PREDICTIVE MODELING TECHNIQUES USING BIG DATA

THESIS

submitted in fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

J.C.BOSE UNIVERSITY OF SCIENCE & TECHNOLOGY, YMCA

by

SHRUTI MITTAL

Registration No. YMCAUST / Ph01 / 2015

Under the Supervision of

Dr. CHANDER KUMAR NAGPAL

PROFESSOR

Department of Computer Engineering

Faculty of Informatics and Computing

J.C. Bose University of Science & Technology, YMCA

Sector-6, Mathura Road, Faridabad, Haryana, India

FEBRUARY, 2022

DEDICATED

To the heroes of my life…

DECLARATION

I hereby declare that this thesis entitled "**PREDICTIVE MODELING TECHNIQUES USING BIG DATA**" being submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy in Computer Engineering under Faculty of Informatics and Computing of J.C. Bose University of Science and Technology, YMCA, Faridabad, during the academic year 2021-2022 is a bona fide record of my original work carried out under the guidance and supervision of **Dr. Chander Kumar Nagpal, Professor(Computer Engg.)**, and has not been presented elsewhere.

I further declare that the thesis does not contain any part of any work which has been submitted for the award of any degree either in this university or in any other university.

Shruti Mittal

YMCAUST / Ph01 / 2015

CERTIFICATE

This is to certify that this thesis entitled "**PREDICTIVE MODELING TECHNIQUES USING BIG DATA**" by Shruti Mittal, submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy in Computer Engineering under Faculty of Informatics and Computing of J.C. Bose University of Science and Technology, YMCA, Faridabad, during the academic year 2021-2022, is a bona fide record of work carried out under my guidance and supervision.

I further declare that to the best of my knowledge, the thesis does not contain any part of any work which has been submitted for the award of any degree either in this university or in any other university.

Dr. Chander Kumar Nagpal

Professor

Department of Computer Engineering

Faculty of Informatics and Computing

J.C. Bose University of Science and Technology, YMCA, Faridabad

Date:

The Ph.D. Viva-Voce examination of research scholar Shruti Mittal (YMCAUST / Ph01 / 2015) has been held on 18.02.2022.

(External Examiner) (Supervisor) (Chairman)

ACKNOWLEDGEMENT

It is a genuine pleasure to express my deep sense of thanks and gratitude to my guide and an honest advisor Prof. C.K.Nagpal for his continuous guidance, support and encouragement for all the research work without which this thesis would not have been possible. I am blessed to have him as my guide and would like to thank him again for his persistent efforts, motivation, and immense patience. I could not have imagined having a better advisor and mentor for my work.

Besides my advisor, I would like to thank Prof. Komal Kumar Bhatia, Prof Atul Mishra, Prof. Manjeet Singh and Prof. Ashutosh Dixit for their insightful comments and encouragement, but also for their useful suggestions which motivated me to widen my research from various perspectives.

I thank my roommates and colleagues for their stimulating discussions and all the encouragements in bringing out the best in this work.

Also I would like to thank the Department of Computer Engineering for the kind help and cooperation throughout my research period.

I owe a sincere thanks to Mr. Sanat Nagpal for providing the technical details related to stock market.

I would like to thank my mother, my husband Aman, my brother, my sister–in-law, my nephew Daksh, my kids Vansh and Saesha, my friends and my in-laws for supporting me spiritually throughout this thesis work and my life in general.

I conclude this acknowledgement with a thanks to the almighty without his blessings this work could not have been accomplished.

Shruti Mittal

ABSTRACT

Initial predictive analytics was based upon statistical methods but in recent times machine learning has taken over. Here the system tries to envisage a complex mathematical function depending upon large number of variables. Keeping in view the inherent structures of these systems, one can infer that this strategy can only be applicable in case of the environment, which are fundamentally governed by the mathematical functions. But in the natural environment, most of the times the things are not purely based upon mathematical functions and there is some contribution of natural and human elements as well. Therefore we are of the view that pure machine learning methods are not adequate for designing the predictive systems. They have to be augmented with some other mechanisms to take care of human and natural elements. The work proposed in this thesis is an effort in this direction.

The domain undertaken for the purpose of predictive analytics in the proposed work is price prediction in the Indian Stock Market. The reason for choosing this domain are: availability of data in public domain, ease of verification of input data, ease of verification of results, major characteristics of Big Data are complied.

While taking the inference from the historical data, the working pattern of all the papers is somewhat similar to that of the time series prediction wherein the basic underline philosophy is that the trend will continue. However, the stock market prediction is not a matter of mere time series trend. Moreover, prediction accuracy has not been that good in any case as it varies from 60% to 85%. An error to the tune of 15% to 20% is quite huge and can lead to mega loss in the financial market. All these conventional machine learning mechanisms suffer from the usual drawbacks of opacity and overfitting. Moreover, the random fluctuations in the stock price data which is a very common element, in the stock prices, is a big hindrance to the proper convergence. Most of the papers go for the few prominent stocks without taking care of the spectrum as a whole.

All these issues need to be taken care of for making any stock market prediction. So there is a need to come out of this mindset and efforts are required to put some human elements.

The work carried out resulted in the generation of various mechanisms for predicting the stock price movement to help the investors in making a rationale decision for their investments with an aim for continuous and long term survival in the market. The models / mechanisms proposed in the work are automatable without human intervention, in order to provide a credible advice which is free from human manipulations.

Prediction process captures the macro details from the historic price data before applying the supervised learning process, making it free from the random price fluctuations in the raw data which normally leads to overfitting. Also the proposed work is not biased towards any particular company or sector therefore this thesis takes up the whole of the Nifty50 spectrum of the Indian Stock Market. The work can be used in all the stock exchanges amongst all types of the stocks, across the globe.

The work carried out resulted in the outcome of various prediction mechanisms whose details have been provided in this thesis. The credibility of the proposed mechanisms was verified on the actual scenario and the results were quite encouraging.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BO	Buying Opportunity
BSE	Bombay Stock Exchange
CNN	Convolutional Neural Network
CO	Correlation Coefficient
CPSP	Current Price Sentiment Pointer
FRB	Fuzzy Rule Base
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
IS	Index Sentiment
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MFNN	Multilayer Feedforward Neural Network
MSI	Market Sentiment Indicator
MVS	Moving Average Signal
NMVA	Normalized Moving Average
NSE	National Stock Exchange
NYSE	New York Stock Exchange
OPSI	Off-Peak Status Indicator
PBI	Price/Bookvalue Indicator
PCA	Principal Component Analysis
PEI	Price/Earning Indicator
PGI5Q	Profit Growth Indicator In 5Quarters
PGI5Y	Profit Growth Indicator In 5Years
PS3Y	Price Sentiment Of The Stock With Respect To Past Three Years
RNN	Recurrent Neural Network
RSI	Relative Strength Index
SGD	Stochastic Gradient Descent
SGI5Q	Sales Growth Indicator In 5Quarters
SGI5Y	Sales Growth Indicator In 5Years
SHI	Stock Health Index
SMA	Simple Moving Average
SMVS	Sector Moving Average Signal
SO	Selling Opportunity
SS	Sector Sentiments
SSI	Sector Sentiments Indicator
SVD	Single Value Decomposition
SVM	Support Vector Machine

CHAPTER I
INTRODUCTION

1.1 PREDICTIVE ANALYTICS

Predictive analytics is being used in many fields such as Weather Forecasting [1]–[3], Automated Patient Monitoring [4], [5], Stock Market price prediction [6], [7], Plant maintenance [8], Vehicle Maintenance [9], [10], Diagnostic Systems etc. Initial predictive analytics was based upon statistical methods [11]–[16] but in recent times machine learning has taken over most of the predictive systems are based upon the pure machine learning algorithms and their designer hope that the machine will identify the underlying pattern in the data and make the predictions. Here the system tries to envisage a complex mathematical function depending upon large number of variables. Keeping in view the inherent structures of these systems, one can infer that this strategy can only be applicable in case of the environment, which are fundamentally governed by the mathematical functions. But in the natural environment, most of the times the things are not purely based upon mathematical functions and there is some contribution of natural and human elements as well. Therefore we are of the view that pure machine learning methods are not adequate for designing the predictive systems. They have to be augmented with some other mechanisms to take care of human and natural elements.

1.2 DOMAIN UNDERTAKEN FOR PREDICTIVE ANALYTICS

The domain undertaken for the purpose of predictive analytics in the proposed work is price prediction in the Indian Stock Market. The reason for choosing this domain are as follows:

a) Availability of data in public domain on various websites such as Rediffmoney [17], Yahoo Finance [18], Money Control [19] & NSE India [20]
b) Ease of verification of input Data
c) Ease of verification of results
d) Satisfies the major characteristics of Big Data

Here it is worth mentioning that the chosen domain satisfies the basic big data characteristics as described in Table 1.1

Table 1.1 Compliance of Big Data Characteristic by Stock Market Data		
Big Data Characteristics	Does Stock Market Domain satisfy it?	Remark
Volume	Yes	Data is voluminous.
Velocity	Yes	Data is being continuously generated very fast. For example, Stock Price data.
Variety	Yes	Heterogeneity in data related to company fundamentals, stock price, periodic sales and profit reports.
Variability	No	Normally applicable to define NLP ambiguity
Veracity	Yes	Accurate, complete and trustworthy data available which can be confirmed from websites.
Visualization	Yes	Data is accessible, comprehendible and understandable.
Value	Yes	Data and Prediction is of big financial value to the investors worldwide.

1.3 RELATED WORK AND ITS CLASSIFICATION

After going through various papers related to the stock market domain, it was found that most of the papers are related to price prediction on the short term basis. The concerns for the investors who are likely to stay regularly in the market for making continuous gain have not been addressed. Majority of the papers available in the literature are normally based upon different variants of artificial neural networks (ANN) and a variety of signal functions. The other techniques used, involve the use of time series analysis, ensemble learning and other soft computing techniques like rough sets, genetic algorithms. We now present the literature survey which has been classified on the basis of the techniques used.

1.3.1 ANN based Forecasting

Sirignano et al. [21] proposed a deep learning based mechanism trained on a set of global financial market features. The data used includes the purchasing and selling records for approximately 1000 NASDAQ stocks. NN consists of three layers with LSTM units with ReLUs in the last, with a stochastic gradient descent (SGD) algorithm as optimizer. It was also observed that feature selection process, if done before training the model, can reduce complexity.

Ni et al. [22] predicted stock price trends by using SVM and optimized it by using fractal feature selection process. The data of the Shanghai Stock Exchange Composite Index (SSECI), and 19 technical indicators were used. Feature selection with k-cross validation was done before processing the data set. Only the technical indicators are considered in this work, however major and minor factors of the financial domain have been ignored.

McNally et al. [23], used RNN and LSTM based model for Bitcoin price forecasting. The Bitcoin dataset used for forecasting was from 19 August 2013 to July 19, 2016. The model used the Boruta algorithm of feature engineering and Bayesian optimization to select LSTM parameters.

Kara et al. [24] proposed two models based on ANN and SVM in predicting stock price movements. The data they used covers the period from January 2, 1997, to December 31, 2007, on the Istanbul Stock Exchange. The work has not compared performance accuracy achieved by it with other models.

Fischer et al. [25] measured the effectiveness of a variety of deep learning methods with LSTM for financial market predictions. They analyzed the effectiveness of a variety of LSTM variables built on a daily price, standard values with daily volume, approximate weekly and weekly volume with average values and concluded that LSTM surpasses other deep learning algorithms and that it accurately captures logical information from time series data.

Long et al. [26] developed a deep learning methodology for predicting stock price movements. The data used is Chinese stock market index CSI 300. By predicting stock price movements, they created a multi-filter neural network (MFNN) with a stochastic gradient descent (SGD) and a back propagation optimizer to study NN parameters.

Jingyi Shen et al. [27] proposed a comprehensive customization of feature engineering and deep learning-based model for predicting price trend of stock markets. The system preprocess the stock market data to predict the future trends. The work is carried out in three phases namely pre-processing of data, feature engineering and trend prediction using LSTM.

Kong et al. [28] proposed a data-driven approach with liquidity measures and technical indicators to predict intraday stock jumps. Amongst various models like Random Forest, SVM, ANN, and KNN, used for making predictions, RF outperformed. Hyper parameter optimization and optimal class balancing method can further improve the performance of their work.

1.3.2 General Soft Computing Techniques based forecasting

Lei [29] used the Wavelet Neural Network (WNN) in combination with Rough Set to predict stock price trends. The database for this model contains five well-known stock market indicators, namely, (1) SSE Composite Index (China), (2) CSI 300 Index (China), (3) All Ordinaries Index (Australia), (4) Nikkei 225 Index (Japan) , and (5) the Dow Jones Index (USA).

Pimenta et al. [30] used multi-objective genetic programming for forecasting the stock market. The database was acquired in the Brazilian stock exchange (BOVESPA), and the key strategies they used were a combination of multi-purpose, genetic, and technological rules. They included a historical period, which was a critical period for Brazilian politics and economics when the validation was made. The limitation of the work is that the authors did not make any comparisons with other existing models.

1.3.3 Time Series Analysis based Forecasting

Idrees et al. [31] proposed a time series based model for forecasting stock market volatility. They designed three step feature engineering: Analyze the time series, see if the time-series is stationary or not, observe the ACF and PACF charts and identify input parameters. The work requires the customization of the ARIMA model for performance enhancement.

Weng et al. [32], focused on short-term stock forecasting using ensemble methods. The ensemble constitutes: neural network regression ensemble (NNRE), Random Forest, AdaBoost and vector regression ensemble support (SVRE). The limitation of this work includes price projections from 1 day to next 10 days with no scalability provisions. Moreover the work is based upon 20 U.S. market stocks only.

1.3.4 Others

Nekoeiqachkanloo et al. [33] proposed a model for stock investment. It is comprehensive system with data processing and two different algorithms that suggest the best components for investing. Second, the system has been embedded with the predictive component, which also retains the features of the time series. Finally, their input features are a combination of basic features and technical features that aim to fill the gap between the financial sector and the technology domain. However, their function is weak in the experimental part. Instead of testing the proposed system on a large database, they selected only 25 well-known shares.

Jeon et al. [34] used a large database based on the millisecond interval using pattern graph tracking to make stock price predictions. The data they used is a large database based on milliseconds of historical stock data from KOSCOM, from August 2014 to October 2014. The predictions were made by using ANN and Hadoop and RHive with large-scale data processing. The only limitation of the work done is that it is difficult to access millisecond-based data in real life.

Amit K. Sinha [35] estimated S&P500 index values for different periodic frequencies using drift and diffusion simulations on over nearly 1 lakh values and probabilities. Despite such a complex simulation, it is observed that the quality of the prediction reduces drastically as the period of prediction increases.

He et. al. [36] forecasted stock return volatility based upon a regression model called ARH which is a combination of autoregression (AR) model with Huber loss function. The results show that the inclusion of Huber loss function is more efficient than standalone AR model.

1.4 DEFICIENCIES IDENTIFIED IN THE LITERATURE

After going through the above work available in literature, following deficiencies have been identified which need to be taken care of:

- Most of the work available in the literature is concerned about the stock price prediction for a very short duration generally one day to one week. It may however be noted that in most of the cases the price variation during the day is not much (1% to 2%) unless some extra-ordinary news comes. So any prediction is likely to be quite accurate.
- While taking the inference from the historical data, the working pattern of all the papers is somewhat similar to that of the time series prediction wherein the basic underline philosophy is that the trend will continue. However, the stock market prediction is not a matter of mere time series trend. So there is a need to come out of this mindset and efforts are required to put some human elements.

- The assumption that the machine will be able to predict precisely is basically a wrong notion. The overall environment is a combination of multiple factors that has to be analyzed and evaluated in the holistic manner. The change of the technology from CNN to LSTM or their combination is not going to serve much of the purpose as the trends may not repeat.
- The prediction accuracy has not been that good in any case as it varies from 60% to 85%. An error to the tune of 15% to 20% is quite huge and can lead to mega loss in the financial market.
- An investor in the stock market is there to stay for making the long term fortune. He needs a credible advice in this regard which can only come with the thorough analysis of the fundamentals of the stock under consideration.
- All these papers are using the conventional machine learning mechanisms which suffer from the usual drawbacks of opacity and overfitting. Moreover, the random fluctuations in the stock price data which is a very common element, in the stock prices, is a big hindrance to the proper convergence. Most of the papers go for the few prominent stocks without taking care of the spectrum as a whole. All these issues need to be taken care of for making any stock market prediction.

1.5 PROBLEM DEFINITION AND OBJECTIVES

1.5.1 Problem Definition

To create various mechanisms for predicting the stock price movement to help the investors in making a rationale decision for their investments with an aim for continuous and long term survival in the market.

1.5.2 Objectives

1. To collect the raw data related to Indian Stock Market from various credible sources.
2. To identify the various factors which influence the price movement in stock market.
3. To extract and create different features from this raw data for the purpose of analytics

4. To apply the normalization process wherever it is required for the purpose of parity.
5. To apply ensemble of learning techniques to generate credible prediction models/mechanisms.

1.5.3 Expected outcome

The models / mechanisms proposed in the work should be automatable without human intervention, in order to provide a credible advice which is free from human manipulations.

1.6 SALIENT FEATURES OF THE PROPOSED WORK

a) The work is not biased towards any particular company or sector. To deal with this issue the work carried out in this thesis takes up the whole of the Nifty50 spectrum of the Indian Stock Market.

b) Prediction process captures the macro details from the historic price data before applying the supervised learning process, making it free from the random price fluctuations in the raw data which normally leads to overfitting.

c) The chosen macrofeatures provide different views for looking at the data thereby increasing the dimensionality of the problem and making it more effectively solvable [37].

d) The work can be used in all the stock exchanges amongst all types of the stocks, across the globe.

e) Systems proposed in this work are easily automatable.

1.7 ORGANIZATION OF THE THESIS

Chapter 2 "Literature Survey" contains the details of the study taken up to properly identify the problem to ensure that the problem to be solved is really existing and the proposed solution is state-of-the-art.

Chapter 3 "Reinforcement Learning Based Predictive Analytics Framework for Survival in Stock Market" proposes a framework that partitions the available data into historical and future part. It generates various reinforcement signals by applying statistical and machine learning techniques on the historical data and studies their impact on the stock prices by analyzing the future data. The outcome of the process has been used to generate the rewards, through the use of fuzzy logic, for various actions in a given state of the environment. Fully automated implementation of the proposed framework can help both institutional as well as common investor in taking the rational decision.

Chapter 4 "Predicting a Reliable Stock for Mid and Long Term Investment" proposes a mechanism that assesses the intrinsic health of a stock, in an automated manner and provides credible advice(s). The work contains the development of a regression based supervised learning model using feature extraction from the raw data. The regressive output of the learning model provides a stock health index that has been classified into fuzzy sets. A fuzzy rule base has been created that generates the requisite advice on the basis of stock health index. The work concludes with the identification of some fundamentally good stocks and validates the results through their quantified performance.

Chapter 5 "A Predictive Analytics Framework for Opportunity Sensing in Stock Market" proposes a supervised machine learning approach on statistically learned macrofeatures obtained from gist of input data which is free from drawbacks of raw data, to predict the price band for the upcoming month and a half for almost all NIFTY50 stocks. The predicted bands are tested for precision in comparison with actual stock price bands. Motivating outcomes so obtained were used to sense opportunity for buying / selling / wait. The results showed that the proposed strategy is quite effective and can be successfully monetized.

Chapter 6 "A Dynamically Adapting Framework for Stock Price Prediction" proposes a tailor made mechanism that involves the initial creation of candidate predictions, selection of the best pair and subjecting this pair to further reduction of error using back propagation learning. The results obtained are quite precise and scalable for the extended period.

Chapter 7 "Conclusion and Scope for Future Work" concludes the work carried out during the learning period. Since for every work there are future dimensions to which it can be extended. Therefore the chapter closes with the identification of future extension possibilities.

CHAPTER II
LITERATURE SURVEY

To identify the problem which is really existing and requires an amicable solution, one has to go through and study the available work. To accomplish this task we have gone through various website, articles, blogs and more than 100 research papers. This helped us in identifying the various issues pertaining to stock market scenario. The chapter contains two sections: Section 2.1 provides the details of the research papers and task carried out by them. Section 2.2 identifies the drawbacks and limitations of the existing work which helped us in defining the problem and the corresponding objectives.

2.1 RELATED WORK

Nayak et al. [38], in their work have predicted stock market trends using supervised machine learning algorithms consists of two models for predicting daily stock price and monthly stock price. For daily prediction, three different supervised machine learning algorithms namely Boosted DT, Logistic Regression and SVM have been used. The prediction model amalgamates historical prices with sentiments to achieve an accuracy of nearly 70% for daily price prediction. For monthly trend prediction, model uses a similar pair of two months and expects that the history will repeat. The drawback of the model is a low accuracy of daily prediction and expecting the history to repeat which may not be the case.

Kim and Kim et al. [39], proposed feature fusion long short-term memory-convolutional neural network (LSTM-CNN) based model that combines features obtained from different representations of the same data, namely, stock time series and stock chart images, to predict stock prices. The proposed model is used for extracting temporal and image features. The performance of the proposed model has been compared with the individual CNN and LSTM models on SPDR S&P 500 ETF data. The implemented framework shows that prediction error can be significantly reduced by using a combination of temporal and image features from the same data rather than using these features separately. Merely relying upon different representations of the same historical data doesn't serve much purpose particularly when the other influencing features like social media, news etc. have been ignored

Stoean et al., [40], in their work have compared the performance of two deep learning architectures namely LSTM and CNN on 25 companies listed in Bucharest Stock Exchange. The experimental dataset is from October 16, 1997 to March 13, 2019. They have reached to the conclusion that LSTM has higher gain in terms of cumulative money gains while CNN has better performance on the number of companies which have made the gain or loss as per prediction. The model gives advice to buy or sell based upon the inference value as and when the required threshold is crossed on either side. The paper admits that the work could have been better had there been an exploitation for the textual data available in the form of social media and financial news.

Bao et al., [41], have presented a novel deep learning framework where wavelet transforms (WT), stacked auto encoders (SAEs) and long-short term memory (LSTM) have been combined for stock price forecasting. The predictive accuracy and profitability of the proposed model outperformed the other three (RNN, LSTM and WLSTM). The prediction period is limited to one day only. The paper utilizes multiple stages wherein a thorough mechanism for hyper-parameters tuning could have been adopted for performance improvement.

