_t is a random variable with zero mean and unit variance.

$$Y_t = \sqrt{H_t}\varepsilon_t \tag{5.6}$$

$$H_t = \omega + \beta_0 H_{t-1} + \alpha_0 \varepsilon_{t-1}^2 \tag{5.7}$$

$Y_t \leftarrow$ Time varying volatility.
$\sqrt{H_t}\varepsilon_t \leftarrow$ Unexpected return.
$H_t \leftarrow$ Predictable variance changing over time.
$\omega \leftarrow Constant.$

$\beta \leftarrow Constant.$
$\alpha \leftarrow Constant.$

It is found from the literature that the GARCH models do not capture the variations of the volatility periods. Most GARCH models considered in literature are one term Autoregressive Conditional Heteroscedasticity (ARCH) and one GARCH, i.e.,GARCH(1.1). However GARCH model ignores the regime-switching in volatility estimation. Because in the equation 5.7 GARCH model alpha and beta parameters can be more than one when the structural break is the present. Therefore, we have considered regime-switching using the MSGARCH and SETAR models.

5.2.3 Regime Switching based on MSGARCH

Structural changes in time series data are referred to as regime-switching. The overall proposed work is depicted in Figure 5.7. Regime switching is essential when there is higher volatility in stock prices. Therefore, we have applied a Markov Switching-based GARCH(MSGARCH) model to estimate the stock price volatility. In this work, we have considered two MSGARCH models to estimates the volatility in stock price. In the first model, we have used homogeneous MSGARCH regime-switching. In homogeneous MSGARCH, GARCH conditional variance is considered. Normal distribution is used to analyze the error distribution in the models.

In the second model, we have used heterogeneous MSGARCH regime-switching. In heterogeneous MSGARCH, GARCH, EGARCH, TGARCH, conditional variances are considered. Normal and Student t are used to analyze the error distribution in the models.

In the proposed work, stock return data are given as input for the MSGARCH model for estimating the volatility. The calculation of stock returns is discussed in Section 3. In MSGARCH model stock returns are defined as Y_t at time t. We assumed that Y_t has zero mean and it is not serially correlated. The MSGRCH is defined in equation 5.8.

$$Y_t \mid (S_t = r, P_{t-1}) \sim D(0, H_{r,t,\varepsilon r}) \tag{5.8}$$

where $D(0, H_{,t,\varepsilon r}) \leftarrow$ Continuous distribution

$H_{r,t} \leftarrow$ Continuous Variance

$r \leftarrow$ regime k

$\varepsilon r \leftarrow$ vector k

$P_{t-1} \leftarrow$ Stock return information set upto $t-1$

$D(0, H_{r,t,\varepsilon r})$ denotes continuous distribution. It has mean value is zero. conditional variance is denoted by $H_{r,t}$ with regime r state. εr vector represents the regime-switching of markov process. The regime switching from one state to another state is evolved using the first order of homogeneous markov chain with state S_t. Here S_t is an integer value which has discrete state $\{1..r\}$. In Equation 5.8, $S_t = r$ represents the current state and P_{t-1} is the previous state of markov process.

We have considered the GARCH, EGARCH, and TGARCH conditional variance in the MSGARCH model to estimate stock price volatility.

Algorithm 6: Algorithm

1: **Input :** Input stock prices and stock indices.
2: **Output :** Ten days forecast prediction .
3: Stock price and stock indices returns are computed and given as input to the MSGARCH models.
4: MSGARCH specification for volatility estimation $Y_t \mid (S_t = r, P_{t-1}) \sim D(0, H_{r,t,\varepsilon r})$.
5: a). GARCH conditional variance with Homogeneous regime switching method.
b). GARCH conditional variance with Heterogeneous regime switching method.
6: Estimate conditional distribution.
a) Normal Distribution $fN(\eta) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\eta^2}, \eta \varepsilon \mathbb{R}$
b)Student T $fN(\eta;\nu) \equiv \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu-2\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\eta^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \eta \varepsilon \mathbb{R}$
7: Model estimation using AIC and BIC metric.
8: Ten days forecast prediction.

Student t and Normal distributions are used to analyze the error distribution in the models. The performance of the model is estimated by using the AIC and BIC metric. The details of the steps are described in Algorithm6.

5.2.4 Regime Switching based on Self-Exciting Threshold AutoRegressive(SETAR) model

The SETAR model is one of the popular models in time series to forecast the future trend in data. SETAR model was used when there is a structural break in the datasets. In SETAR(R, AR) model consists of two parts. R represents the number of the regime, and AR represents the order of auto-regression.

Consider a simply auto-regression(P) for stock price Y_t, and it is defined below equation.

$$Y_t = \mu + \phi_1 Y_t -_1 + \phi_2 Y_t -_2 + ... + \phi_P Y_t -_P + \sigma\varepsilon_t \tag{5.9}$$

where $\phi_1 (i = 1, 2, ..P) \leftarrow$ auto-regression(P) Coefficient.
$\varepsilon_t \sim iidN(0, \sigma^2) \leftarrow$ Constant Variance.

TAR allows the model parameters to change according to the value of a weakly exogenous threshold variable z_t for capturing nonlinear trends.

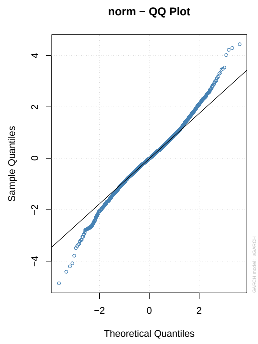
$$Y_t = X_t \phi^{(j)} + \sigma^{(j)} \varepsilon_t if r_j -_1 < z_t < r_j \tag{5.10}$$

where $X_t = (1, Y_t, Y_{t-}1, ...Y_{t-}P) \leftarrow$ Column vector variables
$(r_1, r_2, ..r_k) \leftarrow$ divide the domain of the threshold variable z_t into k different regimes.

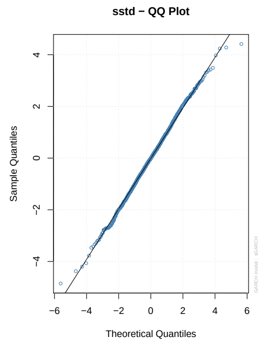
In each different regimes in stock price Y_t follows a different auto-regression(P) model. When the threshold variable $z_t = Y_t -_d$ with the delay parameter d being a positive integer, the dynamics or regime of Y_t is determined by its own lagged value $Y_t -_d$ and the TAR model is called a self-exciting TAR or SETAR model.