The above literature is limited to the use of historical dataset for making future stock price prediction. However the historical dataset solely cannot make accurate future price predictions. Various factors such as company fundaments, stock exchange indices (national and international), sector indices, social media and news shall also be considered. Many researchers have explored these aspects but the coverage is only qualitative [32], [42]–[46]. Many researchers have also considered the effect of Google trends and financial features in predicting the future prices of the stock [44], [47], [48]. A combination of historical price with technical indicators [24], [49]–[51] and company fundamentals have improved the overall performance of the framework. Khan et al. [52], in their work have undertaken a study for stock market prediction based upon ensemble learning, wherein multiple machine learning classifiers are combined. Ironically, the predictions are not that accurate neither in isolation mode nor in ensemble mode and the accuracy varies from 65% to 85% which cannot be of much use for making financial investments.

Sim et al. [53], proposed a stock price prediction model based on CNN. The model takes up 9 technical indicators and converts them into images of time series graph and optimizes it using CNN. The performance of the system has been compared with ANN and SVM based models. The drawback of the model is overdependence on the technical indicators and ignorance of the external factors like financial news, social media effects and the sector sentiments.

Jiawei and Murata, [54], in their work have studied and analyzed the impact of various factors such as Sentiment Analysis, Feature Reduction, Macroeconomics indices which can influence the prediction performance. The drawback of the model is that it merely works upon the 30 day historical data (which is quite a short period) and predicts the outcome for coming 3 days only.

Zhong and Enke, [55] have presented a big data analytics process to predict the daily return direction of the SPDR S&P 500 ETF based on 60 financial and economic features. The inability of handling such huge number of features has led them to go for dimensionality reduction using PCA that helps in removing the insignificant features. They have found out that the prediction accuracy increases quite significantly after the adoption of PCA. They have also explored the variation in the number of hidden layers in the DNN. The major contribution of the paper is to consider the large number of features and the removal of the insignificant ones to improve the performance of the model.

Minh et al. [56], in their work have proposed a framework to predict the directions of stock prices by using financial news and sentiment dictionary. The study involves a novel two stream Gated Recurrent Unit Network and Stock2Vec - a sentiment word embedding trained on financial news dataset and Harvard IV-4. The drawbacks of the proposed scheme are the non-consideration of vital factors like P/E ratio, very high time complexity of the system, requirement of huge computational resources and very long training process.

Ngo and Truong [57], in their work proposed a hybrid time series forecast model based on moving-window firefly algorithm (FA)-based least squares support vector regression (MFA-LSSVR). It uses patterns of historical data to predict future values of time series data. In this model, FA is used to optimize the parameters (fine tune) to improve the accuracy of prediction. The results of the proposed model are compared with the LSSVR and the ARIMA (autoregressive integrated moving average) in predicting energy demand and stock price and were found to be more accurate. However, the proposed model has not considered all factors that affect the values of time series data like weather, temperature of the day etc in energy demand and political events, new government rules and regulations etc in stock market. Moreover the work is limited to make only one-day-ahead forecasts which is a very short period and may not generate fruitful results.

Arora and Saha, [58], in their work, proposed a hybrid Software Fault Prediction model based on artificial neural network (ANN) using firefly Algorithm for optimization of connection weights. The performance of genetic, particle swarm and firefly based evolutionary algorithms for the optimization of the connections weights in an ANN on 7 different real time project's datasets was compared. It was observed that ANN model whose optimal set of weights and biases were achieved by FA performed better in comparison to the genetic, particle swarm based adaptive ANN and BPNN fault prediction models. This technique can reduce the software project's cost and hence improve the software quality.

Ouhibi et al. [59], in their work, proposed three neural networks based methods for fault detection and isolation of asynchronous machine: a probabilistic neural network (PNN), multi-layer perceptron (MLP), and generalized regression neural network (GRNN). The efficiency of these three neural based methods is compared and PNN method was observed to outperform the other two methods.

Kumar an Inbarani, [60], in their work, proposed the covering rough set (CRS)-based classification method for classification of heartbeats to sign interior cardiac arrhythmia in ECG signals. Time-domain features are extracted from the ECG signal morphological features using PT and statistical features using WT method. These features show its high predictiveness and good accuracy in identifying different ECG activities. The proposed CRS-based classification outperforms the decision table, back propagation network, multi-layer perceptron, multiclass classification, JRip and J48 classifiers in terms of accuracy, precision, recall, F-measure etc.

Khalouli et al. [61], in their work, proposed a hybrid approach combining an ant colony optimization (ACO) and variable neighborhood search (VNS) The performance the proposed algorithms was tested on a large number of randomly generated instances. Comparisons with optimal solutions are presented. The results show the effectiveness of our proposed methods.

The results show the number of problems solved, and efficiency, in computational time, of hybrid ACO-VNS was better than that of VNS and ACOLS

Zhao, et al. [62], in their work, proposed a dynamic search strategy based on quantum particle swarm optimization to strengthen the ability of escaping from local optima, and replaced the attractor with beta distribution for faster convergence speed. This algorithm was then compared with PSO and QPSO on general optimization benchmark functions. It was observed that the proposed algorithm outperforms the traditional PSO and QPSO algorithms.

In the current scenario, most of the available recent literature related to stock market prediction is based upon machine learning. The researchers have tried to identify many features which contribute to the change in the stock price. These features include various macroeconomic factors, stock fundamentals, market sentiments, news and social media etc.

Despite all these issues, many people feel that the stock market behavior is random by nature. Meese and Rogoff [63], amazed the researchers by showing that a random walk ''no change" forecast for exchange rates is more accurate than model based forecasting. Similarly, Duffee [64], [65] showed that a random walk forecast for future interest rates is much superior than the model based predictions.

Eicher et al, [66] in their work evaluated predictions utilizing three measurements: (i) predisposition, which estimates predictions from actual values; (ii) effective-ness, which estimates whether the predictions errors were unusual and (iii) content, which estimates the worth of predictions compared to the naïve prediction models. They concluded that these predictions are ideal when they are impartial and effective.

Goyal and Welch [67] in their work conducted a comprehensive evaluation of many predictors for stock market index returns, but find that none of them can convincingly beat the unconditional mean.

Nikou et al. [68] tried to assess the forecasting power of machine learning models for stock price prediction using daily close price value of iShares MSCI United Kingdom exchange-traded fund from January 2015 to June 2018. The results indicate that the machine learning model outperforms the conventional stock prediction techniques.

Baek and Kim [69] proposed a hybrid model ModAug-Net having two LSTM modules: overfitting Prevention Module and Prediction Module. Their model uses S&P500 and Korea Composite Stock Price Index 200 (KOSPI200) for evaluation. They confirmed that Mod-AugNet performed better than any other similar model without Prevention Module. They concluded that the test performance was entirely dependent on the prediction LSTM module.

Kim and Won [70] in their work proposed a new hybrid long short-term memory (LSTM) model with various Generalized Auto Regressive Conditional Heteroskedasticity (GARCH)-type models to improve the stock price prediction. They compared the performance of their system with other single models, like the GARCH, exponential GARCH, exponentially weighted moving average, a deep feed forward neural network (DFN), and the LSTM, as well as the hybrid DFN models combining a DFN with one GARCH-type model. The work concludes with the fact that it improve the overall prediction performance in stock market volatility.

Pang et al. [71], in their work proposed a LSTM based neural network with automatic encoder to predict the stock market price based on the concept of "stock vector" using deep learning. The experimental results depict that the deep LSTM with embedded layer outperforms for the Shanghai A-shares composite index.

Feng et al. [72], in their work, introduced a deep learning solution based on relational stock ranking model using temporal graph convolution for stock price pre-diction. The results are validated on the historical data of two stock markets, NYSE and NASDAQ. The experimental results showed that the model supersedes the state-of-the-art stock prediction solutions achieving an average return ratio of 98% and 71% on NYSE and NASDAQ, respectively.

JingyiShen et al. [27] proposed a model based on feature engineering and deep learning for predicting stock price. Their work includes pre-processing of the stock market dataset, utilization of multiple feature engineering techniques combined with a customized deep learning based system for stock market price trend prediction. The work proposed has performed well when compared with other machine learning models.

Kara et al. [24] in their work proposed two models based on ANN & SVM and compared their performance on Istanbul Stock Exchange. Ten technical indicators were selected as inputs of the proposed models. The experimental results showed that ANN performed better than SVM

Qiu and Song et al. [45] in their work proposed a model based on artificial neural network optimized using genetic algorithm to predict the direction of the Japanese stock market. The work categorizes the technical indicators of the stock market into two major input categories and predict the direction of the daily stock market index. The optimization of the ANN with the help of Genetic Algorithm has enhanced the performance of the system.

Guresen et al. [73] analyzed the effectiveness of various neural networks like multi-layer perceptron (MLP), dynamic artificial neural network (DAN2) and the hybrid neural networks which use generalized autoregressive conditional heteroskedasticity (GARCH) to extract new input variables on daily exchange rate values of NASDAQ Stock Exchange index. It was observed in their work that classical ANN model MLP outperforms DAN2 and GARCH-MLP.

Fischer et al. [25] in their work equated the performance of various deep learning techniques with long short-term memory for financial market predictions. They analyzed the performance of various LSTM model variants built upon daily close price, daily volume weighted average prices, weekly close price and Weekly volume weighted average prices and concluded that LSTM outperforms other deep learning networks and that it effectively extracts the useful information from the noisy financial time series data.

McNally S. et al. [23] have compared the performance accuracy of various deep learning algorithms(optimized RNN and LSTM) with ARIMA model for Bitcoin Price Index. Their major observation is that the deep learning algorithms have better accuracy than ARIMA model and also that LSTM takes longer time to train than RNN.

Seo M. et al. [74] Proposed an ANN based model containing multiple hidden layers combined with Google Domestic Trends (GDT). They concluded that with more number of hidden layers and varying activation functions, accuracy of model can be improved. This, in our view, may not be always true. They have claimed that the performance of the hybrid model with GDTs outperformed GARCH model and the model without GDTs.

2.2 DRAWBACKS AND LIMITATIONS OF THE EXISTING WORK

After going through the above work available in literature, following deficiencies have been identified which need to be taken care of:

- Most of the work available in the literature is concerned about the stock price prediction for a very short duration generally one day to one week. It may however be noted that in most of the cases the price variation during the day is not much (1% to 2%) unless some extra-ordinary news comes. So any prediction is likely to be quite accurate.
- While taking the inference from the historical data, the working pattern of all the papers is somewhat similar to that of the time series prediction wherein the basic underline philosophy is that the trend will continue. However, the stock market prediction is not a matter of mere time series trend. So there is a need to come out of this mindset and efforts are required to put some human elements.
- The assumption that the machine will be able to predict precisely is basically a wrong notion. The overall environment is a combination of multiple factors that has to be analyzed and evaluated in the holistic manner. The change of the technology from CNN to LSTM or their combination is not going to serve much of the purpose as the trends may not repeat.
- The prediction accuracy has not been that good in any case as it varies from 60% to 85%. An error to the tune of 15% to 20% is quite huge and can lead to mega loss in the financial market.
- Huge emphasis has been laid on the price prediction using machine learning [71], [75], [76], in the hope that the machine will be able to explore the underlying pattern in the data and predict the price for the short term.

- Normally there is not much variation in the stock price in a single day unless something exceptional happens. So all these predicted results are likely to be close to the actual price.
- If there is a sharp dip or rise in a single day, none of the above mentioned mechanism works.
- An investor in the stock market is there to stay for making the long term fortune. He needs a credible advice in this regard which can only come with the thorough analysis of the fundamentals of the stock under consideration.
- Conventional machine learning mechanisms suffer from the usual drawbacks of opacity, overfitting and short term prediction scenarios.
- Moreover, the random fluctuations in the stock price data which is a very common element, in the stock prices, is a big hindrance to the proper convergence.
- Most of the papers go for the few prominent stocks without taking care of the spectrum as a whole.

Keeping the above drawbacks and limitations of the available work, this thesis proposes multiple strategies for the predictive analytics in the stock market scenario that can help the investors for the long term continuous survival. These strategies and associated mechanisms have been discussed in next chapter onwards.

CHAPTER-III

REINFORCEMENT LEARNING BASED PREDICTIVE ANALYTICS FRAMEWORK FOR SURVIVAL IN STOCK MARKET

Many people believe that stock markets are essentially a random walk and it is a fool's game to try and predict them [77]. This view is based upon the fact that the stock markets are being governed not only by fundamentals but also by the human greed and manipulations. However, due to the involvement of huge amount of money, the research in stock market price and trend predictions has always been a challenging task and the efforts are being made for the same using the state-of-the-art techniques. This has led to moving from the conventional statistical analysis to the machine learning algorithms, but the challenge still continues and more and more solutions are being provided.

The conventional research in this domain, involved the prediction of future prices based upon conventional statistical models like ARMA [11], ARIMA [12] , ARCH [14], GARCH [15] and ES [16]. These models were primarily based upon linear time series process [78], [79]. These traditional techniques were unable to handle the noisy, non-linear, complex, dynamic, nonparametric, and chaotic nature of the stock market [13], [80].

This led to application of different machine learning techniques on the historical stock price data to predict the future trends of the stock. Atsalakis and Valavanis, [6], surveyed scientific articles that make use of neural and neuro-fuzzy techniques to solve stock market forecasting problem. Further the efforts were made to create hybrid models which combined different machine learning techniques [25], [32], [81] to attain better results. Availability of the Deep neural networks and variety of other machine learning techniques have created a belief that it is possible to accurately predict the stock price with the optimal choice of parameters relating to regression/classification algorithm, signal functions, number of hidden layers, number of cascading stages etc. [7], [71], [74]–[76], [82]–[90]. Expecting that a pure machine learning based model can predict the stock price movement is not appropriate. Stock markets don't follow any linear or nonlinear mathematical function/trend as such and are governed by the human manipulations and greed as well. Most of these research works predict the stock price on the short term basis (from one day to week or fortnight) with no long term goal as such. An investor is not there in the market for one day or one transaction, he/she is there to stay for making the gain on the continuous basis. No paper is taking care of this fact. The major focus of all the papers is on the price movement, no paper has gone for the holistic evaluation of the stock or market related situation for providing the investment advice. No paper has provided an automatable framework which can keep continuous track of stock market situation and generate advice on continuous basis. Thus, there is a need to design and create a framework that can holistically evaluate the current stock price position with respect to historical price data, ongoing company's performance, stock worthiness with respect to earnings and book value, stock market position and sentiments, sector sentiments and provide a credible advice for making the appropriate decision which may be buy/ sell/ hold (no action). The framework should be fully automatable and free from human bias. The work proposed here meets these goals and does the needful. Fig. 3.1 shows the philosophy behind the proposed framework.

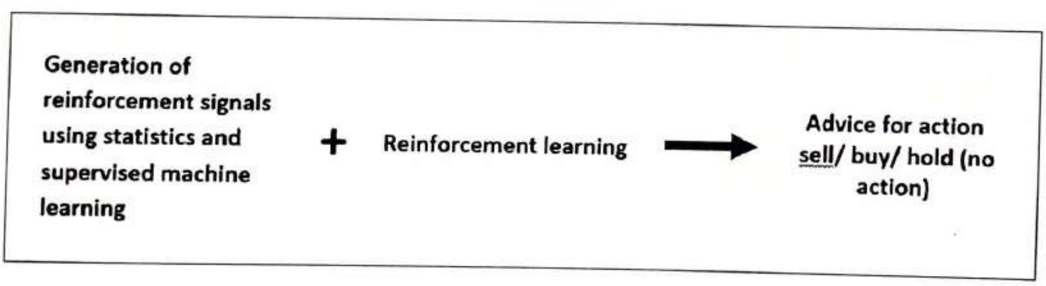


Fig. 3.1 Philosophy behind the proposed approach

3.1 PROBLEM DESCRIPTION AND OBJECTIVE

3.1.1 Problem Description

An investor in the stock market is not investing for a single day. He/she is likely to stay invested in the market looking for a continuous gain by sensing the opportunity as and when it arrives. Every stock in the market undergoes price variation which over a period of one year is normally 1.25 to 2 times of the lowest value. For example, ONGC a NIFTY50 stock in Indian stock market had an annual variation of 64.5 – 122.3 rupees making it 122.3/64.5=1.89 times. Thus there is a need to propose a system that holistically evaluates the current situation and generates advice that helps the investor in taking the rational decisions for surviving in the stock market significant gain with minimal risk.

3.1.2 Objective

To develop a framework to holistically evaluate the current stock position and to generate reinforcement signal which indicates the action to be taken (buy / sell / no action) with possible reward points.

3.2 OVERVIEW OF THE PROPOSED FRAMEWORK

Fig. 3.2 shows the basic design of the predictive analytics framework. The system begins with the identification and listing of possible reinforcement signals and their listing. Historical data is collected and partitioned into historical and future data. The historical data undergoes transformation and preparation to create the desired formats as per the requirement of intended reinforcement signals. Thereafter, normalization and cleansing procedures are applied on the data to ensure the parity of situation and removal of the misleading information. To assess the various reinforcement signal machine learning and statistical learning is applied. Impact of various reinforcement signals is studied on the future data and inferences are drawn. The various inferences drawn are used to acquire the knowledge which forms the basis for the design of predictive analytics framework based upon fuzzy logic. The results are reported and analyzed.

3.3 CREATING THE FRAMEWORK ELEMENTS

Before taking up the basic design of the proposed framework, let us have a look over reinforcement learning that is a key component of this framework.

3.3.1 Reinforcement Learning

Fig. 3.3 shows the basic reinforcement learning framework. Here an agent, at a given instant, existing in state S_t and having reward points as R_t takes an action A_t depending upon the current environmental situation. Consequent upon the action, the state of the agent changes to S_{t+1} and reward R is generated leading cumulative award to $R_{t+1} \leftarrow R_t + R$. The work, presented, uses a model free reinforcement learning algorithm known as Q learning wherein given an environment, the agent tries to learn a policy that maximizes the total reward it gets from the environment at the end of an experiment which involves multiple number of actions taken.

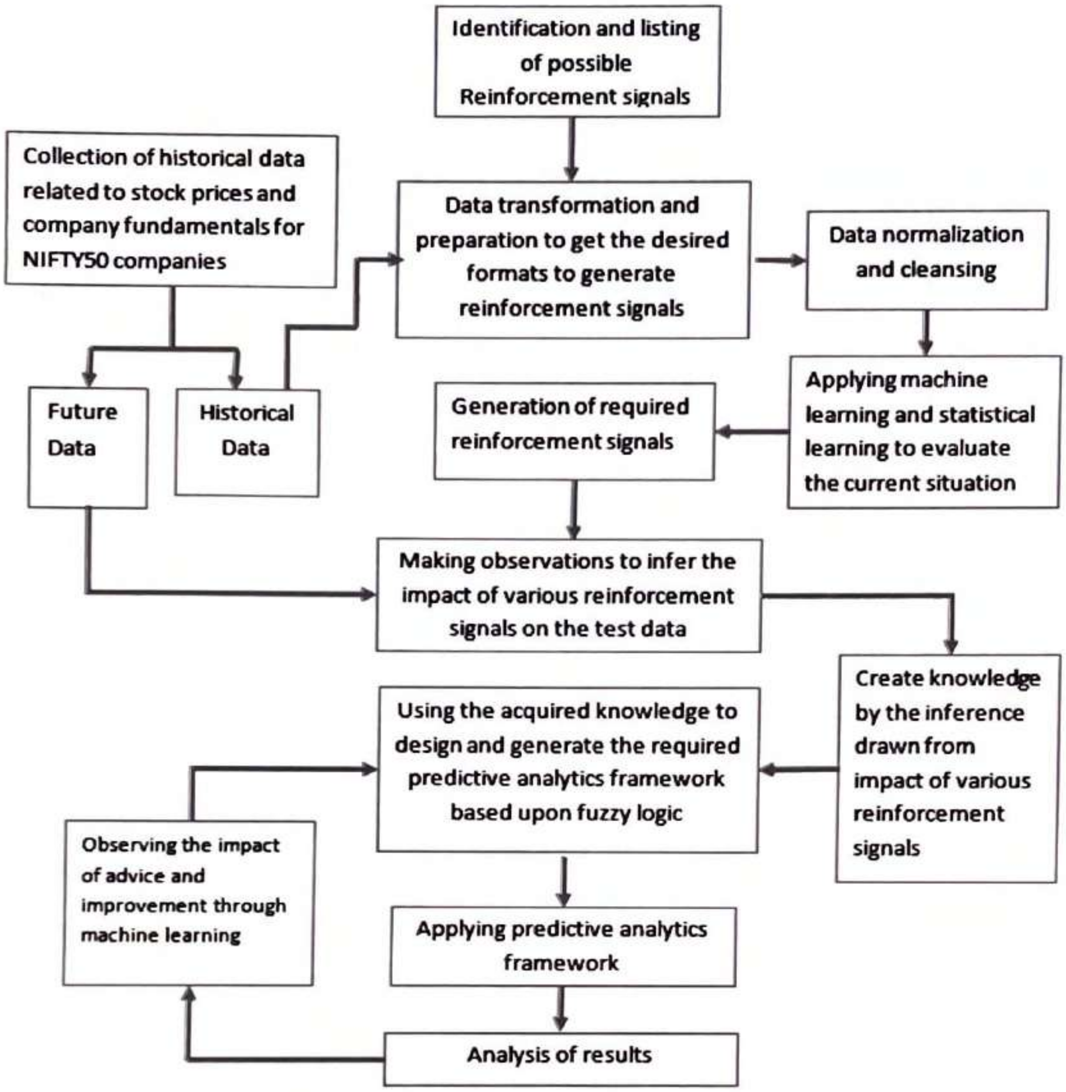


Fig. 3.2: Design of the Predictive Analytics Framework

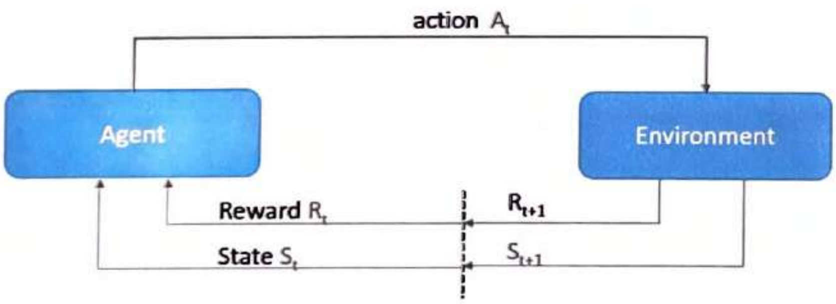


Fig. 3.3 Basic Reinforcement Learning Framework

For optimal performance, the agent must ideally follow the best possible option available to it. But if it is done, then it will be a somewhat mechanical approach and the worse thing would be that the agent won't learn the value of taking the other options available with it in that state. This dilemma is known as the exploration and exploitation trade-off. To come out of this dilemma and to be more humanistic it is advisable to go for greedy approach wherein other decision is taken other than recommended with small probability say 0.1 - 0.2. This will also add an element of conviction. This finally brings us to the Q-learning equation given by Eq. (3.1), which updates the action values of each state and action pair

$$\boxed{New\ Q(s,a) = Q(s,a) + \alpha[R(s,a) + \gamma \max Q'(s',a') - Q(s,a)]} \tag{3.1}$$

New Q value for the state and action is calculated as sum of Current Q value and learning rate,α times sum of reward function R and difference of discount rate (γ) times maximum expected future reward (Q') from current Q values.