5.3 Experiment Results

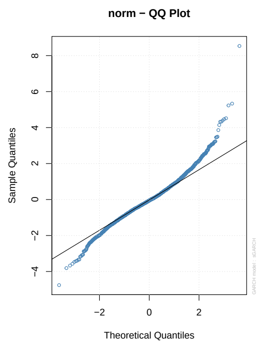
The experiments are carried out in R studio. We have used the MSGARCH R package developed by (Ardia et al., 2016). There were around 3259 data samples collected



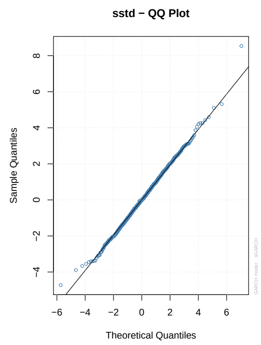
(a) Tata Steel Stock Returns



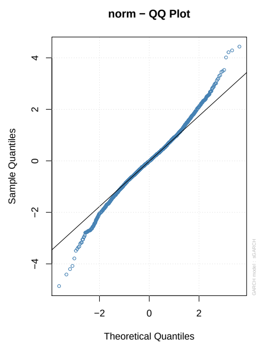
(b) Tata Steel Stock Returns



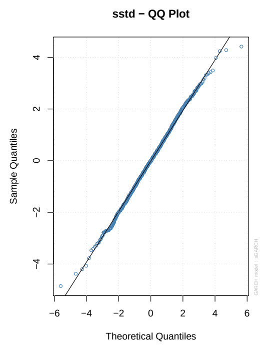
(c) Bharti Airtel Stock Returns



(d) Bharti Airtel Stock Returns



(e) Hcl Tech Index Returns



(f) Hcl Tech Stock Returns

Figure 5.9 Normal Distribution and Student T Distribution fit

from the year January 2007 to April 2021. The data are given as input to the proposed model. We have considered two MSGARCH models: homogeneous MSGARCH and the second is heterogeneous MSGARCH. The proposed model conditional distributions are verified by using the normal distribution and Student t distribution. We have found that Student t distribution performs better than the normal distribution, depicted

Table 5.2 Non Regime Switching Results

Stock	Garch-AIC	Garch-AIC	Garch-GJR-AIC	Garch-GJR-BIC
Tata Steel	-4.5835	-4.5728	-4.596	-4.5835
Bharti Airtel	-4.8611	-4.8504	-4.8633	-4.8508
HCL Tech	-5.0173	-5.0066	-5.0194	-5.007
NTPC	-5.3148	-5.3041	-5.3142	-5.3018
Nifty50 Index	-6.177	-6.1658	-6.2035	-6.1904
Bank Nifty Index	-5.4354	-5.4233	- 5.455	-5.441
Sensex Index	-6.1583	-6.1508	-6.2307	-6.2176
Nifty500 Index	-6.2202	-6.2095	-6.2419	-6.2294
Nifty100 Index	-6.1737	-6.163	-6.1975	-6.185

in Figure 5.9. Therefore further work, we have considered Student t distribution instead of normal distribution. The performance of each model is estimated using the Akaike Informations Criteria (AIC) and Bayesian Information Criteria(BIC) metric and it is defined below equations 5.11 and 5.12. L represents the value of the likelihood function. The number of estimated parameters of the model is represented using P.

$$AIC = -2 * Ln(L) + 2 * P \tag{5.11}$$

$$BIC = Ln(n) * P - 2 * Ln(L) \tag{5.12}$$

The proposed work results are described in Table 5.2 and 5.3. The results compared with the GARCH and MSGARCH models. The heterogeneous MSGARCH model, AIC and BIC value is lowest compare to GARCH and SETAR model. The MSGARCH

Table 5.3 Regime Switching Results

Stock	Homogeneous-MSGARCH-AIC	Homogeneous-MSGARCH-BIC	Heterogeneous-MSGARCH-AIC	Heterogeneous-MSGARCH-BIC	SETAR-AIC	SETAR-BIC
Tata Steel	-15695.5334	-15646.3551	-14650.8153	-14541.4885	-24785.61	-24742.58
Bharti Airtel	-16754.0173	-16704.8366	-16786.6075	-16675.9511	-26032.09	-25989.06
HCL Tech	-17277.0615	-17227.8809	-17316.1259	-17205.4695	-25920.46	-25877.43
NTPC	-18298.7579	-18249.5773	-18365.7547	-18255.0983	-27370.75	-27327.71
Nifty50 Index	-20052.3353	-20003.6219	-18709.9485	-18601.7441	-27659.02	-27616.4
Bank Nifty Index	-16176.0249	-16128.0034	-16216.5752	-16108.5267	-23338.19	-23296.18
Sensex Index	-21163.0729	-21113.9271	-21177.4857	-21066.9076	-27535.16	-27492.57
Nifty500 Index	-21334.2762	-21285.1095	-21393.1934	-21282.5683	-27525.73	-27483.14
Nifty100 Index	-21198.7449	-21149.5782	-21244.5276	-21133.9025	-29333.57	-29290.55

```
R> pred$draw[, 1:3]
               Sim #1    Sim #2    Sim #3
2021-04-01  0.0071768  0.012383  0.014152
2021-04-02  0.0260905 -0.033743 -0.052634
2021-04-03  0.0114869  0.001914 -0.010834
2021-04-04  0.0070288  0.056103  0.003977
2021-04-05  0.0026817 -0.019628  0.026509
2021-04-06 -0.0196014  0.006993 -0.034076
2021-04-07 -0.0023552 -0.038165 -0.055889
2021-04-08 -0.0248717  0.018686  0.004676
2021-04-09  0.0004089  0.015510  0.016780
2021-04-10  0.0438651 -0.027666  0.022284
```

(a) Tata Steel Stock Returns

```
R> pred$draw[, 1:3]
               Sim #1    Sim #2    Sim #3
2021-04-01  0.0044179  0.004458  0.010804
2021-04-02  0.0162481 -0.020962 -0.041353
2021-04-03  0.0043475  0.001188 -0.008916
2021-04-04  0.0027401  0.021625  0.002515
2021-04-05  0.0017124 -0.006912  0.022498
2021-04-06 -0.0169288  0.004210 -0.029657
2021-04-07 -0.0015407 -0.023250 -0.049989
2021-04-08 -0.0164686  0.011343  0.004343
2021-04-09  0.0002717  0.012632  0.011252
2021-04-10  0.0295067 -0.023173  0.009281
```

(b) Bharthi Airtel Stock Returns

```
R> pred$draw[, 1:3]
               Sim #1    Sim #2    Sim #3
2021-04-01  0.0050726  0.005119  0.010809
2021-04-02  0.0179672 -0.023181 -0.041587
2021-04-03  0.0047972  0.001341 -0.008284
2021-04-04  0.0029175  0.023456  0.003003
2021-04-05  0.0017597 -0.007688  0.011871
2021-04-06 -0.0174142  0.004535 -0.014936
2021-04-07 -0.0015494 -0.024195 -0.023757
2021-04-08 -0.0160296  0.012256  0.003514
2021-04-09  0.0002668  0.012540  0.011062
2021-04-10  0.0280868 -0.023033  0.008924
```

(c) Hcl Tech Stock Returns

```
R> pred$draw[, 1:3]
               Sim #1    Sim #2    Sim #3
2021-04-01  0.0061017  0.006157  0.007037
2021-04-02  0.0217785 -0.028098 -0.026197
2021-04-03  0.0058496  0.001635 -0.005136
2021-04-04  0.0035790  0.028795  0.003175
2021-04-05  0.0021685 -0.009485  0.012688
2021-04-06 -0.0099890  0.005629 -0.016114
2021-04-07 -0.0018360 -0.030179 -0.025842
2021-04-08 -0.0190391  0.015336  0.002044
2021-04-09  0.0003172  0.007817  0.012209
2021-04-10  0.0334213 -0.013900  0.009918
```

(d) NTPC Stock Returns

Figure 5.10 Ten days' forecast prediction

Table 5.4 RMSE and MAPE

Stock	Regime Switching Model	RMSE	MAPE
Tata Steel	SETAR	5.82	110.5
Tata Steel	Heterogeneous-MSGARCH	0.0224	0.0945
Bharti Airtel	SETAR	2.83	111.8
Bharti Airtel	Heterogeneous-MSGARCH	0.0210	0.2209
HCL Tech	SETAR	3.04	128.4
HCL Tech	Heterogeneous-MSGARCH	0.0333	0.0374
NTPC	SETAR	1.30	102.3
NTPC	Heterogeneous-MSGARCH	0.0174	0.0038
Nifty50 Index	SETAR	4.19	131.8
Nifty50 Index	Heterogeneous-MSGARCH	0.0167	0.2545
Bank Nifty Index	SETAR	1.62	127.6
Bank Nifty Index	Heterogeneous-MSGARCH	0.0223	0.3148
Sensex Index	SETAR	4.18	140.4
Sensex Index	Heterogeneous-MSGARCH	0.0167	0.4205
Nifty500 Index	SETAR	3.57	122.90
Nifty500 Index	Heterogeneous-MSGARCH	0.0190	2.8114
Nifty100 Index	SETAR	4.06	115.50
Nifty100 Index	Heterogeneous-MSGARCH	0.0202	0.4720

and SETAR models' forecasting performance is calculated using the Root mean square error(RMSE) and mean absolute percentage error (MAPE) metric. The RMSE value of the MSGARCH model is the lowest. The MSGARCH model is outperforms compared to SETAR model, and it is described in Table 5.4. The ten days' forecast prediction is described in Figure 5.10. MSGARCH regime-switching performs better than the GARCH and SETAR models. The reason is that the SETAR model regime-switching is based on the Auto-Regressive process, and it ignores the regime changes when the stock price has higher volatility. However, there are performance variations in the homogeneous MSGARCH and heterogeneous MSGARCH models. Some stock Homogeneous MSGARCH model is fitted well, and some stock heterogeneous MSGARCH model is fit better. The reason is that each stock follows its volatility.

5.4 Summary

The stock price is volatile during the day trade due to many factors impacting stock prices such as demand, supply, volume growth, economic policy, and company earnings. Due to this reason, stock price volatility capturing is a complex and challenging task. Most of the existing work considered GARCH models to capture volatility, but it fails to capture different volatility variations. To capture the volatility of stock price, we have considered regime-switching based on MSGARCH and SETAR models. In this work, we have considered two MSGARCH models to estimates the volatility in stock price. The first is homogeneous switching, and the second is heterogeneous switching. The experiments considered the Indian stock market data. The performance of the MSGARCH and SETAR models is calculated using RMSE and MAPE metrics. The RMSE value of the MSGARCH model is the lowest. We have found that MSGARCH regime-switching is performed better than the GARCH and SETAR models. The reason is that stock prices are volatile and have a structural break in data; hence, GARCH models cannot fit correctly. Future work could be optimizing the GARCH parameters using the meta-heuristic method. We can also apply a machine learning algorithm for GARCH parameter selections.

Chapter 6

Conclusions and Future Work

Stock price prediction and stock price movement classification are difficult due to volatility in the stock market. The stock prices are affected by many events such as company balance sheet variation, political uncertainty, bond market rate, and global market trends. This work addresses three research objectives as stated below.

1) Stock price prediction during the stock crisis using Hybrid Feature Selection (HFS) method.

2) Feature selection based stock price movement classification and prediction.

3) Stock price volatility estimation using the regime switching models.

To the best of our knowledge, this is the first approach is to address stock price prediction during the stock crisis based on stock financial parameters and stock price. We have proposed the Hybrid Feature Selection (HFS) algorithm to remove irrelevant stock financial parameters features. The NB classifier method is considered to find the fundamentally strong stock. Later, stock bubbles are identified by using the RSI method. Moving average statics are considered to identify the stock crisis points. The stock price crisis is predicted using the XGBoost and DNN regression method. The performance of the model is evaluated based on MSE, MAE, and RMSE. The effectiveness of XGBoost and DNN models is quantified by using the Friedman test. We have concluded that HFS based XGBoost performs better than HFS based DNN method.