3.3.2 Transformation of the data to create various reinforcement signals

As described earlier, the basis of the proposed framework is reinforcement learning wherein agent has to be rewarded / penalized for the actions taken by it on the basis of reinforcement signals. The various possible constituent of reinforcement signals for the reward function are follows:

- Current price sentiment of the stock in past three years context
- Trends in moving averages
- Off peak status of the price with respect to recent, mid-term and long term perspective
- Annual sales and profit growth performance of the company in past 3 years
- Price/ earnings and Price/Book Value ratio of the company
- Standing of the company with respect to its peers
- Sector market sentiments
- General market sentiments

We now go for the Experimental set up to generate the above mentioned reinforcement signals to examine their impact on the future stock prices. Basic details of the set up are shown in Table 3.1.

Table 3.1: Experiment Set Up	
Stock Market Under consideration	Indian Stock Market
Stock Exchange	NSE
Companies under consideration	All NIFTY50 companies
Current Date reference	31st March,2019
Past history period under review to generate reinforcement signals	1April 2016- 31 March 2019
Observing the impact of reinforcement signals on future data	1April 2019-31 March 2020
References for data collection	www.moneycontrol.com [19]
	www.money.rediff.com [17]
	www.nseindia.com [20]
	https://in.finance.yahoo.com [18]
	https://www.niftyindices.com/reports/historical-data [91]

3.3.2.1 Price sentiment of the stock with respect to past three years (PS3Y)

To create the PS3Y windows of past 7days, 15 days, 30 days, 3 Months, 6Months, 1year and 3 year prior to 31 March 2019 were used. To evaluate the standing of current price with respect to historical data a normalization process was used given by the Eq. (3.2):

$$\text{Normalized value} = 1 + 4 \times \frac{cp - \min}{\max - \min} \tag{3.2}$$

Here *cp* represents the current price, *min* and *max* represent the minimum and maximum price over the period under consideration. The normalization process generates a value in the interval [1, 5]. The output 1 indicates that the stock is trading at the lowest price of the period under consideration while an output 5 indicates that the stock is trading at the highest price of the period under consideration. Output near 3 indicates that stock is trading in the vicinity of the average price of the period under consideration. A value above 3 indicates the inclination of the price towards maximum side and vice versa.

To get the inference from this normalized data, single value decomposition process was applied. To obtain the single value, the machine learning in the form of multivariate linear regression was applied for which the feature vector contained normalized values obtained for various periods. A feature vector of the form [1, 1, 1, 1, 1, 1, 1] indicates extremely negative sentiments with all-time low price of the stock. It has been mapped to a target sentiment value of -100. Feature vector of the form [5, 5, 5, 5, 5, 5, 5] indicates extremely positive sentiments with all time high price of the stock. It has been mapped to a target sentiment value of +100. Feature vector of the form [3, 3, 3, 3, 3, 3, 3] indicates average price of the stock and has been mapped to a target sentiment value of 0 indicating a neutral sentiment. The system was trained using supervised learning through multivariate linear regression on the Weka Tool to create target sentiment value for a given novel price vector. Table 3.2 shows the price vector and corresponding price sentiment value of some NIFTY50 companies.

Table 3.2 Computation of price sentiment with respect to past 3 years (PS3Y)

Name of the Company	Period taken into consideration prior to March 31st, 2019							Price sentiment over 3 years (PS3Y)
	7 d	15d	30d	3 M	6M	1Y	3Y	
Eicher Motors Ltd.	1.33	1.15	2.20	2.54	2.08	1.52	1.65	-60.93
GAIL (India) Ltd.	1.82	1.70	1.99	3.74	3.27	3.00	4.02	-10.43
Grasim Industries Ltd.	4.85	4.86	4.89	4.96	3.03	2.59	2.05	44.50
HCL Tech. Ltd.	4.50	4.51	4.55	4.75	4.64	4.40	4.65	78.57
HDFC Bank Ltd.	4.40	4.45	4.84	4.86	4.91	4.91	4.97	88.14
Hero MotoCorp Ltd.	2.64	1.62	1.47	1.25	1.18	1.11	1.09	-76.00
Hindalco Ind. Ltd.	2.60	3.27	3.75	3.06	2.18	2.08	3.43	-4.50
Hindustan Uni. Ltd.	4.82	4.39	2.99	1.94	3.34	3.81	4.40	33.50
Housing Dev. Fin. Corp. Ltd.	4.36	3.34	3.99	3.96	4.45	4.17	4.66	56.64
ICICI Bank Ltd.	4.23	4.23	4.69	4.77	4.86	4.89	4.93	82.86
Indian Oil Cor. Ltd.	2.99	3.12	3.90	4.44	4.57	4.02	2.94	35.57
IndusInd Bank Ltd.	3.70	3.70	4.36	4.47	4.58	3.55	4.07	53.07
Infosys Ltd.	4.24	4.49	4.49	4.04	4.32	4.48	4.66	69.43
ITC Ltd.	2.87	2.89	4.07	4.27	4.36	3.45	3.49	31.43
JSW Steel Ltd.	4.06	4.06	4.06	3.71	2.11	1.85	1.42	1.93

3.3.2.2 Trends in moving averages

To explore the price momentum, the conventional concept of moving averages was used. For this purpose, moving average of the stock prices was computed for the period of 5, 10, 30, 50, 100 and 200 days. To ensure the peer comparison between various companies, these moving averages were normalized by fixing the 200 days average price to 100. The reinforcement signal obtained from moving averages has been over the following aspects:

- Overall gain/loss in momentum
- Type of momentum

The overall gain/loss in the momentum is the simple difference between the normalized 5 days moving average and the normalized 200 days moving average. A positive/negative result indicates the overall gain/loss of price strength. Type of the momentum can be monotonically increasing (MI), Generally increasing (GI), fluctuating up (FU), fluctuating down (FD), Generally Decreasing (GD) and monotonically decreasing (MD). Fig. 3.4 shows the algorithm for finding the momentum. Table 3.3 shows the type of momentum and the overall gain/loss in NMVA for some of the NIFTY50 companies.

```
For each company in NIFTY50
Begin
        Extract normalized MVA for (5, 10, 30, 50,100,200) days
        Count = Number of positive differences in consecutive entries
        Switch (count)
                case 5 : TypeofMom="Monotonically Increasing"
                case 4 : TypeofMom="Generally Increasing"
                case 3 : TypeofMom="Fluctuating Up"
                case 2 : TypeofMom="Fluctuating Down"
                case 1 : TypeofMom="Generally Decreasing"
                case 0 : TypeofMom="Monotonically Decreasing"
End
```

Fig.3.4 Algorithm for finding the type of momentum

3.3.2.3 Off-Peak Analysis

To survive in the stock market, one should have the idea of possible upward movement for a particular stock. Though it is not possible to forecast precisely, but the past track record can be a guide in this matter.

Table 3.3 Computation of Type of Momentum

Name of the Company	Number of days used for computing the moving Average						Overall Gain/ loss in NMVA	Type of the momentum
	5	10	30	50	100	200		
Eicher Motors Ltd.	86.41	88.26	87.16	85.52	89.50	100	-13.59	FD
GAIL (India) Ltd.	100.59	100.97	97.38	95.97	97.57	100	0.59	FD
Grasim Industries Ltd.	93.36	92.63	88.86	86.90	90.11	100	-6.64	FU
HCL Technologies Ltd.	105.02	103.81	104.04	102.71	100.32	100	5.02	GI
HDFC Bank Ltd.	110.43	109.69	104.37	103.10	101.26	100	10.43	MI
Hero MotoCorp Ltd.	83.80	85.27	87.78	89.18	94.34	100	-16.20	MD

Hindalco Ind. Ltd.	93.86	93.31	90.69	91.78	96.48	100	-6.14	FD
Hindustan Uni. Ltd.	98.94	99.14	101.13	102.42	103.04	100	-1.06	GD
Housing Dev. Fin. Corp. Ltd.	102.47	103.28	100.61	101.64	101.52	100	2.47	FU
ICICI Bank Ltd.	117.48	117.67	110.64	109.10	108.25	100	17.48	GI
Indian Oil Corp. Ltd.	110.19	109.16	100.29	97.05	95.03	100	10.19	GI
IndusInd Bank Ltd.	103.83	102.40	93.87	91.69	91.92	100	3.83	FU
Infosys Ltd.	106.12	105.52	105.47	106.32	101.38	100	6.12	GI
ITC Ltd.	103.71	103.73	100.54	99.63	99.10	100	3.71	FU

In the off-peak analysis, we have tried to check, how far the current price *cp* is from the local and the global peak. To find out the off-peak values, we used Eq (3.3)

$$off\_peak\% = 100 * \frac{cp}{max} - 100 \tag{3.3}$$

Here *max* represents the maximum value in the period under consideration. For example, a current price of 96 with the local peak for the period as 105 will be off-peak by -8.57%. The off-peak analysis helps in identifying the Short term local fluctuations to make a decision for the short term gain. The reinforcement signals drawn from the off peak analysis are:

- High recent fluctuation (*yes/no*). Yes, if more than 6%.
- Max off peak

The off-peak analysis of some NIFTY 50 companies has been shown in Table 3.4.

Table 3.4 Off peak status of NIFTY 50 stocks as on 31/3/2019

Period→	7 d	15d	30d	3M	6M	1Y	3Y
Perspective→	Recent		Short Term		Mid term		Long term
Eicher Motors Ltd.	-4.60	-9.84	-12.01	-12.01	-18.82	-36.24	-38.66
GAIL (India) Ltd.	-4.33	-5.21	-5.21	-5.21	-10.16	-12.91	-12.91
Grasim Industries Ltd	-0.22	-0.22	-0.22	-0.22	-16.07	-23.03	-35.56
HCL Technologies Ltd.	-1.00	-1.00	-1.00	-1.00	-1.51	-3.25	-3.25
HDFC Bank Ltd.	-0.45	-0.45	-0.45	-0.45	-0.45	-0.45	-0.45
Hero MotoCorp Ltd.	-2.10	-7.54	-10.03	-18.49	-24.46	-33.84	-39.16
Hindalco Industries Ltd.	-2.47	-2.79	-2.79	-9.50	-20.93	-23.19	-27.68

Hindustan Unilever Ltd.	-0.13	-0.49	-2.69	-8.18	-8.70	-8.70	-8.70
Housing Dev. Fin. Corp.Ltd.	-0.58	-2.13	-2.13	-2.54	-2.54	-4.08	-4.08
ICICI Bank Ltd.	-0.96	-0.96	-0.96	-0.96	-0.96	-0.96	-0.96
Indian Oil Corp. Ltd.	-4.01	-4.01	-4.01	-4.01	-4.01	-10.23	-29.66
IndusInd Bank Ltd.	-2.88	-2.88	-2.88	-2.88	-2.88	-12.55	-12.55
Infosys Ltd.	-0.72	-0.72	-0.72	-3.74	-3.74	-3.74	-3.74
ITC Ltd.	-2.06	-2.06	-2.06	-2.06	-2.06	-8.06	-16.00
JSW Steel Ltd.	-1.73	-1.73	-1.73	-5.52	-24.32	-31.47	-83.27

3.3.2.4 Sales Growth and Gross Profit Growth

The sales and the gross profit data for all the NIFTY50 companies was collected for the years ending on 31st March 2016, 2017, 2018 and 2019. The impact of sales growth and gross profit growth as reinforcement signal was studied with their values at 15% and 25% respectively.

Data cleansing operation was required when a loss (say -30%) got converted to profit (say 50%) and the same was shown as the negative growth due to sign reversal. Data was cleansed by making the appropriate amendments.

3.3.2.5 Price/Earning (P/E) and Price/Book Value (P/B)

A P/E ratio (say under 10) and P/B ratio (under 3) is considered to be healthy. Impact of the healthy P/E and P/B reinforcement signal was studied on the price performance of the stock.

Data cleansing operation was required when a negative earning created the negative P/E ratio making it best due to lowest status. Data was then cleansed by making the appropriate amendments.

3.3.2.6 Sector Sentiments

Share prices of a company can go up/down due to overall strong/weak performance of the sector. For example, share prices of oil companies took a big toll during COVID 19 lockdown period in 2020 due to very low consumption, storage issues and extremely low crude prices. Similarly, the stock prices of the Pharma companies showed significant rise in this period due to increased consumption of the drugs and vaccines. The NIFTY index in India has many sector segments such as Automobile, Bank, IT, Pharma, Finance, FMCG, Media, Metal etc.

Table 3.5 shows sector sentiments of some sectors as on 31st March 2019. The sentiment calculation process is similar to that of the PS3Y wherein the sector index has been used instead of the stock price. It shows that sentiment for the financial services which is almost on the peak. The Pharma segment is showing better sentiments in the recent history than the past history. Private banks are also having quite good sentiment. FMCG is having slightly more than average sentiment.

Table 3.5 Sector Sentiments as on 31st March,2019

Period→ / Sector ↓	7 d	15 d	30 d	3M	6M	1Y	3Y	Sector Sentiments (SS)
FMCG	4.25	3.68	4.05	3.66	4.29	3.38	4.19	46.42
Pvt. Banks	4.71	4.73	4.89	4.91	4.94	4.94	4.97	93.5
Pharma	4.72	4.72	4.83	4.93	3.32	3.00	2.40	49.4
Fin. Services	4.93	4.93	4.97	4.98	4.99	4.99	4.99	98.4

Table 3.6 shows the moving average trend for the sectors taken in Table 3.5. The Table shows the best Normalized Moving Average (NMVA) for the Private banks with Generally i)ncreasing trend. NMVA for the Financial services is also very good with Monotonically increasing trend. The Pharma segment is showing a downfall in NMVA in Generally Decreasing trend. FMCG is having a steady neutral trend.

Table 3.6 Sector Moving Average Trend (SMVA)

Sector	% Gain in NMVA	Trend
FMCG	0.63%	GI
Pvt. Banks	12.09%	MI
Pharma	-0.51%	GD
Fin. Services	9.41%	MI

3.3.2.7 Standing of the company with respect to its peers

Identification of peer group performance can be of great help. While a sector performance can be a generic indicator, the peer group performance can act as a specific indicator. A comparison of current standing w.r.t. peer group can be of great help.

- A low price in line with the peer group and sector performance will indicate that stock can be held in the hands of the investor until the situation improves.
- A low price not in line with the peer group and sector performance can be indicative of two situations
  - Something is ailing with the stock and if it is so stock can be parted with to avoid further loss. The ailing situation can be checked with the company fundamentals.

- If company fundamentals are strong then the situation is conducive for making the investment through buying.

- A high price not in line with the peer group and sector performance will indicate that stock can be sold for making short term profit.
- A high price in line with the peer group and sector performance will indicate that stock can be stayed with for making long profit.

For assessing the situation w.r.t. peers and the sector, a table was created for some private using 52 week low and 52 week high values. Table 3.7 shows the normalized values calculated as Eq 3.2

Table 3.7 Comparison of stock prices of various private banks as on 31st March, 2019

	HDFC	Kotak Bank	ICICI bank	AXIS Bank	IndusInd bank
52 week low	940.13	1002.3	256.5	477.5	1333.9
52 week High	1163.5	1424	402.7	788.55	2037.9
Current Value	1158.25	1335.75	398.85	776.1	1782.1
Price Sentiment	4.91	4.16	4.89	4.84	3.55
Type of NMVA	MI	FU	GI	MI	FU
Change in NMVA	10.43	6.49	17.48	22.76	3.83

Table 3.7 indicates that IndusInd bank is unable to perform at par with its peer group and needs to be parted with

3.3.2.8 General Stock Market Sentiments

To assess the market sentiments on the current date, historical data of various stock indices was taken and analyzed in the fashion similar to the price sentiment. It was observed the market sentiments were highly positive with all the three indices having the sentiment value more than 90 as shown in Table 3.8.

Table 3.8 Standing of Market with respect to past history

	7d	15d	30d	3M	6M	1Y	3Y	Index Sentiment (IS)
NIFTY 500	4.95	4.95	4.98	4.98	4.99	4.08	4.59	89.42
NIFTY50	4.92	4.92	4.97	4.98	4.98	4.69	4.87	95.21
MIDCAP150	4.92	4.92	4.96	4.97	4.98	3.36	3.84	78.21
SMALLCAP250	4.86	4.86	4.95	4.96	4.96	2.48	2.60	61.92

The moving average trend for these indices also showed a gain as shown in Table 3.9. The Gain in NIFTY50 was the highest followed by NIFTY500 and MIDCAP150 thereafter indicating the highest support for the blue-chip shares. Also the type of the moving average (Generally Increasing) was best in the NIFTY50 indicating a strong momentum for the blue chip shares and a strong but positive momentum for the Midcap shares.

Table 3.9 Moving Average Trend for various stock indices

Index	Gain in NMVA	Trend
NIFTY 500	36.2 (3.62%)	FU
NIFTY50	51.9 (5.19%)	GI
MIDCAP150	11.7 (1.17%)	FD
SMALLCAP250	-10.8(-1.08%)	FD

3.4 OBSERVATIONS AND INFERENCES

3.4.1 Observations

After generating the various reinforcement signals from the historical data (1/4/2016-31/3/2019) their impact was studied on the future price during the period (1/4/2019-18/9/2020) and following observations were made.

1. In addition to the merit, the shares in the stock market are also governed by the sentiments. A very low sentiment wave can swipe the whole of the stock market spectrum. It can be seen that in the period from 1st Feb. 2020 to 31st March, 2020, shares of 47 NIFTY50 companies (out of 50) showed a significant fall ranging from 3.33% to 76.75% with the average loss of 31%. Only 3 companies namely Hindustan Unilever, Nestle and Dr. Reddy were the exceptions. 16 companies lost their share values by more than 40% just owing to the sentiment. It was also observed that when the sentiment wave became positive most of the stocks recovered. Fig. 3.5 and Fig. 3.6 show the investment status on the quarterly basis when
   a. Equal investment of Rs. 100000/- was made in each NIFTY50 stock on 31st March 2019. In this case, invested capital reduced to 71% on 31st 2010 and recovered back to 97.8% level on 18 Sep., 2020.
   b. Equal numbers of stocks were bought for each NIFTY50 stock on 31st March 2019. In this case, invested capital reduced to 87.2 % on 31st 2010 and recovered back to 110.6 % level on 18 Sep., 2020.

With no major change in performance of the companies recovery process is merely due to sentiments.

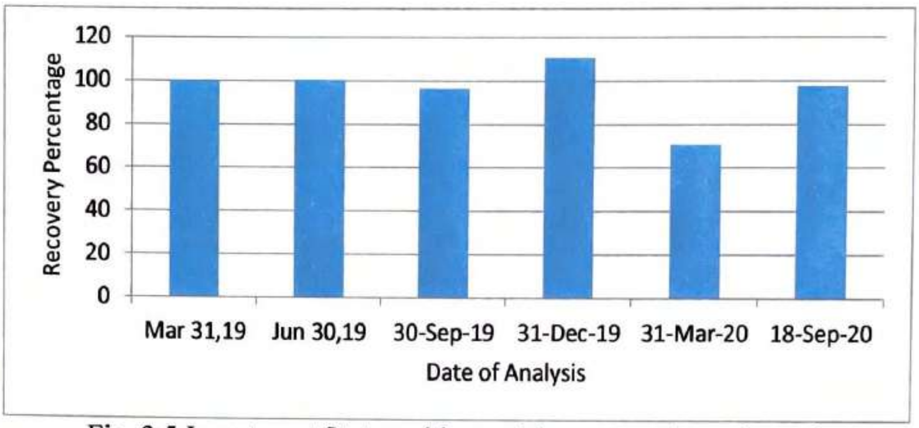


Fig. 3.5 Investment Status with equal investment in each stock

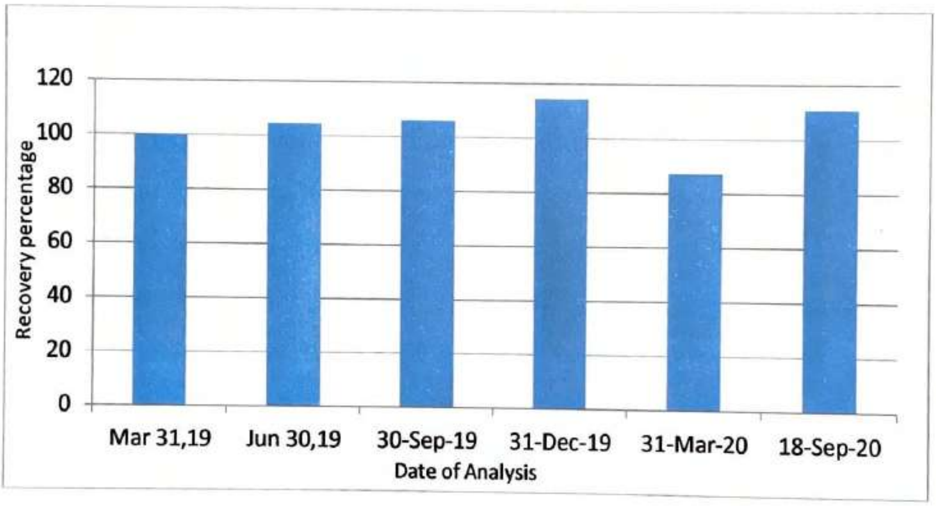


Fig. 3.6 Money recovery percentage with equal number of each stock

Fig. 3.7 depicts the investment recovery percentage after the COVID 19 period by the NIFTY50 companies on 18 Sept. 2020 for the investment made on 31st March 2019. It shows that 32% (16) of Nifty companies recovered to more than 100% w.r.t. 31st March 2019 in Sep. 2020. The downfall in the Feb-May,2020 in the stock has shown the recovery in the next 3-4 months and there has been similar situations in past also e.g. in 2008. So, a very low sentiment time in the market is a good time for long term investment.

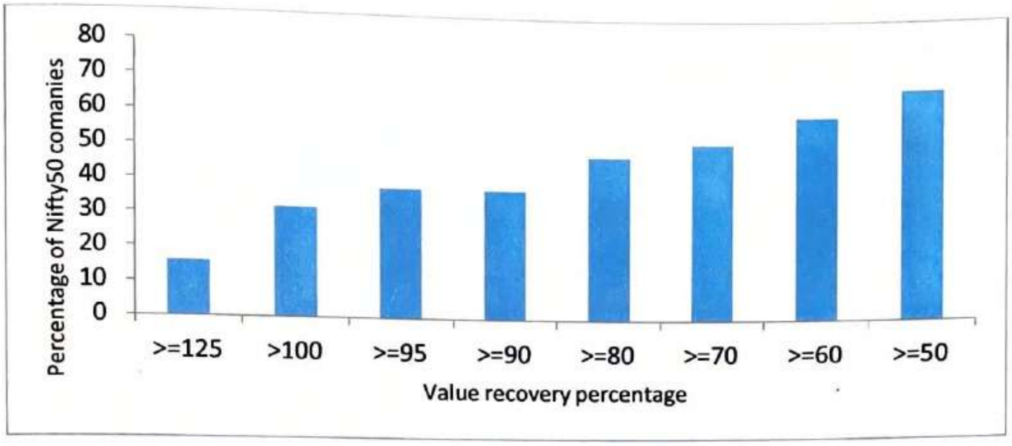


Fig. 3.7 Value recovery percentage by NIFTY50 Companies

2. To check the effect of the company's performance on the stock, the performance was measured using multiple scenarios, with details as follows :

a. All the investments were made on 31st March 2019.

b. In scenario1, on each of the NIFTY50 stocks, a notional investment of Rs 1,00,000 was made amounting to Rs.50,00,000.

c. In scenario2, 300 units of each of the NIFTY50 stocks were notionally purchased amounting to Rs.2,71,94,439.

d. In scenario3, only those NIFTY50 stocks were invested (300 units each), which were having the sales and gross profit growth of more than 15% amounting to Rs. 24669615.

e. In scenario4, only those NIFTY50 stocks were invested (300 units each), which were having the sales and gross profit growth of more than 25% amounting to Rs. 20790825.

f. In scenario5, proportionate investment was made in the NIFTY50 stocks on the basis of their sales and gross profit growth. Larger the sales and gross profit, more is the investment. For the stocks where either the sales growth or the profit growth was negative, no investment was made. The total amount invested was Rs.5177640.

The performance was measured on the quarterly basis in 19-20. To visualize the effect of the recovery, the performance was last measured on 18th September, 2020. The graph in Fig 3.8 shows the status of the investments on various dates with the investment on the 31st March 2019 being normalized to 100 as the base value. The various observations are as follows:

The fall on 31st March 2020, was maximum in case of scenario1. The stock values reduced by nearly 30%. The fall was much lesser (nearly 13%) in case of scenario2 which indicates that the loss in the expensive stocks was much lesser. When the investment was made in case of scenario3, the fall was much lesser (nearly 6.5%) showing that merit of the stock counts in the days of bad sentiments. When the investment was made in case of scenario4, the fall was slightly more than the scenario3 (nearly 9.85%). However, the gain was maximum on all the observation dates except 31st March, 2020 again showing that merit counts. In case of proportionate investment, the gain was not significant and was at par with other scenarios. However, the loss was more (nearly 17%) on 31st March, 2020. These observations indicate that merit of the stock counts and hence the stocks with the history of sales growth and the gross profit growth in the past are safer. Also the more expensive stocks are safer than their low value counterparts.

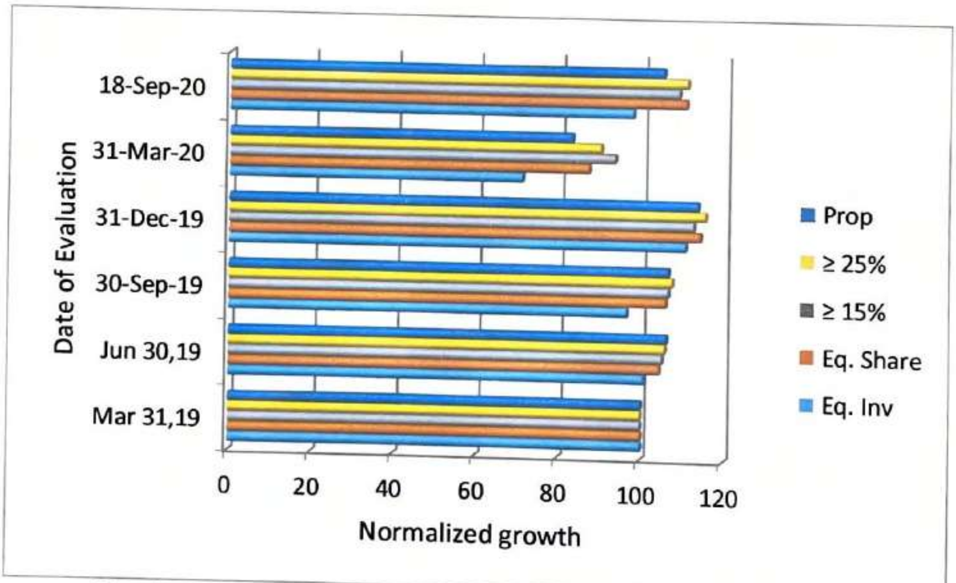


Fig. 3.8 Money Growth with time with different type of investments

3. The observation about the companies which made the significant gain with respect to 31st March, 2019 as the base, we took a look on the NIFTY50 companies which registered a stock price growth of more than 10%, 15% and 20%. Their number came out to be 34, 21 and 15 respectively. The in-depth analysis showed that the companies with moving average type as Monotonically Increasing, Generally Increasing and Fluctuating Up average had quite

significant advantage over other categories. It was also observed that the companies which showed significant price growth were having the positive price sentiment. The various graphs related to this aspect have been shown from Fig 3.9 to 3.11. It was also observed that amongst the 15 companies 10 were 10% below their 3 year peak. This indicates that companies having increasing type of moving average and near their peak maintained their momentum.

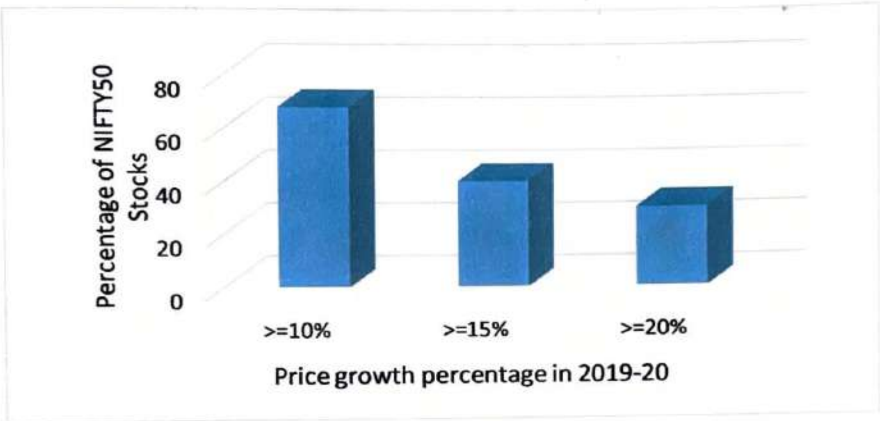


Fig. 3.9 Percentage of NIFTY50 stocks with price growth

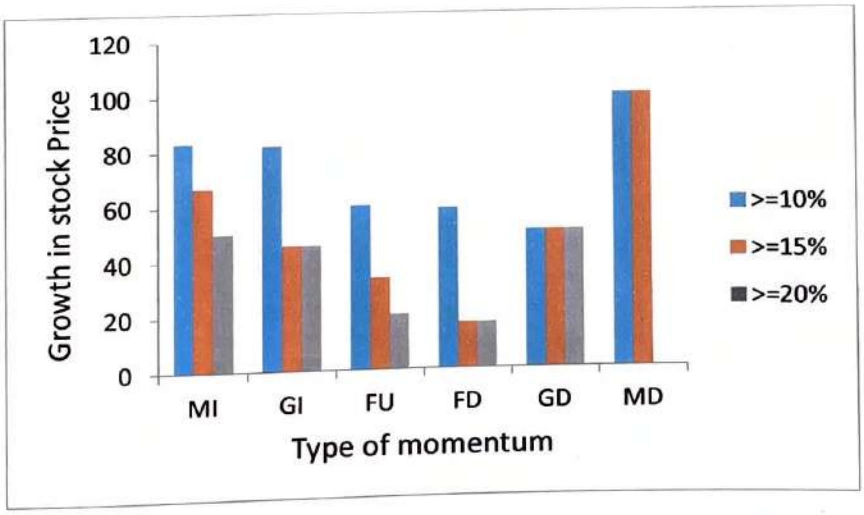


Fig 3.10 MVA momentum type Vs Stock Price Growth

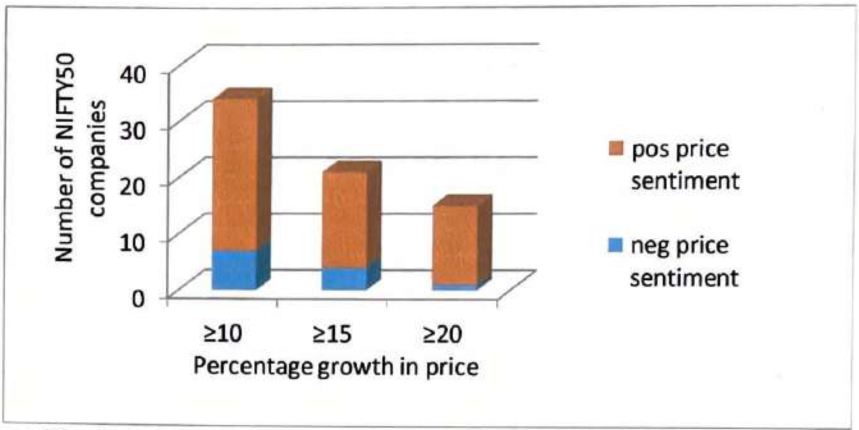


Fig 3.11 Effect of Historic Price Sentiment on Price Growth

In the off-peak analysis, it was observed that a high fluctuation (>6%) in the short term period context (up to 30 days) can be used for gain in short term period. The analysis showed that out of 8, 7 companies exhibited this feature. The experiment was repeated 5 times at an interval of 2 months wherein each time it was found to be a consistent feature. The Fig 3.12 shows the details of the experiment. Also it was observed that long term off peak reference (3 years), if large (more than 30%), did not help in investing as the previous levels were not touched again in almost all the cases.

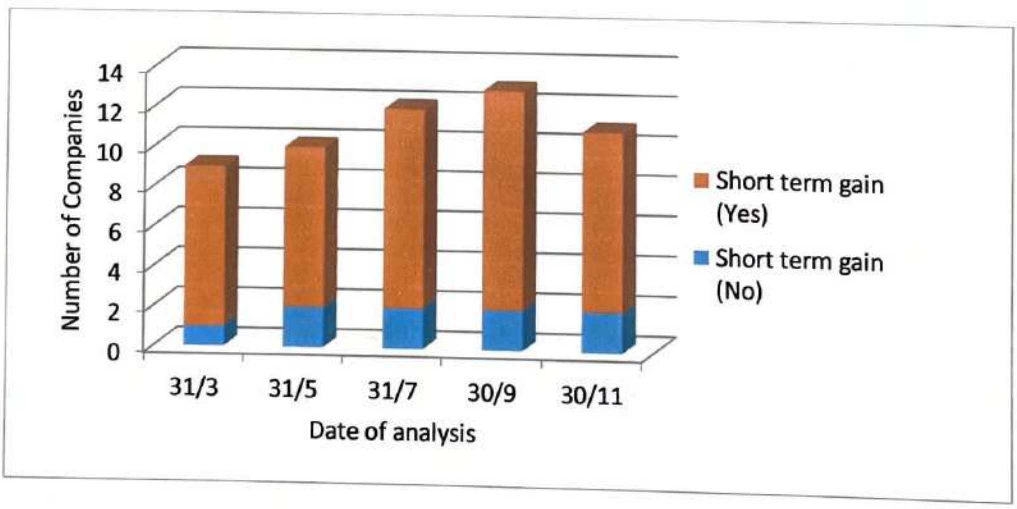


Fig. 3.12 No of Companies showing high short term fluctuation

4. The experiment was also conducted on the low valued safe shares having the P/E ratio below 10 and the P/B value under 5. For each of these companies (11 numbers) 300 shares were invested on 31st March, 2019. The total investment was Rs. 59,99,190. The investment status on various dates is with normalized

value on 31st March, 2019 at 100 is as shown in the Fig. 3.13. The results have been compared with global investments wherein the investment was made in all NIFTY50 was made with 300 stocks each. The comparison shows that there was no significant difference in recovery between the global investment and the safe investment. However, at the fall on 31st March, 2020 these safe shares proved better had comparatively much less fall compared to global investment indicating that safe companies are good for long term investment.

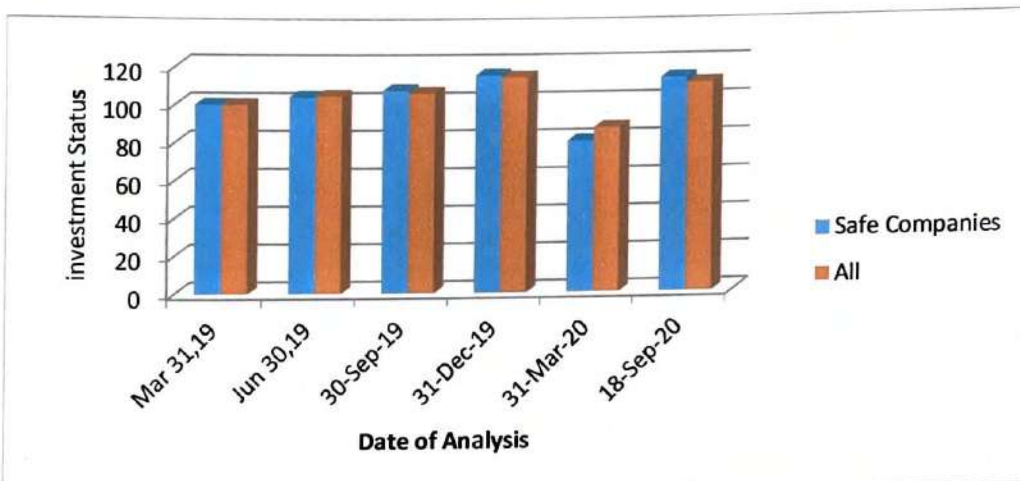


Fig 3.13. Comparision between global and safe investment

5. The Fig. 3.14 shows the status of the investments made in NIFTY50 companies on 31st March 2017.The investment takes two aspects into consideration, equal investment and equal stocks. In case of equal stocks (ES), 300 stocks of each NIFTY50 company were purchased on 31st March, 2017 while in case of equal investment (EI) Rs. 100000 was invested in each stock on the date. The Fig.3.15 shows the investment status checked on different intervals. The observations made are as follows:
    a. In the span of 3 years nearly 40% gain (excluding dividend) was available to the investor.
    b. ES has generally outperformed the EI indicating that high value stocks are more reliable
    c. Even at the time of distress (March 2020) money was recoverable to a good extent 92% in case of EI and 108% in case of ES.

On the change of sentiments to positive, recovery was good and gain was significant (more than bank interest etc.).

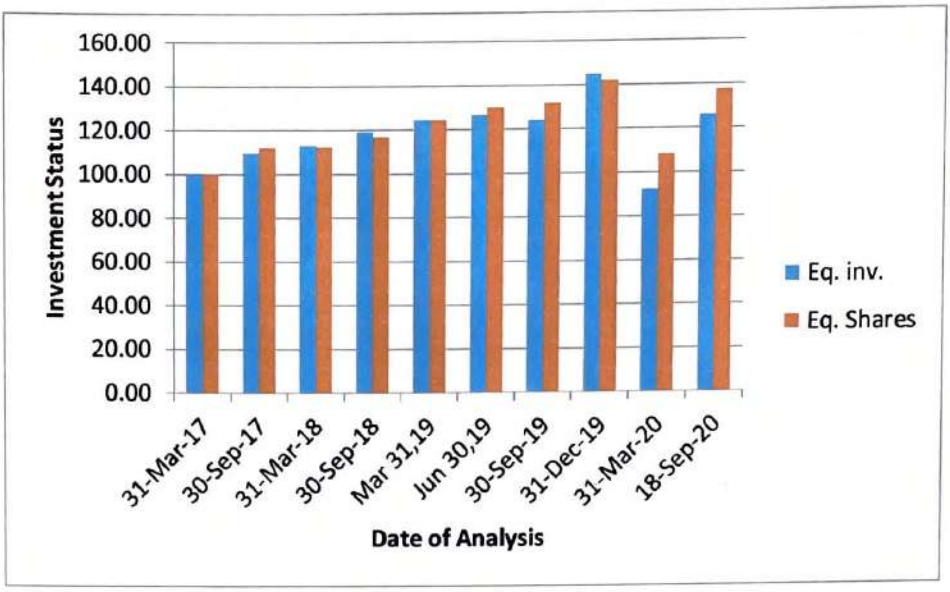


Fig. 3.14 Long Term Investment Status

6. After considering the sentiments and moving average various sectors in section 3.2.2.6, it was observed that positive momentum became stagnant in case of FMCG and Pvt. Banks sectors. Financial Service sectors showed a significant improvement and the downtrend continued in Pharma sectors.

7. After studying the various banking sector stocks in section 3.2.2.7, it was observed that positive momentum was maintained by all the banks. Kotak had a much significant gain due to lesser off peak. However, Indusind Bank lost the momentum due to news about its exposure to vulnerable sectors.

8. While observing sentiments and moving average various indices in section 3.2.2.8, it was observed that the GI trend in NIFTY became almost stagnant, fluctuating up trend in NIFTY50 resulted in a small drop of 1-2%, fluctuating decreasing momentum in MIDCAP and SMALLCAP indices resulted in significant loss in their moving average 5-6% in MIDCAP and 10% in SMALLCAP.

3.4.2 Inferences

After making the above observations, following inferences were drawn to develop the reward system of reinforcement learning.

- The blue chip Stocks in the market are suitable for investment than the mid-cap and the small-cap stocks.
- The share market is governed by the sentiments and is likely to recover when the sentiments improve. Thus there is no need for panic selling in the days of low sentiments.
- The positive price sentiment and the increasing moving averages provide a conducive environment for short-term and mid-term investment.
- A highly fluctuating moving average on the short-term basis provides an opportunity for making significant short-term gains.
- For the long-term investments, a company with strong fundamentals in terms of sales and gross profit growth is relatively more suitable.
- P/E ratio and P/B ratio do not have much impact on the performance of the stock but are helpful in deciding for long term investment.
- Among the blue chip stocks, the stocks with higher values are more reliable.
- Since the prices keep on fluctuating for one reason or the other, a regular exit at the local peaks can help in increasing gains.

3.5 IMPLEMENTATION OF THE PROPOSED PREDICTIVE ANALYTICS FRAMEWORK

As described earlier and shown in the Fig. 3.3, a reinforcement learning framework is based upon an entity called *agent,* a set of states *S*, a set of actions *A* and the rewards values. At a given instant, *agent* can exist in one of the state $s_i \in$ S and can move to the other state $s_j \in$ S depending upon the action $a_k \in A$ taken by it. For every action taken by the agent, a reward $r \in [0,100]$ is generated that indicates the appropriateness of the action taken by the agent. In our framework, an agent

- can exist in 8 states of holding, *S*={H0,H1,H2,H3,H4,H5,H6,H7} where H0 represent state of no holding (say 0 units) and H7 represents the state of maximum holding (say 700 units).
- can take 7 types of actions, *A* = {Sell00, Sell200, Sell300, No action, Buy100, Buy200, Buy300} which represent the selling/buying of 100, 200, 300, 0/0,100,200,300 units of stocks respectively.
- can transit from one state to other by either selling or buying action.

- continues in its current state, if no action is done.
- gets the reward, in the range [0,100], based upon the quality of action in the current state environment.

The limit of 700 is notional which can be scaled as per the status of the investor. Limit serves the two purposes

- Gives caution against the overexposure to a particular stock.
- Helps in generating the diversity in the portfolio.

The selling or buying limit of 300 has been imposed keeping in view that all the stocks are not sold/bought in the heat of the moment. Table 3.10 shows the state transition table for the agent.

Framework design task was divided into 4 phases: Assessment phase, Computation and training phase, Design phase and Learning phase.

Table 3.10 Transition Table for the various actions taken by the agent

Action→ State↓	Sell100	Sell200	Sell 300	No action	Buy100	Buy200	Buy300
H0	X	X	X	H0	H1	H2	H3
H1	H0	X	X	H1	H2	H3	H4
H2	H1	H0	X	H2	H3	H4	H5
H3	H2	H1	H0	H3	H4	H5	H6
H4	H3	H2	H1	H4	H5	H6	H7
H5	H4	H3	H2	H5	H6	H7	X
H6	H5	H4	H3	H6	H7	X	X
H7	H6	H5	H4	H7	X	X	X

In the assessment phase, on the basis of observations made and inferences drawn in the previous sections, different reinforcement signals were partitioned into different segments on the basis of their values/types. The details of the process have been shown in Table 3.11. Further for different types of investment perspectives namely: *Short Term*, *Mid Term* and *Long Term* various reinforcement signals were assigned the impact value on the basis of the observations and inferences made. Table 3.12 shows these values in the form of linguistic variables. Table 3.13 shows the numerical weight factors corresponding to the linguistic entries in Table 3.12.

In the computation phase, two vectors are taken.

- Weight vector from Table 3.13 that depends upon the investment perspective.
- Points vector generated for the various reinforcement signals depending upon their value/type taken from Table 3.11

Total weighted score was computed by the dot product of the above two vectors. In the training phase, this weighted score forms the basis for the regression based supervised machine learning. The supervised learning used in the design of the framework involves the group of reinforcement signals used in Table 3.11 to act as feature vector with some target values. The empirical weights used in Table 3.13 help us assigning the target values for different feature vectors. The assignment of target values in such a manner help us in incorporating the observation, inference and conviction.

Since the above point calculation mechanism cannot be precise, therefore in the design phase, we adopted a soft computing tool that naturally handles this type of the uncertainty or imprecision. The tool is *fuzzy sets and logic.* In this phase, overall weighted score was classified into different fuzzy sets, namely: *Buy300, Buy200, Buy100, No-action, Sell100, Sell200, and Sell300* which represent the action to be taken by the agent. The reward points for an action are created through its membership of the corresponding fuzzy set. Normally, the fuzzy membership is in the range [0, 1] but for the sake of easier conveying sense the reward value was generated in the range [0,100]. Fig.3.15 shows the reward generation scheme corresponding to various actions performed by the agent the Table 3.10 on the basis of weighted score. The advantage of fuzzy logic approach is that it takes into consideration the overall prospect and naturally nullifies the minor adjustments. If the weighted score is above 110, then the corresponding action is *Buy300* with 100 reward points. If the score is 75, then there are 2 possible actions *Buy200* and *Buy100*, with the reward points as 40 and 60 respectively. If the score is only 0, then the corresponding actions are *Sell100* and *Sell200* with reward points as 20 and 80 respectively. All these actions are subject to the bounds imposed in Table 3.10 which relate to the possibility in the current situation as well as the saturation.