The second approach is to address the stock price movements classification and stock price prediction using the technical indicators. Most of the existing work is based on machine learning techniques and considered very few technical indicators to predict and classify the stock price movements. This work addresses three tasks (a) Feature selection and stock price classification (b) Stock price classification using a combination of candlestick data and technical indicators (c) Stock price prediction. This task started with selection and identification of technical indicators using the Boruta feature selection technique, followed by stock price classification using the DNN, ANN and SVM technique. The performance of the DNN model is better than the ANN and SVM model. The proposed DNN method improved the classification accuracy rate by by 5% to 6%. In this work, the second task was to classify the stock price movements using a combination of candlestick data and technical indicators. The trend in data is identified using the combination of candlestick data and technical indicators values. National Stock Exchange (NSE), datasets are considered for an experiment from the years 2008 to 2021. The proposed DNN model outperformed the ANN method by 8-11%. The third subtask in this work is prediction of stock price using the ANN method with and without Boruta feature selection. The model performance was evaluated using metrics like MAE and RMSE. The experimental results showed that ANN with Boruta feature selection method outperformed the ANN model without Boruta feature selection.

The third research objectives in this work is to address the stock price volatility estimation is using the regime-switching models. Most of the existing work considered GARCH models to capture volatility, but these models fails to capture different volatility variations. To capture the volatility of stock price during the structural break, we have considered regime-switching based on MSGARCH and SETAR models. In this work, we have considered two MSGARCH models to estimates the volatility in stock price namely, MSGARCH model with homogeneous switching and MSGARCH model with heterogeneous switching. The performance of the MSGARCH and SETAR models is calculated using RMSE and MAPE metrics. The RMSE value of the MSGARCH model is the lowest. It is found that regime-switching models performed better than the GARCH models. The reason is that stock prices are volatile and have a structural break in data; hence, the regime switching models performed better.

During this research work, we have identified two research gaps which are left unaddressed due to lack of time.

- In future, the researchers can explore other technical indicators to predict the crisis point. It is possible to improve and fine-tune the XGBoost method with different optimizers.
- In future GARCH parameters can be optimized using the meta-heuristic methods and machine learning techniques.

6.1 Appendix

6.1.1 Course Work

SI No	Name of Subject	Subject Code	No. of Credit	Grade
1	Research Methodology	HU800	2	S
2	Advanced Database System	IT701	4	BC
3	Data Warehousing and Data Mining	IT905	4	BB
4	Optimization Techniques and Random Processes	MA712	4	CD
Completed Course work with CGPA 6.67				

Figure 6.1 Course work details

6.1.2 Work Timeline

Stage	Activity	Period	Status
1	Course work	July 2016 - May 2017	Completed
2	Litrature survey	June 2016	Completed
3	Research proposal	January 2018	Completed
4	Research progress seminar I	May 2019	Completed
5	Research progress seminar II	March 2020	Completed
6	Pre-synopsis	June 2021	Completed
7	Thesis submission	July 2021	Completed

Figure 6.2 Tentative research schedule

List of Publications

International Conferences

1. **Nagaraj Naik**, and Biju R Mohan (2019), "Optimal Feature Selection of Technical Indicator and Stock Prediction using Machine Learning Technique", In 2^{nd} *International Conference on Emerging Technologies in Computer Engineering (ICETCE),* held at SKIT Jaipur, India. Pages 261-268, DOI 10.1007/978-981-13-8300-7_22 (Scopus indexed).

2. **Nagaraj Naik**, and Biju R Mohan (2019), "Stock Price Movements Classification using Machine and Deep Learning Techniques-The Case Study of Indian Stock Market", In 20^{th} *International Conference on Engineering Applications of Neural Networks(EANN),* held at University of Piraeus,Greece. Pages 445-452, DOI 10.1007/978-3-030-20257-6_38 (Scopus indexed).

3. **Nagaraj Naik**, and Biju R Mohan (2019), "Study of Stock Return Predictions using Recurrent Neural Networks with LSTM", In 20^{th} *International Conference on Engineering Applications of Neural Networks(EANN),* held at University of Piraeus,Greece. Pages 453-459, DOI 10.1007/978-3-030-20257-6_39 (Scopus indexed).

4. **Nagaraj Naik**, and Biju R Mohan (2019), "GARCH-Model Identification Based on Performance of Information Criteria", In 3^{rd} *International Conference on Computing and Network Communications(CoCoNet),* held at IIITM Kerala, India. Pages 1945-1942, DOI 10.1016/j.procs.2020.04.207 (Scopus indexed).

5. **Nagaraj Naik**, and Biju R Mohan (2019), "GARCH Model identification for Stock Crises Events", In 3^{rd} *International Conference on Computing and Network Communications(CoCoNet),* held at IIITM Kerala, India. Pages 1945-1942, DOI 10.1016/j.procs.2020.04.187 (Scopus indexed).

6. **Nagaraj Naik**, and Biju R Mohan (2020), "Log Periodic Power Law Fitting on Indian Stock Market", In *International Conference on Machine Learning, Image*

Processing, Network Security and Data sciences(MIND), held at NIT Silchar, India. Pages 261-268, DOI 10.1007/978-981-15-6318-8.4 (Scopus indexed).

International Journals

1. **Nagaraj Naik** and Biju R Mohan (2019), "Intraday Stock Prediction Based on Deep Neural Network", *Journal of National Academy Science Letters, Springer*, Pages 241-246, DOI 10.1007/s40009-019-00859-1 (SCIE & Scopus indexed).

2. **Nagaraj Naik** and Biju R Mohan (2021), "Novel Stock Crisis Prediction Technique- A Study on Indian Stock Market", *IEEE Access*, Pages 86230-86242, DOI 10.1109/ACCESS.2021.3088999 (SCIE & Scopus indexed).

3. **Nagaraj Naik** and Biju R Mohan(2021), "Stock Price Volatility estimation using Regime Switching Technique", *Mathematics Journal, MDPI*, Pages 1595, DOI 10.3390/math9141595 (SCIE & Scopus indexed).

References

Aalen, O. O. (1989). “A linear regression model for the analysis of life times”. *Statistics in medicine*, 8(8):907–925.

Aarstol, M. (1999). “Inflation, inflation uncertainty, and relative price variability”. *Southern Economic Journal*, pages 414–423.

Abdullah, S., Siddiqua, S., Siddiquee, M. S. H., and Hossain, N. (2017). “Modeling and forecasting exchange rate volatility in Bangladesh using GARCH models: a comparison based on normal and Student’st-error distribution”. *Financial Innovation*, 3(1):1–19.