Table 3.11

Point allocation to various categories of reinforcement signal types					
Reinforcement Signal	Value/Type	Points	Reinforcement Signal	Value/Type	Points
Company's Av. Sales growth in past 2 Quarters (QSG)	Above 20%	+3	Stock Price Sentiment (PS3Y)	>75	7
	10-20%	+2		50-75	5
	0-10%	+1		20-50	3
	0 to -10%	-1		-20 to 20	0
	<-10%	-2		-20 to -50	-3
Sector MVA (SMVA)	-10 to -20%	-1		-50 to -100	-5
	GI	+2	Stock Price Moving average type (SPMVA)	MI	+5
	FU	+1		GI	+3
	FD	-1		FU	+1
	GD	-2		FD	-1
	MD	-3		GD	-3
Sector Sentiment (SS)	>75	+5		MD	-5
	50-75	+3	Company's Av. Sales growth in past 3 years (CASG)	Above 15%	+5
	20-50	+1		Above 10%	+3
	-20 to 20	0		5-10%	+1
	-50 to -20	-1		-5 to 5%	0
	-75 to -50	-3		-10 to -5%	-1
	-100 to -75	-5		-10 to -20%	-3
Company's Av. Profit growth in past 3 years (CAV)	Above 50%	+5		< -20	-5
	Above 20%	+3	P/E	<=5	+3
	10-20%	+2		5-10	+2
	0-10%	+1		10-25	+1
	0 to -10%	0		>25	0
	-10 to -20%	-1	P/B	< 0	-1
	-20 to -50%	-2		< 3	+2
	Less than -50%	-3		3-5	+1
Short term fluctuation in Stock Price from (STF)	>10%	5		5-10	0
	6-10%			Above 10	-1
	3-6 %	3	News	Highly positive	+5
	0 to 3%	1		Positive	+3
				Negative	-3
				Highly negative	-5
Company's Av. Profit growth in past 2 Quarters (QPG)	Above 30%	+3	IS	>75	+5
	15-30%	+2		50-75	+3
	0-15%	+1		20-50	+1
	0 to -15%	-1		-20 to 20	0
	< -15%	-2		-50 to -20	-1
				-100 to -50	-3

Table 3.12 Impact of various reinforcement signal for different types of investments on the basis of inferences drawn (Very Low, Low, Medium, High, V High, Neg Low, Neg. Medium)

Investment perspective	QSG	SMVA	SS	CAV	STF	PS3Y	SPMPVA	CASG	P/E	P/B	QPG	News
Short term	H	L	VL	L	VH	M	H	L	VL	VL	H	VH
Mid term	M	M	L	M	L	H	M	H	L	VL	M	M
Long term	H	VH	NL	H	VL	H	M	VH	M	M	L	L

Table 3.13 Weight allocation to various reinforcement signal for different types of investments on the basis of their impact shown in Table 3.12

Investment perspective	QSG	SMVA	SS	CAV	STF	PS3Y	SPMPVA	CASG	P/E	P/B	QPG	News	IS
Short term	4	2	1	2	5	3	4	2	1	1	4	5	4
Mid term	3	3	2	3	2	4	3	4	2	1	3	3	3
Long term	4	5	-2	4	1	4	3	5	3	3	2	2	-3

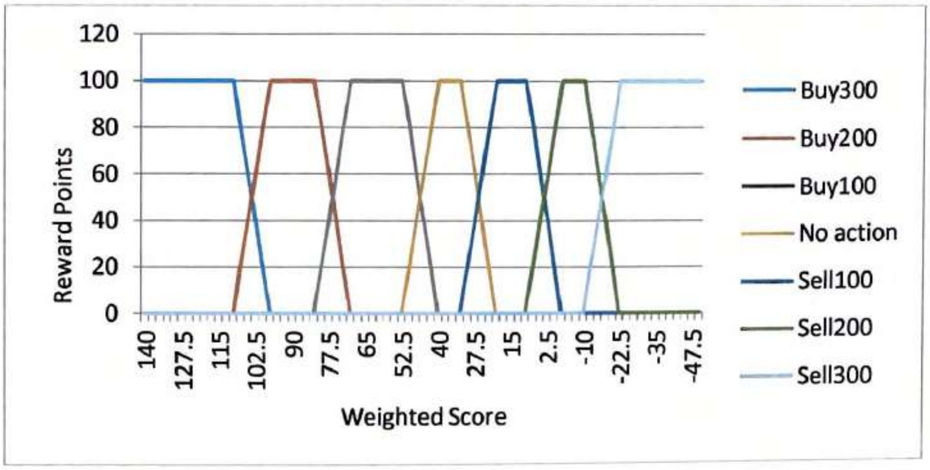


Fig. 3.15 Reward points for various actions

In the learning phase, a mixture of Q learning and supervised learning can be applied. For Q learning exploitation/exploration policy is to be adopted is to with an epsilon value ranging between 0.3 and 0.4. The reward modification was done on the basis of the extent of the success of random action and failure of suggested action and their continued repetition. The modified rewards for a given action lead to change in the expected weight score under a given state of the environment described by the status of various reinforcement signals defined in Table 3.11. This modified weighted score can be used to modify the weight vectors of Table 3.13 by using the supervised learning with environment state acting as the feature vector and weighted score as target value. Fig. 3.16 shows the details of the learning process. The learning process is subject to vary based upon the stock market behavior and choice and success/ failure of random actions. We are leaving it with a qualitative treatment only.

Fig. 3.16 Learning phase of the proposed predictive framework

3.6 RESULTS

To evaluate the efficacy of the proposed scheme, with 31st March 2019 as the base, a notional investment of Rs. 5000000 was made under following mechanisms.

1. Equal investment of Rs. 100000 was made on each stock making the baseline investment of Rs. 500000.
2. Investment was evaluated on the quarterly basis leaving it on the mercy of random movement in stock market.
3. Investment of Rs. 5000000 was made in top 20 stocks on the basis of proposed predictive framework for the short term basis with continuous evaluation and sell/buy/hold decision on the fortnightly basis.
4. Investment of Rs. 5000000 was made in top 20 stocks on the basis of proposed predictive framework for the Mid-term basis with continuous evaluation and sell/buy/hold decision on the 30 days basis.
5. Investment of Rs. 5000000 was made in top 20 stocks on the basis of proposed predictive framework for the Long term basis with continuous evaluation and sell/buy/hold decision on the quarterly basis.

Fig. 3.17 shows the status of the investment made under various types of above mentioned categories. It can be seen that sensing of reinforcement signals and taking the appropriate actions can result in 10 to 15% more gain in the investments.

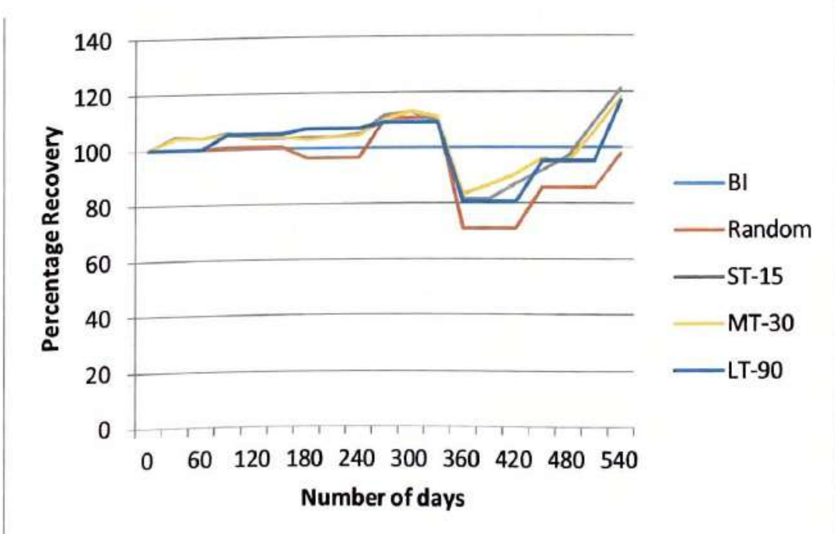


Fig. 3.17 Status of various types of investments

Referring to the observation made in section 3.2.2.6 related to various sectors, it was inferred that investment in pharma sectors is risky due to the very weak sector sentiments. The normalized moving average (NMVA) of different pharma companies in the future period, shown in Fig. 3.18, endorses the inference drawn. In case of FMCG sector, the sector sentiments were neutral and future moving average trend in Fig. 3.19 depict that Britannia and ITC have shown a slight loss while Nestle has made a significant gain. In case of private banking sector and Financial Services Sector the strong sector sentiment continues and almost all the companies show the gain as shown in Fig. 3.20 and Fig. 3.21. Refer section 3.2.2.7, wherein it was observed that due to weak performance amongst its peer group stock of IndusInd Bank should be parted with. The stock prices of the IndusInd Bank in the future study period endorse that observation and inference. Results were also drawn for other sectors also, but only some of them have been shown for representation. We have gone for normalized moving average for the stocks instead of the actual prices as they are the actual reflection of the trend and devoid of the noisy spikes in the stock price.

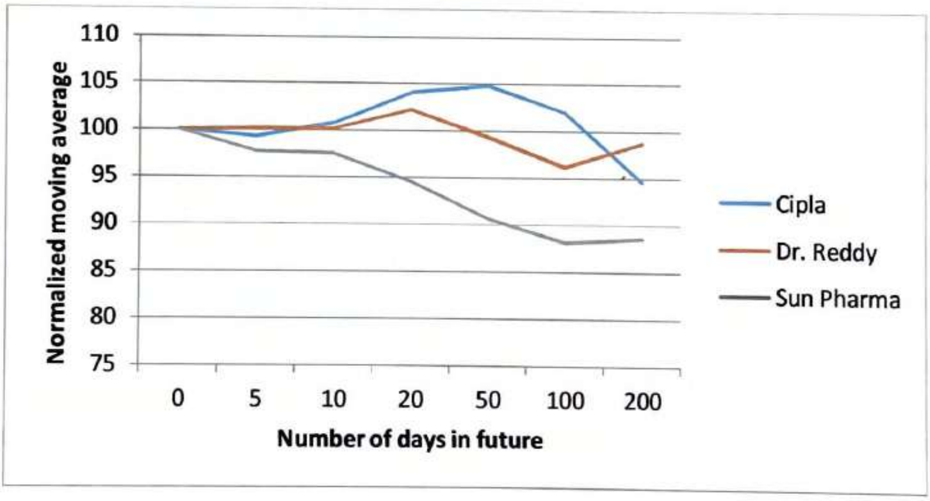


Fig. 3.18 NMVA of different Pharma companies in future

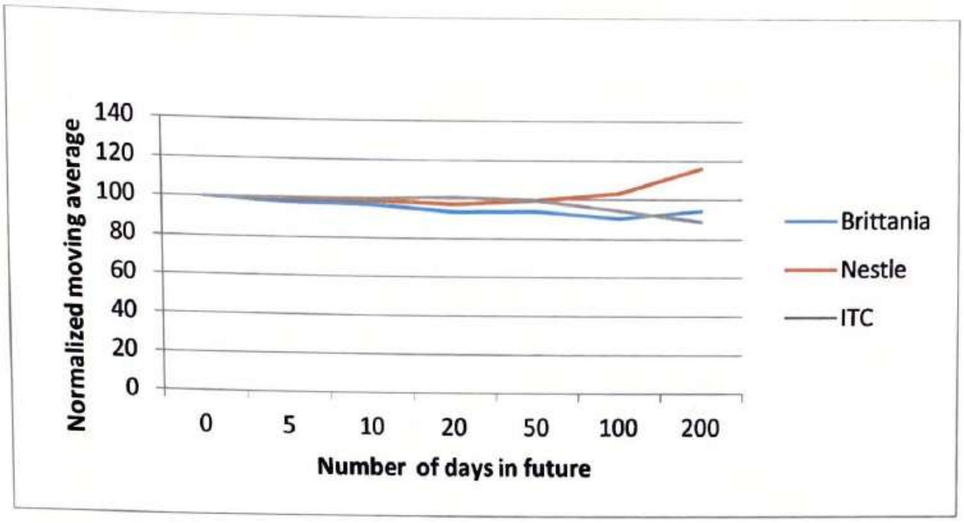


Fig. 3.19 NMVA of different FMCG companies in future

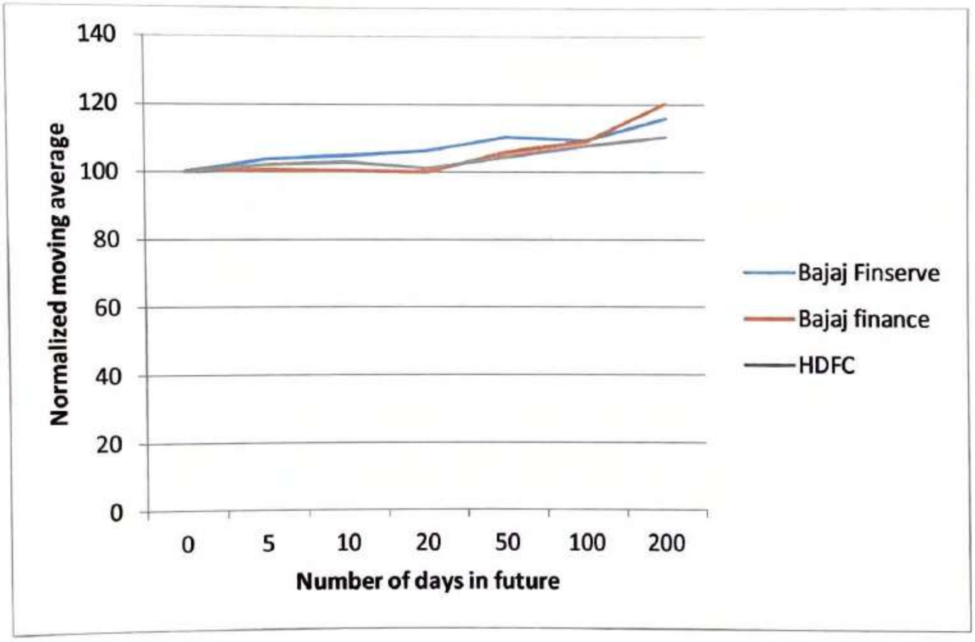


Fig. 3.20 NMVA of different finance companies in future

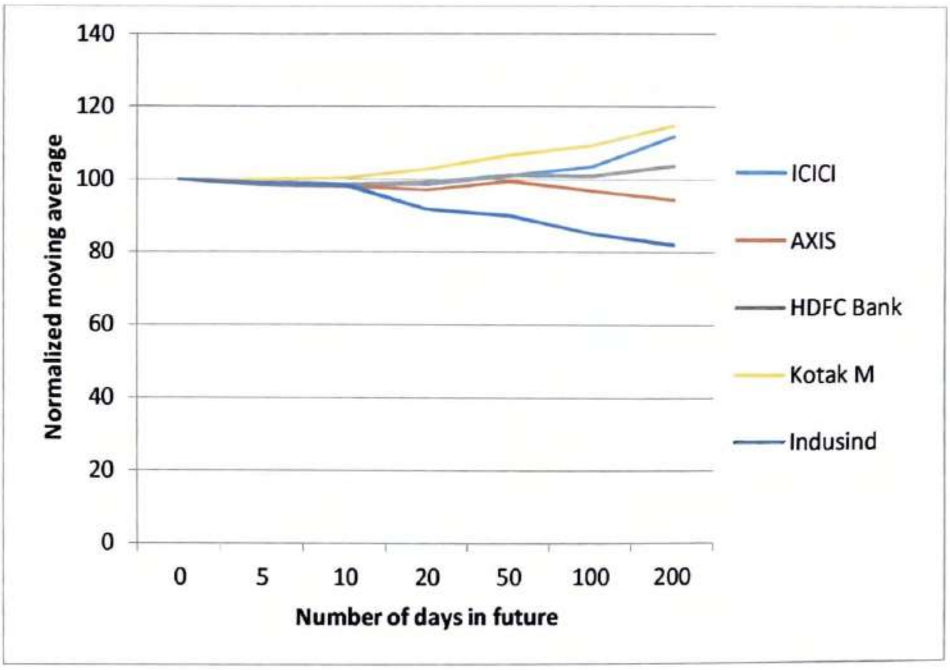


Fig. 3.21 NMVA of different pvt. banks in future

It is not possible to accurately predict the future but reinforcement signals being seen in the current environment can help us in anticipating the outcome of future events provided the circumstances remain normal. Past history can help us in understanding the positive and negative impact of various reinforcement signals and their intensity. We have identified various reinforcement signals and studied their impact on the stock prices. The knowledge acquired in the process has been used to develop an analytics framework wherein different actions relating to the investment, under the given stock market situations, have been associated with the corresponding rewards predict the appropriateness of the action. The results have been quite satisfying and we are sure that the proposed framework is of immense help to both institutional and individual investors.

CHAPTER-V

A PREDICTIVE ANALYTICS FRAMEWORK FOR OPPORTUNITY SENSING IN STOCK MARKET

Easy access to stock market data in electronic form has provided the researchers, a platform for extracting useful patterns for making future price predictions. In India, this data is available on various websites such as NSE [20], rediffmoney [17], moneycontrol [19], yahoo finance [18] etc. Most of the predictive systems are based upon pure machine learning algorithms and their designer hope that the machine will identify the underlying pattern in the data and make the predictions [85], [97]–[102], [102], [103], [103]–[109]. Reasons for the immense popularity of machine learning include: ability to handle large amount of data, ability to map between well-defined inputs and outputs, availability of large digital data sets in current scenario, eradication of long chains of logical reasoning, tolerance for error, special skills not required for the user. The availability of the open source machine learning APIs have fuelled these engines.

The problem with the stock market price data is that the stock prices fluctuate many times a day and over a period of month there are so many random variations. If this data is used as such, then there are two possibilities. A rigorously trained system likely to fall in the trap of overfitting leading to quite erroneous results while in the validation phase. An ordinarily trained system would fall in the trap of underfitting and will not be able to provide accurate results. It is for this reason that most of the research papers available in the stock market prediction domain predict the data for a very small period from one day to one week [41], [54], [57], [110], [111]and are limited to only few prominent stocks [48], [112]–[116].

Keeping this aspect in view, we are of the view that the design of a predictive system should ideally be based upon macrofeatures [117], [118], which can be created through a mathematical function based upon multiple low level features. These new reduced set of features should then be able to summarize most of the information contained in the original set of features. These features shall be more informative and interpretable in a better way. A system designed on the basis of these features shall have the following advantages over the conventionally designed purely machine learning based system: improved data visualization, increase in the explainability of the model, overfitting risk reduction, improved accuracy, easy to debug and transfer learning ability (reusability of modules).

In case of stock market, many such macrofeatures are available, which have been created by the various statistical researchers. These macrofeatures have been developed over time and interpreted for their utility. Number of these macrofeatures is quite large and the information provided by them is quite overlapping many times. It is possible to identify the non-overlapping macrofeatures and to combine the pieces of semi-processed information provided by them to create a meaningful prediction system based upon machine learning. The work carried out undertakes this task by applying supervised learning on the combination of macrofeatures.

The task includes the extraction of non-overlapping macrofeatures from the past price data over a period of 120 working days (almost 6 months) for creating the unified input feature vector and the data of next 30 working days (almost month and a half) for the as desired output. The learning mechanism so created is used for prediction on similar future data. This completes the cycle. The output obtained, is a predicted price band which is likely to prevail in the upcoming month and a half. 4 such cycles have been used to demonstrate the precision of the proposed mechanism. The ensuring of the precision is an academic task that requires monetization aspect for its commercial usage. Keeping this aspect in view, the price band so obtained has been used for signaling the various opportunities like buying/selling/wait keeping in view the prevailing price position. The results show that the system is effective and can be used for continuous gain.

5.1 PROBLEM DEFINITION AND OBJECTIVES

Before taking up details of the proposed work, let us explicitly define the problem and the associated objectives.

5.1.1 Problem Definition

To predict the future price band of the NIFTY50 stocks by using different macro features create on the historic price data.

5.1.2 Objectives

- To design a framework using regression based supervised learning to predict the maximum, minimum and average price of the stock for the upcoming 30 working days from the historical data of past 120 working days.
- Input feature vector of the supervised learning process should be created from component feature vectors obtained from the different macrofeatures.
- The macrofeatures used for the purpose should have the different views in order to represent multiple aspects of the data.

5.2 PROPOSED FRAMEWORK

As described in the previous section, the proposed work is based upon the different macrofeatures capable of providing complete and nearly non-overlapping views. We start the discussion with the selection of macrofeatures.

5.2.1 Selection of Macrofeatures

The first view involves the comparative strength in the price movement wherein it is observed for how many days the price went up or down and by what magnitude in the period under consideration. The stock markets normally use Relative Strength Index RSI (14) for this purpose to measure the relative strength of the price in the past 14 working days. This macrofeature was developed by J. Welles Wilder [59]. RSI is a short term momentum indicator whose value oscillates between the 0 and 100. The value of the index is has been calculated for the recent past using a single-step formula as given in Eq. 5.1. The detailed excel sheet based calculations for RSI can be seen at various websites [33].

$$RSI = 100 - \left[\frac{100}{1 + \frac{Average\ gain}{Average\ loss}} \right] \tag{5.1}$$

RSI was implemented through sliding window process in Microsoft Excel. We have used RSI (120), RSI (60), RSI (30), RSI (15) and RSI (5) to create the component feature set.

The second view involves the moving average for the past period which conveys the basic direction of the price movement. This movement can also be generally increasing, oscillating or generally decreasing. To take the historical account, Simple Moving Average (SMA) was computed for the period of 5, 15, 30, 60 and 120 days and was represented as SMA_k where k is the number of days with current day being referred as nth day. SMA [60] is the arithmetic mean of the close price as in Eq. 5.2.

$$SMA_k = \frac{1}{k} \sum_{i=n-k+1}^{n} p_i \tag{5.2}$$

The third view involves the possible upward or downward movement of the price to anticipate the risk. For this purpose, the stock market uses Bollinger Bands for past 20 days with 2 levels of standard deviation represented as BB (20, 2). The component feature set relating to this view has been created using BB (120, 2), BB (60, 2), BB (30, 2), BB (15, 2) and BB (5, 2).Bollinger Bands [61] are a technical analysis tool, specifically they are a type of trading band or envelope. BB uses central tendency, such as moving average, as the base for defining highs and lows of the band referred to as upper band (UB) and lower band (LB). Formulas for computing the UB and LB are as shown (3) & (4).

$$UpperBB = MA + D\sqrt{\frac{\sum_{i=1}^{n}(y_i - MA)^2}{n}} \tag{5.3}$$

$$LowerBB = MA - D\sqrt{\frac{\sum_{i=1}^{n}(y_i - MA)^2}{n}} \tag{5.4}$$

where MA is the Moving Average and D represents the number of standard deviations. The fourth view of the component feature set is the actual closing price on the start and ending date of the cycle represented as Day1_CP and Day120_CP respectively.

These four views take care of multiple aspects of the price data which consider the basic direction of price movement, dispersion in price movement and relative magnitude of upward / downward movement on the daily basis thereby completing the entire spectrum. Here it is worth mentioning that the input data so created is almost free from random fluctuations, distractive patterns and much lesser in volume. Moreover, supervised learning model created from such a data would be much more transparent than the one created on the raw data.

5.2.2 OVERVIEW AND WORKING OF THE PROPOSED FRAMEWORK

Fig. 5.1 shows the overview of the proposed predictive framework. The work begins with the identification of the macrofeatures keeping their applicability in view. As described in the previous subsection, the identified macrofeatures are RSI, SMA and BB with their computations at various junctures for past 120 working days. Next Step involves the design of unified input feature vector. Fig 5.2 shows the components of three input feature vectors used for predicting the maximum, minimum and the average output. The feature vector for maximum price prediction uses Upper Bollinger Band (UB) only. Feature vector for minimum price prediction uses Lower Bollinger Band (LB) only. Feature vector for average price prediction uses both UB and LB. Now raw data relating to past 120 working days is picked up and the unified input feature vector is created for all the companies under consideration. Thereafter, the data for next 30 working days is taken as target output for training purpose. Training module, so obtained, is applied on the next cycle 120 working days input data to obtain the predicted output. The obtained predicted output is compared with the output of the next 30 working days to validate the results. The process is repeated many times to ensure the consistency and accuracy of results.

The trained module, so obtained, is used for opportunity sensing. The design of the opportunity sensing module is shown in Fig. 5.3. It is based upon the normalization of the min and max value to range [0,100]. Fuzzy sets are created over this normalized range. The current price is then applied to the normalized range to identify the corresponding fuzzy set(s) with its membership. Fuzzy set(s) and their corresponding memberships are used to identify the applicable opportunity with associated rewards. If the rewards obtained exceed the threshold then corresponding opportunity is signaled. The significance of signaled opportunities is validated for their monetization abilities.

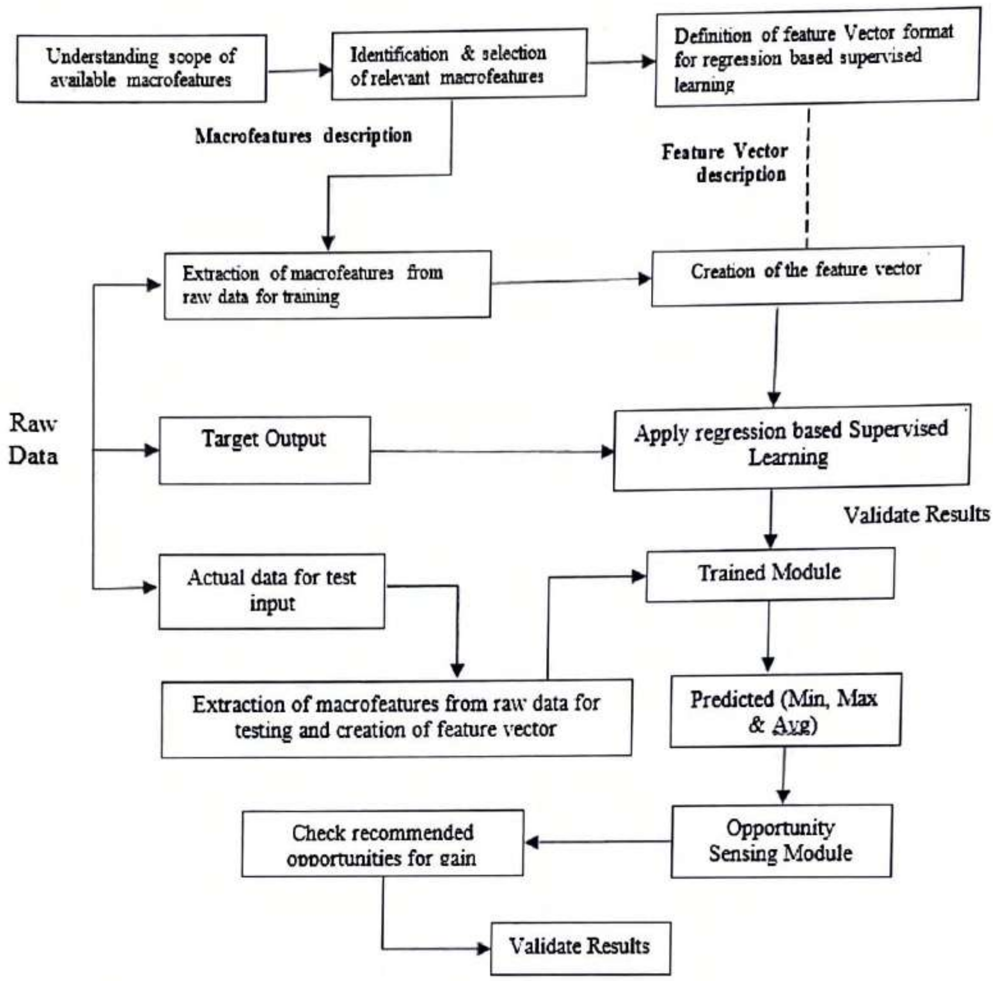


Fig. 5.1: Overview of the Proposed Predictive Framework

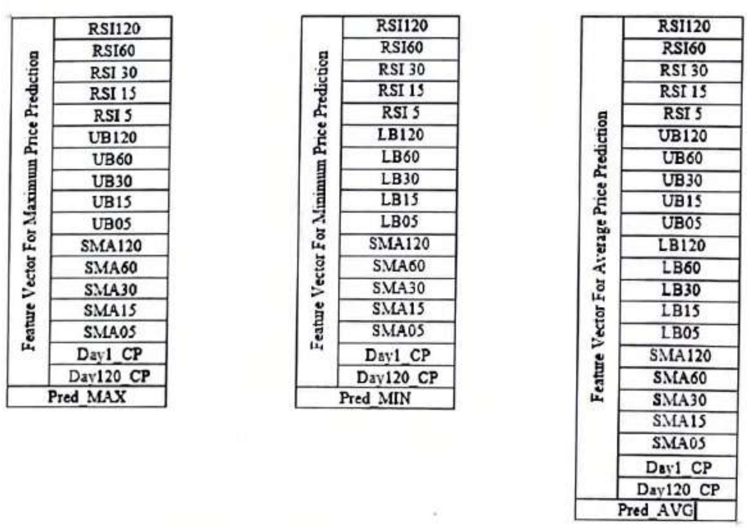


Fig 5.2: Feature Vectors for Maximum, minimum and Average Price Prediction

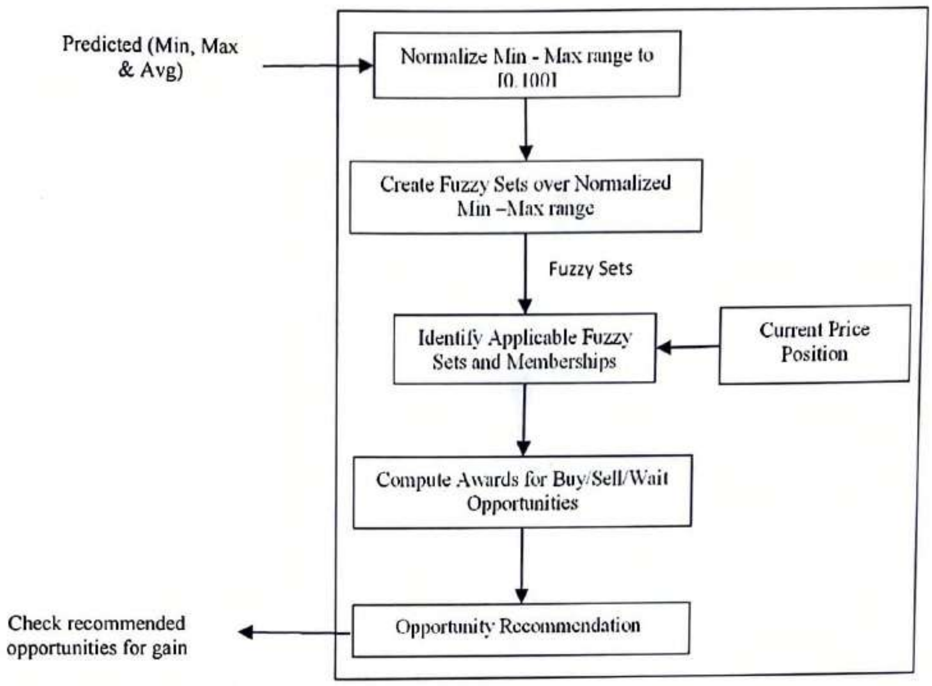


Fig. 5.3: Design of the Opportunity Sensing Module

The subsequent section talks about the conduct of the experiment based upon the proposed predictive framework.

5.3 THE EXPERIMENT

5.3.1 Experimental Setup

As described earlier, the experiment was repeated for 4 cycles by taking the data for 46 NIFTY50 companies, listed in Table 5.1 from the various websites [17]–[20] for evaluating the framework. Detail of various cycles is given in Table 5.2.

An artificial neural network (ANN) was trained on the input/output using Python 3.7 After reducing the mean square error (MSE) to its minimum, trained module was applied on the input data of prediction phase. The results so obtained were checked with the actual output for the purpose of validation. Table 5.3 5.4, 5.5 and 5.6 show the training, testing and validation details for Cycle 1 for Max, Min and Average stock prices.

Table 5.1: List of NIFTY50 companies under consideration			
Co.No.	**STOCK**	**Co.No.**	**STOCK**
1	Adani Ports & Special Economic Zone	24	ITC Ltd.(L)
2	Asian Paints Ltd.(L)	25	JSW Steel Ltd.(L)
3	Axis Bank Ltd.(L)	26	Kotak Mahindra Bank Ltd.(L)
4	Bajaj Auto Ltd.(L)	27	Larsen & Toubro Ltd.(L)
5	Bajaj Finance Ltd.(L)	28	Mahindra & Mahindra Ltd.(L)
6	Bajaj Finserv Ltd.(L)	29	Maruti Suzuki India Ltd.(L)
7	Bharat Petroleum Corporation Ltd.(L)	30	Nestle India Ltd.(L)
8	Bharti Airtel Ltd.(L)	31	NTPC Ltd.(L)
9	Britannia Industries Ltd.(L)	32	Oil & Natural Gas Corporation Ltd.(L)
10	Cipla Ltd.(L)	33	Power Grid Corporation Of India
11	Coal India Ltd.(L)	34	Reliance Industries Ltd.(L)
12	Dr. Reddys Laboratories Ltd.(L)	35	Shree Cement Ltd.(L)
13	Eicher Motors Ltd.(L)	36	State Bank Of India(L)
14	GAIL (India) Ltd.(L)	37	Tata Consultancy Services Ltd.(L)
15	Grasim Industries Ltd.(L)	38	Tata Motors Ltd.(L)
16	HCL Technologies Ltd.(L)	39	Tata Steel Ltd.(L)
17	HDFC Bank Ltd.(L)	40	Tech Mahindra Ltd.(L)
18	Hero MotoCorp Ltd.(L)	41	Titan Company Ltd.(L)
19	Hindalco Industries Ltd.(L)	42	Ultratech Cement Ltd.(L)
20	Hindustan Unilever Ltd.(L)	43	UPL Ltd.(L)
21	HDFC	44	Vedanta Ltd.(L)
22	ICICI Bank Ltd.(L)	45	Wipro Ltd.(L)
23	Infosys Ltd.(L)	46	Zee Entertainment Enterprises Ltd

Table 5.2 Training, Testing and Validation details				
CYCLE	**Training Phase**		**Prediction & Validation Phase**	
	Period for input training data feature extraction (120 working days)	**Period for output data used for supervised learning (30 working days)**	**Input data period for feature extraction for testing (120 working days)**	**Prediction period for validation (30 working days)**
Cycle1	1 July 20-18 Dec. 20	21 Dec.20-2 Feb. 21	3 Aug 20- 21 Jan. 21	22 Jan. 21-5 March 21
Cycle2	3 Aug 20- 21 Jan. 21	22 Jan. 21-5 March 21	1st Sept, 20 – 22nd Feb, 21	23rd Feb, 21 – 8th April, 21
Cycle 3	1 Sept, 20 – 22nd Feb, 21	23rd Feb, 21 – 8th April, 21	1st Oct, 20 – 25th Mar, 21	26th March , 21-12th May, 21
Cycle 4	1st Oct, 20 – 25th Mar, 21	26th March , 21-12th May, 21	2nd Nov, 20- 29th April, 21	30thApril, 21-11th June, 21

Table 5.3 Training Data for Max Price Prediction – Cycle 1

Co.No.	RSI 120	RSI 60	RSI 30	RSI 15	RSI 5	UB 120	UB 60	UB 30	UB 15	UB 5	SMA 120	SMA 60	SMA30	SMA 15	SMA 5	Day1 CP	Day120 CP	MAX
1	58.6	67.0	69.1	63.4	23.1	438.7	472.8	500.7	499.1	480.0	357.3	381.9	414.3	453.0	470.5	343.8	463.3	558.6
2	64.1	69.0	70.7	81.9	59.2	2466.8	2539.5	2626.0	2710.1	2605.9	1997.3	2190.5	2310.5	2432.0	2550.8	1688.0	2601.5	2849.3
3	57.6	66.1	56.8	37.4	15.3	627.6	684.1	655.5	631.6	619.0	486.5	535.1	602.2	612.8	609.6	433.3	609.5	714.2
4	56.4	58.7	66.0	58.2	35.5	3233.3	3342.7	3432.4	3396.3	3340.3	2995.8	3064.8	3151.0	3279.7	3292.4	2842.1	3347.6	4201.5
5	63.1	70.0	76.3	59.6	64.4	4853.1	5417.5	5476.2	5210.1	5427.8	3639.7	3965.8	4617.3	4914.4	5059.7	2953.9	5252.5	5334.1
6	62.0	74.1	81.2	51.2	42.2	8745.1	9852.2	10411.5	9372.0	9508.6	6592.9	7002.8	8178.3	9007.9	9155.5	6169.4	9215.6	9721.8
7	50.9	50.3	54.3	44.3	18.8	449.8	415.1	411.8	410.1	406.9	387.8	370.1	386.9	391.8	399.3	380.1	393.9	426.7
8	47.4	54.2	57.8	60.3	33.0	598.6	521.5	520.4	528.6	521.5	493.4	456.1	483.2	495.6	508.6	560.2	510.6	602.7
9	51.9	53.2	62.7	56.8	22.4	3943.0	3862.6	3783.6	3812.6	3792.4	3671.2	3620.8	3596.9	3673.1	3751.5	3545.9	3720.5	3665.4
10	56.4	51.3	53.1	65.6	65.1	824.2	803.7	797.8	795.4	806.3	732.0	764.8	758.5	765.9	779.4	627.1	793.5	856.8
11	52.0	59.6	61.3	60.2	36.9	143.1	140.6	145.3	148.6	148.1	125.8	122.2	129.2	135.2	142.7	133.9	141.7	146.6
12	59.1	50.9	54.7	60.9	56.9	5427.5	5259.1	5131.3	5148.0	5080.4	4625.4	4984.6	4897.9	4972.0	5036.6	3911.4	5228.8	5416.8
13	57.6	60.2	62.0	30.3	35.3	2571.5	2696.4	2777.4	2605.7	2519.7	2182.1	2313.1	2461.5	2512.3	2456.6	1840.2	2480.9	2972.6
14	56.1	70.6	72.4	63.9	8.6	117.5	125.0	135.3	134.1	126.6	95.9	95.6	105.9	118.8	125.0	102.9	125.1	144.0
15	61.4	64.4	62.6	50.0	25.1	926.6	948.9	966.1	944.5	920.4	725.2	812.9	867.8	911.0	910.6	614.3	909.7	1190.0
16	62.3	54.7	58.0	59.3	56.6	941.1	891.0	891.0	897.2	896.7	758.3	842.1	843.8	860.3	877.1	551.8	890.7	1062.1
17	59.4	69.7	62.8	67.7	71.3	2366.2	2539.0	2454.4	2447.7	2543.5	1940.1	2094.0	2280.1	2310.4	2373.6	1835.2	2479.4	2751.2
18	56.0	52.8	52.8	46.6	11.3	3317.8	3349.5	3238.3	3228.5	3192.2	2943.8	3090.2	3074.7	3137.2	3128.3	2547.4	3135.3	3474.9
19	61.4	67.8	66.4	51.4	35.3	238.2	258.5	267.2	258.6	252.4	185.2	198.8	223.9	241.6	246.4	147.6	249.9	272.9
20	53.4	62.3	66.7	67.3	27.5	2297.2	2319.1	2376.4	2441.2	2400.2	2154.1	2155.5	2192.6	2250.7	2345.0	2170.4	2333.4	2450.6
21	60.9	69.9	62.2	47.4	65.2	71.1	75.2	71.2	71.1	72.2	54.6	61.6	67.6	68.4	69.1	47.9	68.3	80.3
22	58.8	65.8	59.8	57.9	31.0	506.5	547.1	532.6	535.4	528.4	400.0	438.2	486.4	501.2	516.4	364.0	517.2	617.3
23	64.0	60.2	62.4	65.9	59.8	1226.5	1200.9	1179.9	1189.6	1170.5	1002.9	1104.0	1131.1	1146.7	1161.4	731.9	1189.8	1387.2
24	52.3	65.2	69.3	59.6	11.1	213.5	213.7	221.1	222.2	217.5	186.1	181.6	194.4	204.6	214.6	202.9	214.5	219.8
25	66.0	65.9	61.9	46.2	40.5	386.6	386.0	380.4	374.8	374.6	282.0	322.7	349.3	363.3	364.0	191.4	364.2	405.4
26	60.8	71.6	68.3	54.4	50.3	1912.5	2115.2	2009.0	1992.1	1979.6	1468.0	1606.4	1839.4	1888.2	1945.8	1356.8	1949.3	2017.9
27	62.2	74.9	76.2	74.8	64.4	1166.7	1265.1	1298.1	1295.3	1312.6	969.4	1015.3	1120.4	1179.7	1247.5	924.3	1281.6	1520.2
28	60.1	60.8	62.8	37.5	26.4	730.4	771.2	804.4	765.3	741.5	622.5	653.4	697.1	736.8	724.0	499.0	733.6	829.9
29	58.5	61.3	58.2	62.5	17.9	7760.4	7802.7	8031.6	8180.3	7825.5	6759.5	7068.8	7260.3	7567.9	7737.2	5803.1	7559.8	8232.8
30	53.6	62.1	59.3	56.4	39.3	17975.6	18617.9	18625.3	18752.1	18660.1	16540.2	16789.2	17503.0	17896.8	18362.3	16798.8	18362.8	18732.7
31	53.1	59.7	62.0	56.2	29.2	102.6	103.5	106.3	107.7	107.0	90.0	89.6	94.9	99.3	104.4	93.6	103.8	102.6
32	56.1	66.8	68.7	61.3	33.6	92.3	96.4	105.1	107.4	105.3	76.1	74.9	81.7	91.5	100.6	80.4	99.0	105.3
33	54.0	59.6	54.3	30.4	28.9	197.2	206.1	201.0	196.8	196.7	175.4	177.5	190.5	193.6	193.9	173.4	193.9	205.1
34	53.6	43.1	55.5	47.7	17.1	2323.3	2309.7	2055.3	2028.1	2012.1	2051.1	2066.6	1978.3	1975.1	1986.7	1737.6	1991.6	2099.4
35	52.8	65.7	59.6	42.3	47.6	24781.1	25958.6	25603.8	25287.8	24780.9	21649.8	22248.1	23863.6	24434.3	24111.3	22663.5	24574.4	26432.6
36	59.7	67.9	66.1	57.1	19.1	258.4	282.3	286.0	285.0	276.0	206.9	220.4	247.6	262.7	270.9	184.8	271.5	333.1
37	61.1	58.3	60.4	60.6	59.6	2915.1	2901.3	2836.0	2853.3	2857.4	2459.4	2683.6	2720.5	2762.6	2799.1	2092.1	2861.0	3308.8
38	61.8	64.3	64.1	40.1	20.3	183.5	193.2	200.9	186.9	184.1	135.7	150.6	168.1	180.8	179.9	100.8	180.6	322.0
39	67.1	74.1	77.5	59.2	31.3	593.5	665.9	698.3	655.5	649.4	423.7	466.7	549.6	612.6	632.5	323.7	631.3	722.8
40	63.7	58.3	61.5	63.5	41.3	966.9	943.2	961.6	959.9	937.5	761.9	850.8	880.3	915.3	921.5	545.9	933.8	1077.6
41	64.2	69.4	70.4	72.8	61.3	1435.3	1485.1	1519.9	1528.2	1529.3	1172.3	1278.6	1357.4	1423.9	1476.2	948.2	1519.3	1572.6
42	60.0	66.7	60.5	56.6	49.1	5102.8	5356.1	5262.9	5284.2	5233.7	4248.9	4603.3	4901.9	5025.2	5106.1	3898.6	5152.5	6123.5
43	50.7	43.2	53.4	54.2	52.0	532.8	530.3	477.1	490.6	468.1	465.3	459.7	436.7	449.6	448.2	447.0	460.3	588.2
44	57.4	56.2	73.5	65.7	17.0	146.7	153.0	156.3	157.5	148.3	117.7	116.6	120.2	135.3	146.7	107.7	147.4	185.5
45	64.5	58.7	57.0	49.3	50.7	389.8	376.4	364.3	364.6	361.3	305.0	344.1	351.9	356.4	354.2	219.1	363.5	459.0
46	54.7	55.5	63.6	65.5	49.9	234.4	224.2	228.9	237.1	241.8	186.8	196.1	200.5	210.8	223.6	175.4	225.3	242.9