Aditya, M., Helen, A., and Suryana, I. (2021). “Naïve bayes and maximum entropy comparison for translated novel’s genre classification”. In *Journal of Physics: Conference Series*, volume 1722, page 012007. IOP Publishing.

Alexander, S. S. (1961). “Price movements in speculative markets: Trends or random walks”. *Industrial Management Review (pre-1986)*, 2(2):7.

Allaro, H. B., Kassa, B., and Hundie, B. (2011). “A time series analysis of structural break time in the macroeconomic variables in Ethiopia”. *African Journal of Agricultural Research*, 6(2):392–400.

Aloui, C. and Jammazi, R. (2009). “The effects of crude oil shocks on stock market shifts behaviour: A regime switching approach”. *Energy economics*, 31(5):789–799.

Anbalagan, T. and Maheswari, S. U. (2015). “Classification and prediction of stock market index based on fuzzy metagraph”. *Procedia Computer Science*, 47:214–221.

Ardia, D., Bluteau, K., Boudt, K., Catania, L., and Trottier, D.-A. (2016). "Markov-switching GARCH models in R: The MSGARCH package". *Journal of Statistical Software, Forthcoming*.

Ardia, D., Bluteau, K., Boudt, K., Catania, L., and Trottier, D.-A. (2019). "Markov-switching GARCH models in R: The MSGARCH package". *Journal of Statistical Software*, 91(4).

Ariyo, A. A., Adewumi, A. O., and Ayo, C. K. (2014). "Stock price prediction using the ARIMA model". In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pages 106–112. IEEE.

Atoi, N. V. (2014). "Testing volatility in Nigeria stock market using GARCH models". *CBN Journal of Applied Statistics*, 5(2):65–93.

Ballings, M., Van den Poel, D., Hespeels, N., and Gryp, R. (2015). "Evaluating multiple classifiers for stock price direction prediction". *Expert Systems with Applications*, 42(20):7046–7056.

Barak, S., Arjmand, A., and Ortobelli, S. (2017). "Fusion of multiple diverse predictors in stock market". *Information Fusion*, 36:90–102.

Barak, S. and Modarres, M. (2015). "Developing an approach to evaluate stocks by forecasting effective features with data mining methods". *Expert Systems with Applications*, 42(3):1325–1339.

Bauwens, L., Preminger, A., and Rombouts, J. V. (2006). "Regime switching GARCH models". *Available at SSRN 914144*.

BenSaïda, A. (2015). "The frequency of regime switching in financial market volatility". *Journal of Empirical Finance*, 32:63–79.

Berutich, J. M., López, F., Luna, F., and Quintana, D. (2016). "Robust technical trading strategies using GP for algorithmic portfolio selection". *Expert Systems with Applications*, 46:307–315.

Blair, B. J., Poon, S.-H., and Taylor, S. J. (2010). "Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns". pages 1333–1344.

Bollen, J., Mao, H., and Zeng, X. (2011). "Twitter mood predicts the stock market". *Journal of computational science*, 2(1):1–8.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity". *Journal of econometrics*, 31(3):307–327.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). "A training algorithm for optimal margin classifiers". In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM.

Bougoudis, I., Iliadis, L., and Papaleonidas, A. (2014). "Fuzzy inference ANN ensembles for air pollutants modeling in a major urban area: the case of Athens". In *International Conference on Engineering Applications of Neural Networks*, pages 1–14. Springer.

BSE (2021). "Bombay Stock Exchange". `https://www.bseindia.com/`. [Online; accessed 19-July-2008].

Calvet, L. E. and Fisher, A. J. (2004). "How to forecast long-run volatility: Regime switching and the estimation of multifractal processes". *Journal of Financial Econometrics*, 2(1):49–83.

Candel, A., Parmar, V., LeDell, E., and Arora, A. (2016). "Deep learning with H2O". *H2O. ai Inc.*

Cao, H., Badescu, A., Cui, Z., and Jayaraman, S. K. (2020). "Valuation of VIX and target volatility options with affine GARCH models". *Journal of Futures Markets*, 40(12):1880–1917.

Caporale, G. M. and Zekokh, T. (2019). "Modelling volatility of cryptocurrencies using Markov-Switching GARCH models". *Research in International Business and Finance*, 48:143–155.

Cervelló-Royo, R., Guijarro, F., and Michniuk, K. (2015). "Stock market trading rule based on pattern recognition and technical analysis: Forecasting the DJIA index with intraday data". *Expert Systems with Applications*, 42(14):5963–5975.

Chan, W. S. (2003). "Stock price reaction to news and no-news: drift and reversal after headlines". *Journal of Financial Economics*, 70(2):223–260.

Chand, S., Kamal, S., and Ali, I. (2012). "Modeling and volatility analysis of share prices using ARCH and GARCH models". *World Applied Sciences Journal*, 19(1):77–82.

Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., and Vlachogiannakis, N. (2018). "Forecasting stock market crisis events using deep and statistical machine learning techniques". *Expert Systems with Applications*, 112:353–371.

Chen, H., Zhang, J., Tao, Y., and Tan, F. (2019). "Asymmetric GARCH type models for asymmetric volatility characteristics analysis and wind power forecasting". *Protection and Control of Modern Power Systems*, 4(1):1–11.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2015). "Xgboost: extreme gradient boosting". *R package version 0.4-2*, 1(4).

Chen, T.-l. and Chen, F.-y. (2016). "An intelligent pattern recognition model for supporting investment decisions in stock market". *Information Sciences*, 346:261–274.

Chon, K. H. and Cohen, R. J. (1997). "Linear and nonlinear ARMA model parameter estimation using an artificial neural network". *IEEE transactions on biomedical engineering*, 44(3):168–174.

Chourmouziadis, K. and Chatzoglou, P. D. (2016). "An intelligent short term stock trading fuzzy system for assisting investors in portfolio management". *Expert Systems with Applications*, 43:298–311.

Ciregan, D., Meier, U., and Schmidhuber, J. (2012). "Multi-column deep neural networks for image classification". In *2012 IEEE conference on computer vision and pattern recognition*, pages 3642–3649. IEEE.