Table 5.4 Validation and Testing Data for Max Price Prediction - Cycle 1

Co. No	RSI 120	RSI 60	RSI 30	RSI 15	RSI 5	UB 120	UB 60	UB 30	UB 15	UB 5	SMA 120	SMA 60	SMA 30	SMA 15	SMA 5	Day1 CP	Day120 CP	Ac. MAX	Pr. MAX	Err %
1	64.7	68.9	63.4	65.0	49.3	520.7	564.2	548.7	555.1	565.9	390.8	440.7	493.6	516.1	534.4	311.4	547.7	752.5	639.1	15.1
2	64.4	65.8	56.9	37.4	39.6	2801.5	2930.5	2869.0	2899.6	2731.5	2186.4	2444.5	2665.1	2732.6	2637.1	1705.5	2716.4	2596.6	3029.3	-16.7
3	61.2	64.4	53.8	56.7	27.2	699.2	705.6	698.3	703.5	686.7	525.5	606.4	636.0	661.0	670.5	417.4	675.8	794.0	786.7	0.9
4	59.8	74.9	66.4	84.2	97.0	3523.1	3716.5	3699.3	3691.4	3677.7	3096.1	3248.9	3436.9	3547.8	3601.0	2917.4	3703.3	4237.5	4002.5	5.5
5	61.2	68.0	53.5	38.1	56.7	3525.4	3851.2	3386.0	3345.4	3089.9	3996.6	4650.2	5037.0	5030.3	4884.4	3174.5	5118.3	5779.9	5583.5	3.4
6	61.6	72.1	51.2	53.9	64.6	9828.0	10673.8	9333.9	9223.1	9200.3	7085.8	8162.0	8951.3	8852.1	8720.1	6103.1	9052.8	10537.7	9748.7	7.5
7	49.3	59.3	51.6	57.4	11.0	436.2	423.4	423.2	428.3	428.6	384.4	385.0	394.9	401.6	415.6	417.8	408.3	469.0	493.7	-5.3
8	51.7	65.1	65.3	71.9	47.2	584.4	583.4	598.4	621.4	608.7	487.9	500.2	531.7	554.4	591.7	547.3	581.3	608.8	688.4	-13.1
9	48.1	54.9	47.4	55.0	28.6	3905.4	3748.5	3772.0	3661.1	3672.9	3643.4	3586.6	3632.7	3590.5	3615.2	3776.2	3630.8	3613.6	4040.2	-11.8
10	53.9	55.6	56.7	46.4	32.8	833.1	857.7	867.4	856.2	853.6	764.2	783.8	812.9	830.6	824.3	710.4	812.8	863.4	964.1	-11.7
11	52.1	59.2	49.2	46.9	14.7	146.9	150.7	147.4	148.4	148.2	127.1	131.0	139.1	139.9	142.0	128.4	137.4	156.1	176.6	-13.2
12	53.7	51.1	50.8	38.6	10.6	5455.0	5401.9	5408.0	5455.8	5366.2	4850.3	5027.7	5180.8	5247.1	5148.7	4527.5	5052.2	5082.4	5761.2	-13.4
13	60.2	69.6	68.4	87.5	79.0	2779.6	2954.6	2964.8	3018.9	2900.0	2312.6	2492.6	2611.0	2758.0	2876.4	2078.2	2926.6	2972.6	3186.0	-7.2
14	59.9	70.0	58.6	64.4	14.6	136.6	151.0	144.6	147.5	144.7	101.5	112.5	128.5	134.7	139.6	94.1	133.8	150.8	158.4	-5.0
15	63.9	67.7	60.2	66.0	42.8	1024.2	1049.7	1055.8	1066.3	1053.6	793.7	894.3	951.4	995.8	1023.8	633.5	1017.1	1350.4	1176.6	12.9
16	59.7	59.4	63.7	56.9	22.9	1008.3	1033.8	1067.7	1071.1	1034.0	823.3	890.5	946.6	997.3	996.4	706.0	993.3	985.8	1156.4	-17.3
17	62.9	63.3	62.3	52.3	36.0	2716.4	2789.5	2809.4	2767.6	2744.9	2079.9	2365.0	2528.2	2647.6	2645.7	1739.8	2663.0	2860.4	2964.5	-3.6
18	56.5	60.8	53.9	68.5	64.6	3324.6	3289.6	3286.4	3329.2	3321.0	3021.0	3074.7	3134.8	3164.5	3228.3	2831.3	3247.3	3584.0	3636.1	-1.5
19	58.7	63.4	48.0	49.3	22.3	266.9	285.8	270.7	276.2	265.3	202.4	227.7	248.4	255.8	252.3	161.4	248.9	359.4	319.7	11.0
20	53.3	58.2	56.5	51.5	53.4	2399.6	2501.4	2447.6	2453.1	2408.7	2183.3	2255.4	2369.5	2387.7	2360.3	2204.5	2367.6	2409.4	2648.6	-9.9
21	64.2	62.1	52.4	46.4	13.3	78.8	77.5	76.9	77.1	75.8	59.3	68.4	71.4	73.5	74.3	46.3	72.8	83.3	79.3	4.7
22	62.1	65.7	55.7	49.4	33.4	569.3	581.3	562.4	559.1	561.2	432.7	495.0	528.2	542.5	545.4	343.3	552.7	674.0	647.7	3.9
23	61.4	61.9	67.5	63.8	33.0	1332.0	1368.6	1405.5	1407.9	1383.9	1084.6	1184.5	1257.9	1318.2	1336.7	956.9	1339.7	1343.6	1526.0	-13.6
24	54.1	65.2	54.8	51.9	38.2	221.2	230.7	221.4	223.5	221.9	188.6	197.2	210.9	211.1	217.7	192.6	213.7	234.4	277.2	-18.3
25	64.3	63.7	53.4	44.5	31.4	412.4	418.8	416.7	410.0	407.2	317.6	359.5	382.3	397.1	394.0	220.6	393.3	428.1	481.1	-12.4
26	59.9	63.4	49.6	24.3	27.7	2125.1	2115.2	2024.3	2025.6	1912.8	1580.8	1843.3	1936.1	1927.1	1874.8	1308.6	1857.8	2019.7	2144.6	-6.2
27	63.8	71.5	69.1	64.3	43.3	1367.5	1457.6	1407.5	1400.7	1401.5	1044.4	1175.7	1296.1	1340.2	1363.9	915.9	1371.3	1566.4	1544.0	1.4
28	58.8	66.8	55.9	67.8	25.5	790.8	843.3	834.1	857.5	837.4	657.4	707.5	748.9	778.8	818.2	597.2	808.3	928.4	938.6	-1.1
29	59.3	59.6	55.4	61.8	43.1	7977.7	8269.0	8249.4	8380.2	8293.0	7091.3	7423.1	7748.3	7902.8	8014.5	6165.2	8076.8	8048.9	8864.9	-10.1
30	52.7	53.9	50.6	33.3	18.5	18709.0	19001.6	18866.4	18769.9	18202.1	16787.4	17725.9	18251.7	18144.6	17806.7	16524.8	17716.3	17589.6	19737.3	-12.2
31	53.0	55.8	45.7	38.7	17.0	106.0	106.9	105.2	102.6	103.3	91.9	95.8	100.1	99.3	99.6	85.6	96.1	112.7	117.4	-4.2
32	55.5	62.6	51.5	49.5	13.2	102.5	112.4	106.4	106.9	106.6	79.3	85.7	97.2	99.0	100.0	75.9	94.7	119.1	111.9	6.0
33	53.2	59.3	52.0	55.8	20.9	205.0	207.2	205.9	210.3	208.2	179.5	190.3	194.7	197.6	199.9	178.6	195.1	236.5	256.8	-8.6
34	51.4	52.2	53.8	56.7	68.9	2293.7	2064.0	2045.9	2047.0	2083.1	2056.2	1974.4	1973.2	1962.3	1990.6	2009.0	2099.4	2202.1	2305.1	-4.7
35	52.7	58.5	44.8	47.2	23.8	25833.0	26071.4	25466.3	25877.3	25090.9	22077.1	23785.0	24316.3	24577.9	24294.2	21747.3	25719.0	28676.6	26739.9	6.8
36	60.5	67.8	54.1	55.3	20.4	299.2	319.3	308.3	312.4	311.2	224.7	254.9	280.1	291.0	301.3	192.3	294.9	415.2	365.6	11.9
37	64.0	65.4	71.6	80.4	64.5	3180.1	3231.4	3331.6	3381.2	3322.3	2618.3	2835.8	2988.2	3127.6	3255.1	2254.1	3273.9	3303.1	3596.1	-8.9
38	69.3	78.3	77.8	88.6	79.3	222.6	245.5	263.3	284.8	281.4	155.1	178.6	200.0	222.1	257.0	113.1	290.6	348.5	337.1	3.3
39	63.7	69.5	54.3	50.7	22.1	716.4	779.0	732.3	735.4	723.8	484.9	577.3	656.3	688.7	690.0	373.5	666.7	777.2	799.0	-2.8
40	60.7	61.2	55.3	51.3	21.4	1038.5	1069.4	1078.4	1088.1	1061.7	834.8	916.0	977.4	1018.8	1014.0	673.6	1004.7	1010.6	1176.6	-16.4
41	60.6	63.2	55.9	36.0	50.6	1584.9	1656.0	1595.8	1596.6	1548.1	1271.8	1404.0	1512.5	1535.3	1504.4	1077.3	1517.3	1563.2	1730.3	-10.7
42	61.3	63.9	57.8	63.7	43.6	5620.5	5690.0	5722.6	5728.9	5646.5	4526.5	5029.9	5286.7	5470.4	5491.6	4045.6	5543.5	6810.3	6063.2	11.0
43	56.6	63.9	63.8	79.3	81.0	545.2	529.7	554.1	577.5	604.7	472.3	456.2	478.5	504.2	546.7	452.1	584.5	615.5	672.9	-9.3
44	57.9	67.5	61.4	53.3	24.7	176.0	155.7	192.2	189.4	179.7	128.5	135.4	161.9	174.5	173.7	115.3	169.1	219.3	218.6	0.3
45	64.6	67.2	70.3	74.8	38.2	430.5	443.0	468.5	476.1	459.9	335.3	371.7	395.7	425.6	439.9	281.5	445.8	446.5	536.2	-20.1
46	58.3	60.7	52.5	47.0	33.9	241.9	240.8	233.6	232.2	234.2	198.6	206.7	221.8	224.3	225.2	134.9	225.1	249.4	292.4	-17.2

Table 5.5 Training Data for Min Price Prediction – Cycle 1

Co. No.	RSI 120	RSI 60	RSI 30	RSI 15	RSI 5	UB 120	UB 60	UB 30	UB 15	UB 5	SMA 120	SMA 60	SMA30	SMA 15	SMA 5	Day1 CP	Day120 CP	MIN
1	58.6	67.0	69.1	63.4	23.1	276.0	291.0	325.0	406.9	461.0	357.3	381.9	414.3	453.0	470.5	343.8	463.3	444.3
2	64.1	69.0	70.7	81.9	59.2	1527.8	1841.5	1995.1	2154.0	2495.7	1997.3	2190.5	2310.5	2432.0	2550.8	1688.0	2601.5	2407.4
3	57.6	66.1	56.8	37.4	15.3	345.3	386.1	548.8	594.0	600.1	486.5	535.1	602.2	612.8	609.6	433.3	609.5	582.5
4	56.4	58.7	66.0	58.2	35.5	2758.2	2786.9	2869.6	3165.2	3244.6	2995.8	3064.8	3151.0	3279.7	3292.4	2842.1	3347.6	3238.3
5	63.1	70.0	76.3	59.6	64.4	2426.3	2514.0	3758.3	4618.8	4691.5	3639.7	3965.8	4617.3	4914.4	5059.7	2953.9	5252.5	4719.9
6	62.0	74.1	81.2	51.2	42.2	4440.7	4153.5	5945.0	5643.9	8802.3	6592.9	7002.8	8178.3	9007.9	9155.5	6169.4	9215.6	8354.6
7	50.9	50.3	54.3	44.3	18.8	325.9	325.1	362.0	373.5	391.7	387.8	370.1	386.9	391.8	399.3	380.1	393.9	367.4
8	47.4	54.2	57.8	60.3	33.0	388.2	390.6	446.1	462.6	495.6	493.4	456.1	483.2	495.6	508.6	560.2	510.6	490.9
9	51.9	53.2	62.7	56.8	22.4	3399.5	3379.0	3410.2	3533.6	3710.6	3671.2	3620.8	3596.9	3673.1	3751.5	3545.9	3720.5	3501.1
10	56.4	51.3	53.1	65.6	65.1	639.8	725.9	719.2	736.4	752.4	732.0	764.8	758.5	765.9	779.4	627.1	793.5	778.5
11	52.0	59.6	61.3	60.2	36.9	108.4	103.8	113.1	121.8	137.4	125.8	122.2	129.2	135.2	142.7	133.9	141.7	125.9
12	59.1	50.9	54.7	60.9	56.9	3823.2	4710.1	4664.5	4796.0	4992.7	4625.4	4984.6	4897.9	4972.0	5036.6	3911.4	5228.8	4428.1
13	57.6	60.2	62.0	50.3	35.3	1792.7	1929.8	2145.5	2418.9	2393.5	2182.1	2313.1	2461.5	2512.5	2456.6	1840.2	2480.9	2396.4
14	56.1	70.6	72.4	63.9	8.6	74.3	66.3	76.4	103.5	123.4	95.9	95.6	105.9	118.8	125.0	102.9	125.1	114.7
15	61.4	64.4	62.6	50.0	25.1	523.8	676.8	769.6	877.4	900.9	725.2	812.9	857.8	911.0	910.6	614.3	909.7	877.3
16	62.3	54.7	58.0	59.3	56.6	575.4	793.1	796.6	823.4	857.5	758.3	842.1	843.8	860.3	877.1	551.8	890.7	870.5
17	59.4	69.7	62.8	67.7	71.3	1514.0	1649.1	2105.8	2173.0	2203.8	1940.1	2094.0	2280.1	2310.4	2373.6	1835.2	2479.4	2377.8
18	56.0	52.8	52.8	46.6	11.3	2569.7	2830.9	2911.1	3046.0	3064.4	2943.8	3090.2	3074.7	3137.2	3128.3	2547.4	3135.3	2994.1
19	61.4	67.8	66.4	51.4	35.3	132.2	139.1	180.6	224.7	240.3	185.2	198.8	223.9	241.6	246.4	147.6	249.9	226.3
20	53.4	62.3	66.7	67.3	27.5	2011.1	1991.8	2008.8	2060.2	2289.7	2154.1	2155.5	2192.6	2250.7	2345.0	2170.4	2333.4	2230.4
21	60.9	69.9	62.2	47.4	65.2	38.2	48.0	64.0	65.7	66.0	54.6	61.6	67.6	68.4	69.1	47.9	68.3	68.4
22	58.8	65.8	59.8	57.9	31.0	293.5	329.3	440.3	467.0	504.4	400.0	438.2	486.4	501.2	516.4	364.0	517.2	494.5
23	64.0	60.2	62.4	65.9	59.8	779.4	1007.1	1082.2	1103.7	1152.4	1002.9	1104.0	1131.1	1146.7	1161.4	731.9	1189.8	1177.3
24	52.3	65.2	69.3	59.6	11.1	158.6	149.5	167.6	187.1	211.7	186.1	181.6	194.4	204.6	214.6	202.9	214.5	201.5
25	66.0	65.9	61.9	46.2	40.5	177.5	259.4	318.3	351.7	353.3	281.0	322.7	349.3	363.3	364.0	191.4	364.2	353.5
26	60.8	71.6	68.3	54.4	50.3	1023.6	1097.5	1669.9	1784.4	1911.9	1468.0	1606.4	1839.4	1888.2	1945.8	1355.8	1949.3	1712.9
27	62.2	74.9	76.2	74.8	64.4	772.1	765.5	942.6	1064.1	1182.3	969.4	1015.3	1120.4	1179.7	1247.5	924.3	1281.6	1234.8
28	60.1	60.8	62.8	37.5	26.4	514.6	535.6	589.8	708.3	706.4	622.5	653.4	697.1	736.8	724.0	499.0	733.6	686.0
29	58.5	61.3	58.2	62.5	17.9	5758.7	6334.9	6488.9	6955.4	7648.9	6759.5	7068.8	7260.3	7567.9	7737.2	5803.1	7559.8	7206.6
30	53.6	62.1	59.3	56.4	39.3	15104.8	14960.5	16380.7	17041.5	18064.6	16540.2	16789.2	17503.0	17896.8	18362.3	16798.8	18362.8	17060.0
31	53.1	59.7	62.0	56.2	29.2	77.5	75.8	83.4	91.0	101.9	90.0	88.6	94.9	99.3	104.4	93.6	103.8	88.9
32	56.1	66.8	68.7	61.3	33.6	59.9	53.4	58.3	75.5	96.0	76.1	74.9	81.7	91.5	100.6	80.4	99.0	83.3
33	54.0	59.6	54.3	30.4	28.9	153.5	148.9	180.0	190.5	191.2	175.4	177.5	190.5	193.6	193.9	173.4	193.9	184.4
34	53.6	43.1	55.5	47.7	17.1	1778.9	1823.4	1901.2	1922.0	1961.4	2051.1	2066.6	1978.3	1975.1	1986.7	1737.6	1991.6	1841.9
35	52.8	65.7	59.6	42.3	47.6	18518.5	18537.6	22123.3	23580.8	23441.7	21649.8	22248.1	23863.6	24434.3	24111.3	22663.5	24574.4	22773.3
36	59.7	67.9	66.1	57.1	19.1	155.3	158.5	209.1	240.4	265.8	206.9	220.4	247.6	262.7	270.9	184.8	271.5	254.7
37	61.1	58.3	60.4	60.6	59.6	2003.8	2465.9	2605.0	2671.9	2740.7	2459.4	2683.6	2720.5	2762.6	2799.1	2092.1	2861.0	2820.1
38	61.8	64.3	64.1	40.1	20.3	88.0	108.1	135.3	174.7	175.8	135.7	150.6	168.1	180.8	179.9	100.8	180.6	164.6
39	67.1	74.1	77.5	59.2	31.3	253.9	267.5	401.0	569.7	615.6	423.7	466.7	549.6	612.6	632.5	321.7	631.3	595.8
40	63.7	58.3	61.5	63.5	41.3	557.0	758.3	799.0	870.7	905.5	761.9	850.8	880.3	915.3	921.5	545.9	933.8	906.8
41	64.2	69.4	70.4	72.8	61.3	909.4	1072.0	1195.0	1319.7	1423.1	1172.3	1278.6	1357.4	1423.9	1476.2	948.2	1519.3	1420.3
42	60.0	66.7	60.5	56.6	49.1	3395.0	3850.4	4541.0	4766.2	4978.5	4248.9	4603.3	4901.9	5025.2	5106.1	3898.6	5152.5	5007.9
43	50.7	43.2	53.4	54.2	52.0	397.7	389.2	396.4	408.7	428.3	465.3	459.7	436.7	449.6	448.2	447.0	460.3	437.3
44	57.4	56.2	73.5	65.7	17.0	88.6	80.3	84.0	113.1	145.0	117.7	116.6	120.2	135.3	146.7	107.7	147.4	133.6
45	64.5	58.7	57.0	49.3	50.7	220.2	311.8	339.4	348.3	347.0	305.0	344.1	351.9	356.4	354.2	219.1	363.5	354.0
46	54.7	55.5	63.6	65.5	49.9	139.2	168.1	172.0	184.5	205.4	186.8	196.1	200.5	210.8	223.6	175.4	225.3	212.6