Demertzis, K., Iliadis, L. S., and Anezakis, V.-D. (2018). "Extreme deep learning in biosecurity: the case of machine hearing for marine species identification". *Journal of Information and Telecommunication*, 2(4):492–510.

Dhaene, G. and Wu, J. (2020). "Incorporating overnight and intraday returns into multivariate GARCH volatility models". *Journal of Econometrics*, 217(2):471–495.

Dieobold, F. X. (1986). "Modeling the persistence of conditional variances: A comment". *Econometric Reviews*, 5(1):51–56.

Emenogu, N. G., Adenomon, M. O., and Nweze, N. O. (2020). "On the volatility of daily stock returns of Total Nigeria Plc: evidence from GARCH models, value-at-risk and backtesting". *Financial Innovation*, 6(1):1–25.

Endri, E., Abidin, Z., Simanjuntak, T. P., Nurhayati, I., et al. (2020). "Indonesian Stock Market Volatility: GARCH Model". *Montenegrin Journal of Economics*, 16(2):7–17.

Engle, R. F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Enke, D. and Mehdiyev, N. (2013). "Stock market prediction using a combination of stepwise regression analysis, differential evolution-based fuzzy clustering, and a fuzzy inference neural network". *Intelligent Automation & Soft Computing*, 19(4):636–648.

Enke, D. and Thawornwong, S. (2005). "The use of data mining and neural networks for forecasting stock market returns". *Expert Systems with applications*, 29(4):927–940.

Escobar-Anel, M., Rastegari, J., and Stentoft, L. (2021). "Option pricing with conditional GARCH models". *European Journal of Operational Research*, 289(1):350–363.

Fakhfekh, M. and Jeribi, A. (2020). "Volatility dynamics of crypto-currencies' returns: Evidence from asymmetric and long memory GARCH models". *Research in International Business and Finance*, 51:101075.

Fan, J., Zheng, J., Wu, L., and Zhang, F. (2021). "Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models". *Agricultural Water Management*, 245:106547.

Filimonov, V. and Sornette, D. (2013). "A stable and robust calibration scheme of the log-periodic power law model". *Physica A: Statistical Mechanics and its Applications*, 392(17):3698–3707.

Fleitas, S., Fishback, P., and Snowden, K. (2018). "Economic crisis and the demise of a popular contractual form: Building & Loans in the 1930s". *Journal of Financial Intermediation*, 36:28–44.

Franses, P. H. and Van Dijk, D. (1996). "Forecasting stock market volatility using (non-linear) Garch models". *Journal of Forecasting*, 15(3):229–235.

Gabriel, A. S. (2012). "Evaluating the forecasting performance of GARCH models. Evidence from Romania". *Procedia-Social and Behavioral Sciences*, 62:1006–1010.

Glezakos, T. J., Tsiligiridis, T. A., Iliadis, L. S., Yialouris, C. P., Maris, F. P., and Ferentinos, K. P. (2009). "Feature extraction for time-series data: An artificial neural network evolutionary training model for the management of mountainous watersheds". *Neurocomputing*, 73(1-3):49–59.

Göçken, M., Özçalıcı, M., Boru, A., and Dosdoğru, A. T. (2019). "Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection". *Neural Computing and Applications*, 31(2):577–592.

Graham, B. (1965). *"The intelligent investor"*. Prabhat Prakashan.

Graham, B. and McGolrick, C. (1975). *"The interpretation of financial statements"*. HarperCollins Publishers.

Gulisashvili, A. (2012). *"Analytically tractable stochastic stock price models"*. Springer Science & Business Media.

Gummesson, E. (2000). *"Qualitative methods in management research"*. Sage.

Gunduz, H., Yaslan, Y., and Cataltepe, Z. (2017). "Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations". *Knowledge-Based Systems*, 137:138–148.

Hajizadeh, E., Seifi, A., Zarandi, M. F., and Turksen, I. (2012). "A hybrid modeling approach for forecasting the volatility of S&P 500 index return". *Expert Systems with Applications*, 39(1):431–436.

Hansen, P. R. and Lunde, A. (2005). "A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?" *Journal of applied econometrics*, 20(7):873–889.

Hassan, M. R., Nath, B., and Kirley, M. (2007). "A fusion model of HMM, ANN and GA for stock market forecasting". *Expert systems with Applications*, 33(1):171–180.

Hsu, H.-H., Hsieh, C.-W., and Lu, M.-D. (2011). "Hybrid feature selection by combining filters and wrappers". *Expert Systems with Applications*, 38(7):8144–8150.

Hu, W., Si, Y.-W., Fong, S., and Lau, R. Y. K. (2019). "A formal approach to candlestick pattern classification in financial time series". *Applied Soft Computing*, 84:105700.

Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E., and Liu, M. (2015). "Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review". *Applied Soft Computing*, 36:534–551.

Hu, Y., Ni, J., and Wen, L. (2020). "A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction". *Physica A: Statistical Mechanics and its Applications*, 557:124907.

Huang, Y., Dai, X., Wang, Q., and Zhou, D. (2021). "A hybrid model for carbon price forecasting using GARCH and long short-term memory network". *Applied Energy*, 285:116485.

Iliadis, L., Spartalis, S., Paschalidou, A., and Kassomenos, P. (2007). "Artificial neural network modelling of the surface ozone concentration". *International Journal of Computational and Applied Mathematics*, 2(2):125–139.

Ince, H. and Trafalis, T. B. (2008). "Short term forecasting with support vector machines and application to stock price prediction". *International Journal of General Systems*, 37(6):677–687.

Ioannidis, F., Kosmidou, K., Savva, C., and Theodossiou, P. (2021). "Electricity pricing using a periodic GARCH model with conditional skewness and kurtosis components". *Energy Economics*, page 105110.

Jacobsson, E. (2009). "How to predict crashes in financial markets with the Log-Periodic Power Law". *Master diss., Department of Mathematical Statistics, Stockholm University*.

Jin, Z., Yang, Y., and Liu, Y. (2019). "Stock closing price prediction based on sentiment analysis and LSTM". *Neural Computing and Applications*, pages 1–17.

Johansen, A. and Sornette, D. (1999). "Log-periodic power law bubbles in Latin-American and Asian markets and correlated anti-bubbles in Western stock markets: An empirical study". *arXiv preprint cond-mat/9907270*.

Junyu, H. (2020). "Prediction of Financial Crisis Based on Machine Learning". pages 71–75.

Kara, Y., Boyacioglu, M. A., and Baykan, Ö. K. (2011). "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange". *Expert systems with Applications*, 38(5):5311–5319.

Keown, A. J. and Pinkerton, J. M. (1981). "Merger announcements and insider trading activity: An empirical investigation". *The journal of finance*, 36(4):855–869.

Kiersz, A. (2015). "Here's how badly warren buffett beat the market".

Klaassen, F. (2002). "Improving GARCH volatility forecasts with regime-switching GARCH". In *Advances in Markov-switching models*, pages 223–254. Springer.

Kristjanpoller, W. and Minutolo, M. C. (2015). "Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model". *Expert systems with applications*, 42(20):7245–7251.

Kursa, M. B., Rudnicki, W. R., et al. (2010). "Feature selection with the Boruta package". *J Stat Softw*, 36(11):1–13.

Laboissiere, L. A., Fernandes, R. A., and Lage, G. G. (2015). "Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks". *Applied Soft Computing*, 35:66–74.

Lamoureux, C. G. and Lastrapes, W. D. (1990). "Persistence in variance, structural change, and the GARCH model". *Journal of Business & Economic Statistics*, 8(2):225–234.

Lee, T. K., Cho, J. H., Kwon, D. S., and Sohn, S. Y. (2019). "Global stock market investment strategies based on financial network indicators using machine learning techniques". *Expert Systems with Applications*, 117:228–242.

Leivo, T. H. and Pätäri, E. J. (2011). "Enhancement of value portfolio performance using momentum and the long-short strategy: The Finnish evidence". *Journal of Asset Management*, 11(6):401–416.

Li, A. W. and Bastos, G. S. (2020). "Stock market forecasting using deep learning and technical analysis: a systematic review". *IEEE Access*, 8:185232–185242.

Li, C. (2017). "Log-periodic view on critical dates of the Chinese stock market bubbles". *Physica A: Statistical Mechanics and its Applications*, 465:305–311.

Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., Min, H., and Deng, X. (2016). "Empirical analysis: stock market prediction via extreme learning machine". *Neural Computing and Applications*, 27(1):67–78.

Lin, F., Liang, D., and Chen, E. (2011). "Financial ratio selection for business crisis prediction". *Expert Systems with Applications*, 38(12):15094–15102.

Loomis, C. (2012). “Buffett beats the sp for the 39th year”.

Maciel, L. (2020). “Cryptocurrencies value-at-risk and expected shortfall: Do regime-switching volatility models improve forecasting?” *International Journal of Finance & Economics*.

Malkiel, B. G. (2003). “The efficient market hypothesis and its critics”. *Journal of economic perspectives*, 17(1):59–82.

Malkiel, B. G. and Fama, E. F. (1970). “Efficient capital markets: A review of theory and empirical work”. *The journal of Finance*, 25(2):383–417.

Marcucci, J. et al. (2005). “Forecasting stock market volatility with regime-switching GARCH models”. *Studies in Nonlinear Dynamics & Econometrics*, 9(4):1–53.

Mitchell, M. L. and Mulherin, J. H. (1994). “The impact of public information on the stock market”. *The Journal of Finance*, 49(3):923–950.

Molnár, P. (2016). “High-low range in GARCH models of stock return volatility”. *Applied Economics*, 48(51):4977–4991.

Naeem, M., Tiwari, A. K., Mubashra, S., and Shahbaz, M. (2019). “Modeling volatility of precious metals markets by using regime-switching GARCH models”. *Resources Policy*, 64:101497.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., and Muharemagic, E. (2015). “Deep learning applications and challenges in big data analytics”. *Journal of big data*, 2(1):1–21.

Najjar, E. and Al-augby, S. (2021). “Sentiment Analysis Combination in Terrorist Detection on Twitter: A Brief Survey of Approaches and Techniques”. *Research in Intelligent and Computing in Engineering*, pages 231–240.

Nayak, R. K., Mishra, D., and Rath, A. K. (2015). “A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices”. *Applied Soft Computing*, 35:670–680.

Niaki, S. T. A. and Hoseinzade, S. (2013). "Forecasting S&P 500 index using artificial neural networks and design of experiments". *Journal of Industrial Engineering International*, 9(1):1.

Noda, T., Nagaoka, N., and Ametani, A. (1996). "Phase domain modeling of frequency-dependent transmission lines by means of an ARMA model". *IEEE Transactions on Power Delivery*, 11(1):401–411.

NSE (2021). "National Stock Exchange". `https://www.nseindia.com/`. [Online; accessed 19-July-2008].

O'Gorman, T. W. (2001). "A comparison of the F-test, Friedman's test, and several aligned rank tests for the analysis of randomized complete blocks". *Journal of Agricultural, Biological, and Environmental Statistics*, 6(3):367–378.

Pai, P.-F. and Lin, C.-S. (2005). "A hybrid ARIMA and support vector machines model in stock price forecasting". *Omega*, 33(6):497–505.

Pan, Z., Wang, Y., Wu, C., and Yin, L. (2017). "Oil price volatility and macroeconomic fundamentals: A regime switching GARCH-MIDAS model". *Journal of Empirical Finance*, 43:130–142.

Patel, J., Shah, S., Thakkar, P., and Kotecha, K. (2015a). "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques". *Expert Systems with Applications*, 42(1):259–268.

Patel, J., Shah, S., Thakkar, P., and Kotecha, K. (2015b). "Predicting stock market index using fusion of machine learning techniques". *Expert Systems with Applications*, 42(4):2162–2172.

Pouya, A. R., Solimanpur, M., and Rezaee, M. J. (2016). "Solving multi-objective portfolio optimization problem using invasive weed optimization". *Swarm and Evolutionary Computation*, 28:42–57.

Prechter Jr, R. R. and Parker, W. D. (2007). "The financial/economic dichotomy in social behavioral dynamics: the socionomic perspective". *The Journal of Behavioral Finance*, 8(2):84–108.

Preis, T., Moat, H. S., and Stanley, H. E. (2013). "Quantifying trading behavior in financial markets using Google Trends". *Scientific reports*, 3:srep01684.

Qiu, M., Song, Y., and Akagi, F. (2016). "Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market". *Chaos, Solitons & Fractals*, 85:1–7.

RBI (2021). "Reseve Bank Of India". `https://rbi.org.in/scripts/BS_ViewBulletin.aspx?Id=18878#11`. [Online; accessed 19-July-2008].

Robert, E. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*.

Rodríguez-González, A., García-Crespo, Á., Colomo-Palacios, R., Iglesias, F. G., and Gómez-Berbís, J. M. (2011). "CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator". *Expert systems with Applications*, 38(9):11489–11500.

Roll, R. (1988). "The international crash of October 1987". *Financial analysts journal*, 44(5):19–35.

Ronaghi, F., Salimibeni, M., Naderkhani, F., and Mohammadi, A. (2020). "ND-SMPF: A Noisy Deep Neural Network Fusion Framework for Stock Price Movement Prediction". In *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, pages 1–7. IEEE.

Sapuric, S., Kokkinaki, A., and Georgiou, I. (2020). "The relationship between Bitcoin returns, volatility and volume: asymmetric GARCH modeling". *Journal of Enterprise Information Management*.

Sezer, O. B. and Ozbayoglu, A. M. (2018). "Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach". *Applied Soft Computing*, 70:525–538.

Sharma, P. et al. (2015). "Forecasting stock index volatility with GARCH models: international evidence". *Studies in Economics and Finance*.

Singh, S., Parmar, K. S., and Kumar, J. "Soft computing model coupled with statistical models to estimate future of stock market". *Neural Computing and Applications*, pages 1–19.

Smith, V. L. (2003). "Constructivist and ecological rationality in economics". *American economic review*, 93(3):465–508.

Song, X., Kim, D., Yuan, H., Cui, X., Lu, Z., Zhou, Y., and Wang, Y. (2020). "Volatility analysis with realized GARCH-Itô models". *Journal of Econometrics*.

Sornette, D. (2009). "Dragon-kings, black swans and the prediction of crises". *arXiv preprint arXiv:0907.4290*.

Srikanth, K., Murthy, N., and Reddy, P. P. (2021). "Sentiment Classification on Online Retailer Reviews". In *ICCCE 2020*, pages 1557–1563. Springer.

Stephen, O., Sain, M., Maduh, U. J., and Jeong, D.-U. (2019). "An efficient deep learning approach to pneumonia classification in healthcare". *Journal of healthcare engineering*, 2019.

Sun, H. and Yu, B. (2020). "Volatility asymmetry in functional threshold GARCH model". *Journal of Time Series Analysis*, 41(1):95–109.

Suzuki, T. and Ohkura, Y. (2016). "Financial technical indicator based on chaotic bagging predictors for adaptive stock selection in Japanese and American markets". *Physica A: Statistical Mechanics and its Applications*, 442:50–66.

TRINH, Q. T., NGUYEN, A. P., NGUYEN, H. A., and NGO, P. T. (2020). "Determinants of Vietnam government bond yield volatility: A GARCH approach". *The Journal of Asian Finance, Economics, and Business*, 7(7):15–25.

Tsai, C.-F., Lin, Y.-C., Yen, D. C., and Chen, Y.-M. (2011). "Predicting stock returns by classifier ensembles". *Applied Soft Computing*, 11(2):2452–2459.

Vaisla, K. S. and Bhatt, A. K. (2010). "An analysis of the performance of artificial neural network technique for stock market forecasting". *International Journal on Computer Science and Engineering*, 2(6):2104–2109.

Vapnik, V. N. and Chervonenkis, A. J. (1974). "Theory of pattern recognition".

Vu, T.-T., Chang, S., Ha, Q. T., and Collier, N. (2012). "An experiment in integrating sentiment features for tech stock prediction in twitter".

Wang, Q. and Wang, Z. (2021). "VIX futures and its closed-form pricing through an affine GARCH model with realized variance". *Journal of Futures Markets*, 41(1):135–156.

Wosnitza, J. H. and Denz, C. (2013). "Liquidity crisis detection: An application of log-periodic power law structures to default prediction". *Physica A: Statistical Mechanics and its Applications*, 392(17):3666–3681.

Xing, D.-Z., Li, H.-F., Li, J.-C., and Long, C. (2021). "Forecasting price of financial market crash via a new nonlinear potential GARCH model". *Physica A: Statistical Mechanics and its Applications*, 566:125649.

Yamaguchi, K. (2008). "Reexamination of stock price reaction to environmental performance: A GARCH application". *Ecological Economics*, 68(1-2):345–352.

Yürekli, K., Kurunç, A., and Öztürk, F. (2005). "Testing the residuals of an ARIMA model on the Cekerek Stream Watershed in Turkey". *Turkish Journal of Engineering and Environmental Sciences*, 29(2):61–74.

Zhang, L., Wang, F., Xu, B., Chi, W., Wang, Q., and Sun, T. (2018). "Prediction of stock prices based on LM-BP neural network and the estimation of overfitting point by RDCI". *Neural Computing and Applications*, 30(5):1425–1444.

Zhang, Q., Sornette, D., Balcilar, M., Gupta, R., Ozdemir, Z. A., and Yetkiner, H. (2016a). "LPPLS bubble indicators over two centuries of the S&P 500 index". *Physica A: statistical Mechanics and its Applications*, 458:126–139.

Zhang, Q., Zhang, Q., and Sornette, D. (2016b). "Early warning signals of financial crises with multi-scale quantile regressions of Log-Periodic Power Law Singularities". *PloS one*, 11(11):e0165819.

Zhang, Y.-J., Yao, T., He, L.-Y., and Ripple, R. (2019). "Volatility forecasting of crude oil market: Can the regime switching GARCH model beat the single-regime GARCH models?" *International Review of Economics & Finance*, 59:302–317.

Zhong, X. and Enke, D. (2017). "A comprehensive cluster and classification mining procedure for daily stock market return forecasting". *Neurocomputing*, 267:152–168.

Brief Bio-Data

Nagaraj Naik

Research Scholar

Department of Information Technology

National Institute of Technology Surathkal

Surathkal, Mangalore

Karanataka 575025

My research interests are Financial Data Analysis, Stock Price Analysis, and Statistical Techniques.

Address of Communication

203, Sapthshree Apartment,

Gudde Kopla Main Road, Iddy,

Surathkal, Karanataka

Phone Number : 8123449690

e-mail: nagaraj21.naik@gmail.com