Table 5.6 Training Data for Average Price Prediction – Cycle 1

Co.No	RSI 120	RSI 60	RSI 30	RSI 15	RSI 5	UB 120	UB 60	UB 30	UB 15	UB 5	LB 120	LB 60	LB 30	LB 15	LB 5	SMA 120	SMA 60	SMA30	SMA 15	SMA 5	Day1 CP	Day120 CP	AVG
1	58.6	67.0	69.1	63.4	23.1	438.7	472.8	500.7	499.1	480.0	276.0	291.0	328.0	406.9	461.0	357.3	381.9	414.3	453.0	470.5	343.8	463.3	510.1
2	64.1	69.0	70.7	81.9	59.2	2466.8	2539.5	2626.0	2710.1	2605.9	1527.8	1841.5	1995.1	2154.0	2495.7	1997.3	2190.5	2310.5	2432.0	2550.8	1688.0	2601.5	2652.5
3	57.6	66.1	56.8	37.4	15.3	627.6	684.1	655.5	631.6	619.0	345.3	386.1	548.8	594.0	600.1	486.5	535.1	602.2	612.8	609.6	433.3	609.5	651.0
4	56.4	51.7	66.0	58.2	35.5	3233.3	3342.7	3432.4	3396.3	3340.3	2758.2	2786.9	2869.6	3163.2	3244.6	2995.8	3064.8	3151.0	3279.7	3292.4	2842.1	3347.6	3642.2
5	63.1	70.0	76.3	59.6	64.4	4853.1	5417.5	5476.2	5210.1	5427.8	2426.3	2514.0	3758.3	4618.8	4691.5	3639.7	3965.8	4617.3	4914.4	5059.7	2953.9	5252.5	5038.1
6	62.0	74.1	81.2	51.2	42.2	8745.1	9852.2	10411.5	9372.0	9508.6	4440.7	4153.5	5945.0	8643.9	8802.3	6592.9	7002.8	8178.3	9007.9	9155.5	6169.4	9215.6	8961.0
7	50.9	50.3	54.3	44.3	18.8	449.8	415.1	411.8	410.1	406.9	325.9	325.1	362.0	373.5	391.7	387.8	370.1	386.9	391.8	399.3	380.1	393.9	393.5
8	47.4	54.2	57.8	60.3	33.0	598.6	521.5	520.4	528.6	521.5	388.2	390.6	446.1	462.6	495.6	493.4	456.1	483.2	495.6	508.6	560.2	510.6	550.1
9	51.9	53.2	62.7	56.8	22.4	3943.0	3862.6	3783.6	3812.6	3792.4	3399.5	3379.0	3410.2	3533.6	3710.6	3671.2	3620.8	3596.9	3673.1	3751.5	3545.9	3720.5	3590.9
10	56.4	51.3	53.1	65.6	65.1	824.2	803.7	797.8	795.4	806.3	639.8	725.9	719.2	736.4	752.4	732.0	764.8	758.5	765.9	779.4	627.1	793.5	825.6
11	52.0	59.6	61.3	60.2	36.9	143.1	140.6	145.3	148.6	148.1	108.4	103.8	113.1	121.8	137.4	125.8	122.2	129.2	135.2	142.7	133.9	141.7	136.4
12	59.1	50.9	54.7	60.9	56.9	5427.5	5259.1	5131.3	5148.0	5080.4	3823.2	4710.1	4664.5	4796.0	4992.7	4625.4	4984.6	4897.9	4972.0	5036.6	3911.4	5228.8	5112.3
13	57.6	60.2	62.0	30.3	35.3	2571.5	2696.4	2777.4	2805.7	2519.7	1792.7	1929.8	2145.5	2418.9	2393.5	2182.1	2313.1	2461.5	2512.3	2456.6	1840.2	2480.9	2712.5
14	56.1	70.6	72.4	63.9	8.6	117.5	125.0	135.3	134.1	126.6	74.3	66.3	76.4	103.5	123.4	95.9	95.6	105.9	118.8	125.0	102.9	125.1	130.0
15	61.4	64.4	62.6	50.8	25.1	926.6	948.9	966.1	944.5	920.4	523.8	676.8	769.6	877.4	900.9	725.2	812.9	867.8	911.0	910.6	614.3	909.7	994.7
16	62.3	54.7	58.0	59.3	56.6	941.1	891.0	891.0	897.2	896.7	575.4	793.1	796.6	823.4	857.5	758.3	842.1	843.8	860.3	877.1	551.8	890.7	966.6
17	59.4	69.7	62.8	67.7	71.3	2366.2	2539.0	2454.4	2447.7	2543.5	1514.0	1649.1	2105.8	2173.0	2203.8	1940.1	2094.0	2280.1	2310.4	2373.6	1835.2	2479.4	2573.9
18	56.0	52.8	52.8	46.6	11.3	3317.8	3349.5	3238.3	3228.5	3192.2	2569.7	2830.9	2911.1	3046.0	3064.4	2943.8	3090.2	3074.7	3137.2	3128.3	2547.4	3135.3	3187.5
19	61.4	67.8	66.4	51.4	35.3	238.2	258.5	267.2	258.6	252.4	132.2	139.1	180.6	224.7	240.3	185.2	198.8	223.9	241.6	246.4	147.6	249.9	246.6
20	53.4	62.3	66.7	67.3	27.5	2297.2	2319.1	2376.4	2441.2	2400.2	2011.1	1991.8	2008.8	2060.2	2289.7	2154.1	2155.5	2192.6	2250.7	2345.0	2170.4	2333.4	2366.6
21	60.9	69.9	62.2	47.4	65.2	71.1	75.2	71.2	71.1	72.2	38.2	48.0	64.0	65.7	66.0	54.6	61.6	67.6	68.4	69.1	47.9	68.3	73.3
22	58.8	65.8	59.8	57.9	31.0	506.5	547.1	532.6	535.4	528.4	293.5	329.3	440.3	467.0	504.4	400.0	438.2	486.4	501.2	516.4	364.0	517.2	538.7
23	64.0	60.2	62.4	65.9	59.8	1226.5	1200.9	1179.9	1189.6	1170.5	779.4	1007.1	1082.2	1183.7	1152.4	1002.9	1184.0	1131.1	1146.7	1161.4	731.9	1189.8	1291.7
24	52.3	65.2	69.3	58.6	11.1	213.5	213.7	221.1	222.2	217.5	158.6	149.5	167.6	187.1	211.7	186.1	181.6	194.4	204.6	214.6	202.9	214.5	210.3
25	66.0	65.9	61.9	46.2	40.5	386.6	386.0	380.4	374.8	374.6	177.5	259.4	318.3	351.7	353.3	282.0	322.7	349.3	363.3	364.0	191.4	364.2	386.7
26	60.8	71.6	68.3	54.4	50.3	1912.5	2115.2	2009.8	1992.1	1979.6	1023.6	1097.5	1669.9	1784.4	1911.9	1468.8	1606.4	1839.4	1881.2	1945.8	1356.8	1949.3	1891.7
27	62.2	74.9	76.2	74.8	64.4	1166.7	1265.1	1291.1	1295.3	1312.6	772.1	765.5	942.6	1064.1	1182.3	989.4	1015.3	1120.4	1179.7	1247.5	924.3	1281.6	1336.6
28	60.1	60.8	62.8	37.5	26.4	730.4	771.2	804.4	765.3	741.5	514.6	535.6	589.8	708.3	706.4	622.5	653.4	697.1	736.8	724.0	499.0	733.6	764.0
29	58.5	61.3	58.2	62.5	17.9	7780.4	7882.7	8091.6	8180.3	7825.5	5758.7	6334.9	6438.9	6955.4	7648.9	6759.5	7068.8	7260.3	7567.9	7737.2	5808.1	7559.8	7752.8
30	53.6	62.1	59.3	56.4	39.3	17975.6	18817.9	18625.3	18752.1	18860.1	15104.8	14960.5	16380.7	17041.5	18064.6	16540.2	16789.2	17503.0	17896.8	18362.3	16798.8	18362.8	18014.6
31	53.1	59.7	62.0	56.2	29.2	102.6	108.5	106.3	107.7	107.0	77.5	75.8	83.4	91.0	101.9	90.6	89.6	94.9	99.3	104.4	99.6	103.8	97.5
32	56.8	66.8	68.7	61.3	33.6	92.3	96.4	105.1	107.4	105.3	59.9	53.4	58.3	75.5	96.0	76.1	74.9	81.7	91.5	100.6	80.4	99.0	95.4
33	54.0	59.6	54.3	30.4	28.9	197.2	206.1	201.0	196.8	196.7	153.5	148.9	180.0	190.5	191.2	175.4	177.5	190.5	195.6	195.9	173.4	193.9	193.9
34	53.6	43.1	55.5	47.7	17.1	2523.3	2309.7	2055.3	2028.1	2012.1	1778.9	1823.4	1901.2	1922.0	1961.4	2051.1	2066.6	1978.3	1975.1	1986.7	1737.6	1991.6	1958.8
35	52.8	65.7	59.6	42.3	47.6	24781.1	25958.6	25603.8	25287.8	24780.9	18518.5	18537.6	22123.3	23580.8	23411.7	21649.8	22248.1	23863.6	24434.3	24111.3	22663.5	24574.4	24237.5
36	59.7	67.9	66.1	57.1	19.1	258.4	292.3	286.0	285.0	276.0	155.3	158.5	209.1	240.4	265.8	206.9	220.4	247.6	262.7	270.9	184.8	271.5	286.0
37	61.1	58.3	60.4	60.6	59.6	2915.1	2901.3	2836.0	2853.3	2857.4	2003.8	2465.9	2685.0	2671.9	2740.7	2459.4	2683.6	2720.5	2762.6	2799.1	2092.1	2861.0	3099.1
38	61.1	64.3	64.1	40.1	20.3	183.5	199.2	200.9	186.9	184.1	88.0	108.1	135.3	174.7	175.8	135.7	150.6	168.1	180.8	179.9	100.8	180.6	227.2
39	67.1	74.1	77.5	59.2	31.3	595.5	665.9	698.3	655.5	648.4	253.9	267.5	401.0	569.7	615.6	423.7	466.7	549.6	612.6	632.5	323.7	631.3	659.8
40	63.7	58.3	61.5	63.5	41.3	966.9	943.2	961.6	959.9	937.5	557.0	758.3	799.0	870.7	905.5	761.9	850.8	880.3	915.3	921.5	545.9	933.8	990.7
41	64.2	69.4	70.4	72.8	61.3	1435.3	1485.1	1519.9	1528.2	1529.3	909.4	1072.0	1195.8	1319.7	1423.1	1172.3	1278.6	1357.4	1423.9	1476.2	948.2	1519.3	1514.3
42	60.0	66.7	60.5	56.6	48.1	5102.8	5356.1	5262.9	5284.2	5293.7	3395.0	3850.4	4543.8	4766.2	4978.5	4249.9	4603.3	4901.9	5025.2	5106.1	3898.6	5152.5	5409.9
43	50.7	43.2	53.4	54.2	52.0	532.8	530.3	477.1	490.6	462.1	397.7	389.2	396.4	408.7	428.3	465.3	459.7	436.7	449.6	448.2	447.0	460.3	508.7
44	57.4	56.2	73.5	65.7	17.0	146.7	153.0	156.3	157.5	148.3	83.6	80.3	84.0	113.1	145.0	117.7	116.6	120.2	135.3	146.7	107.7	147.4	167.4
45	64.5	58.7	57.0	48.3	50.7	389.8	376.4	364.3	364.6	361.3	220.2	311.8	338.4	348.3	347.0	305.0	344.1	351.9	356.4	354.2	219.1	363.5	416.5
46	54.7	55.5	65.6	65.5	49.9	234.4	224.2	228.9	237.1	241.8	139.2	168.1	172.0	184.5	205.4	186.8	196.1	200.5	210.8	223.6	175.4	225.3	222.8

5.3.2 Experimental Observations

The output of the experiment for various cycles has been depicted in Fig. 5.4 to Fig. 5.10. The results can be divided into two segments:

First Segment from Fig. 5.4 to Fig. 5.7 depicts the cycle wise performance of the proposed model. To evaluate the performance, the obtained output has been classified into Prediction Accuracy Bands (PABs) of size ±3%, ±5%, ±7%, ±10% and ±15%. The results of the min prediction are quite remarkable and majority of the stocks are covered within ±10% band. Next performer is average price prediction. Max price prediction comes in the last. The reason for the same is that the Indian stock markets have seen significantly high upward trend in the post pandemic (COVID) scenario. The beauty of the output is that it is consistent amongst all the cycles.

Second Segment from Fig. 5.8 to Fig. 5.10 depicts the consolidated performance which quantifies the average performance across all cycles and endorses the above mentioned results.

Fig. 5.11 to Fig. 5.13 show the stock wise actual performance through the use of scatter diagram. The purpose of these diagrams is to show that a significant portion of predictions results are bound in the narrow band and numbers of outliers are very few ensuring the good fitness of results. Since the stock price are different for different stocks, a stock may have the price as 230 and other as 15600 therefore while plotting these graph the actual price (max, min or average) was normalized to 100. Fig. 5.13 shows a particular case where the numbers of outliers are quite high. This is the worst graph we obtained. But such things are quite likely in a random environment like stock market.

Fig. 5.4: Prediction Accuracy in Cycle 1

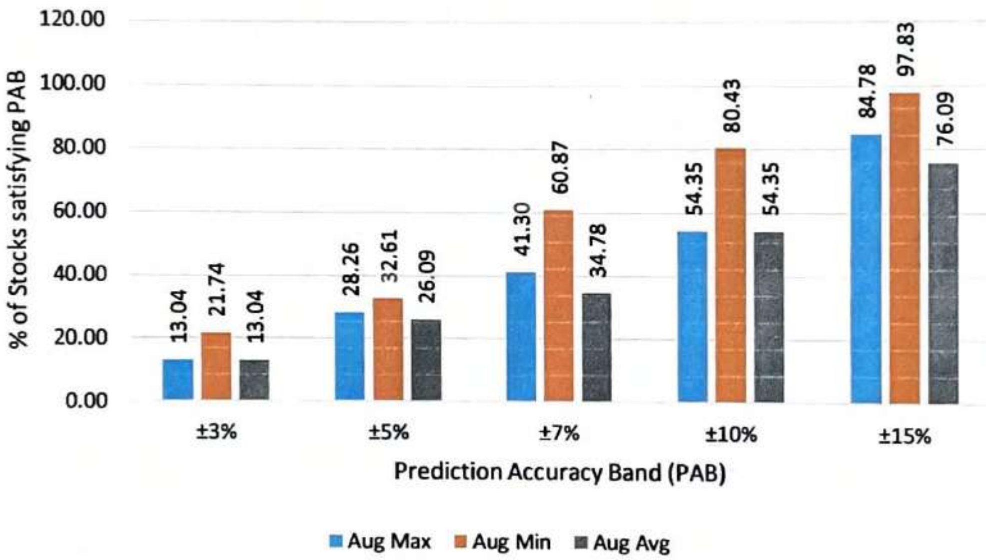


Fig. 5.5: Prediction Accuracy in Cycle 2

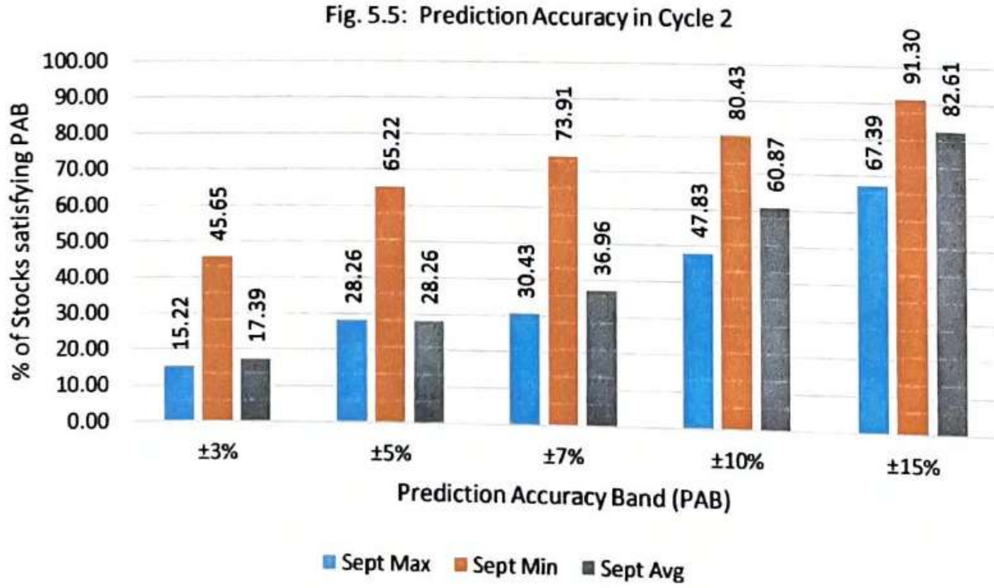


Fig. 5.6: Prediction Accuracy in Cycle 3

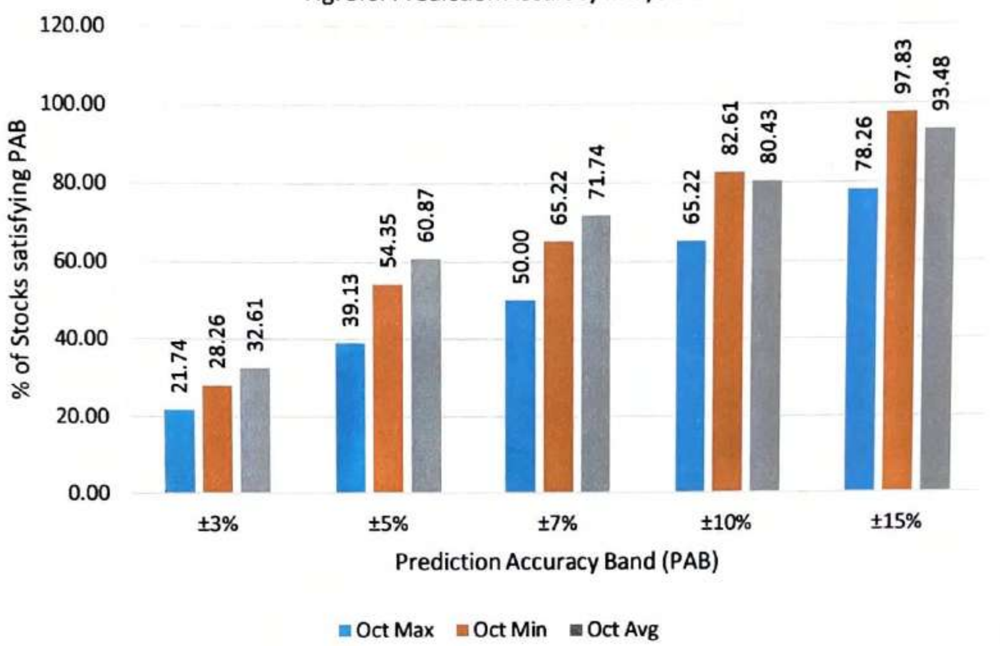


Fig. 5.7: Prediction Accuracy in Cycle 4

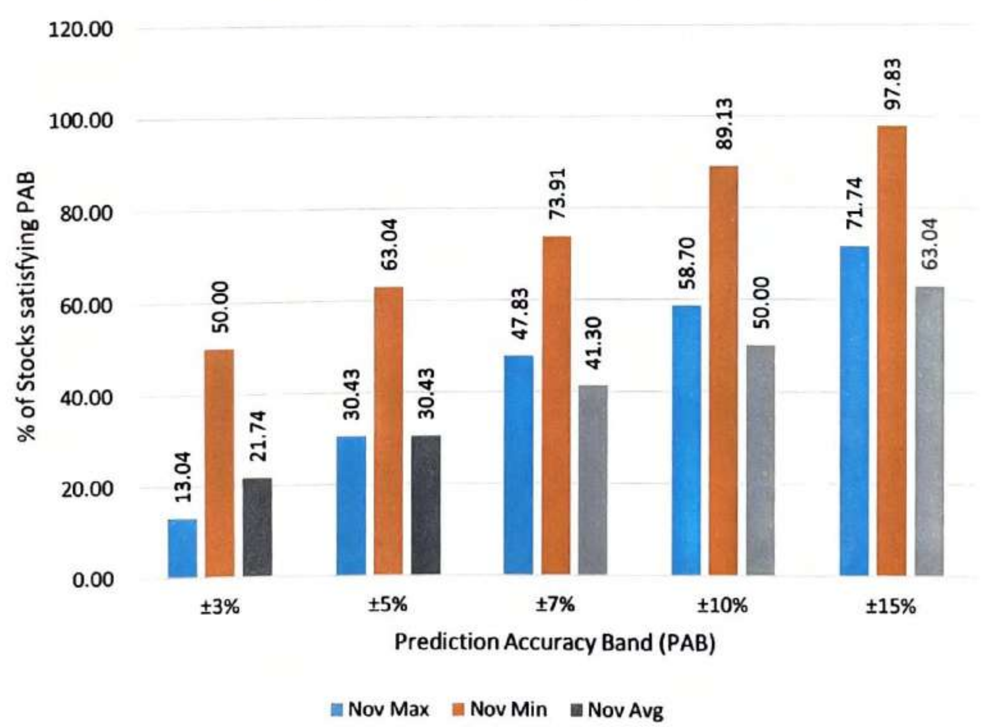


Fig. 5.8: Maximum Price Prediction Trend across all cycles

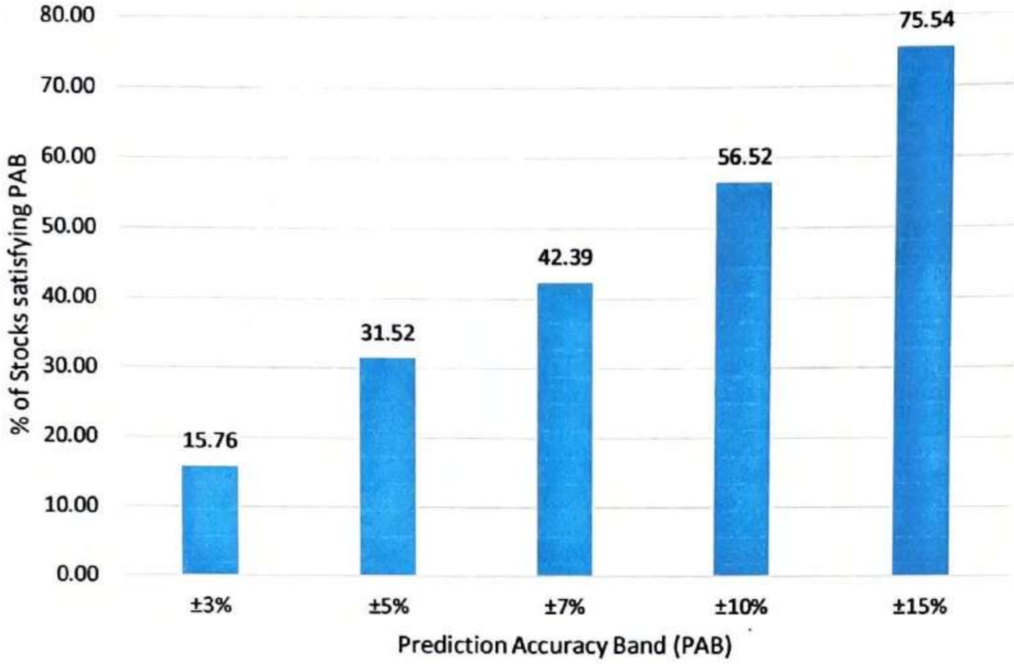
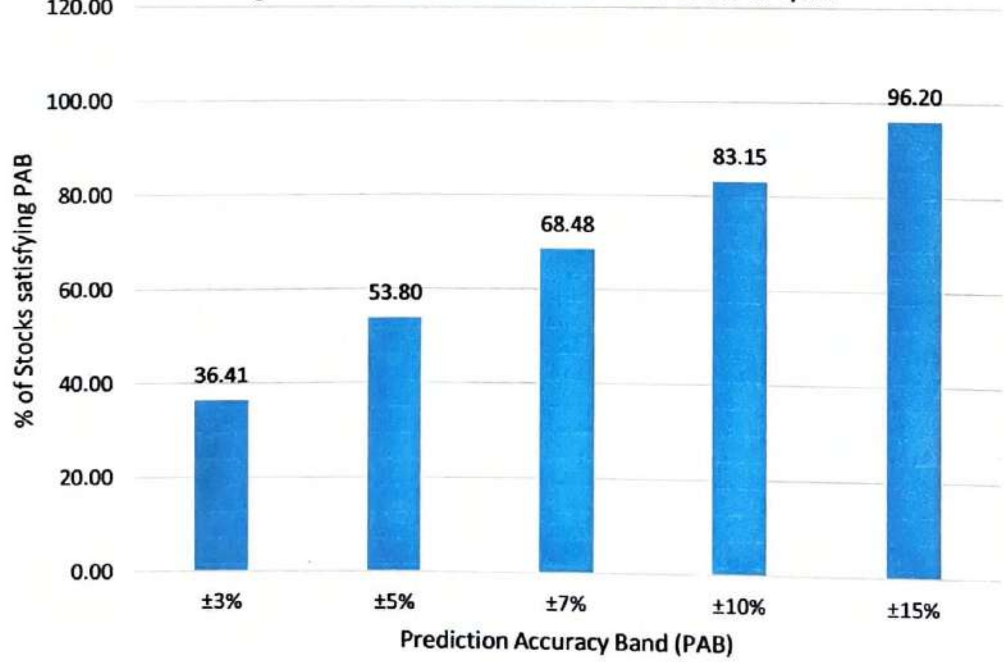
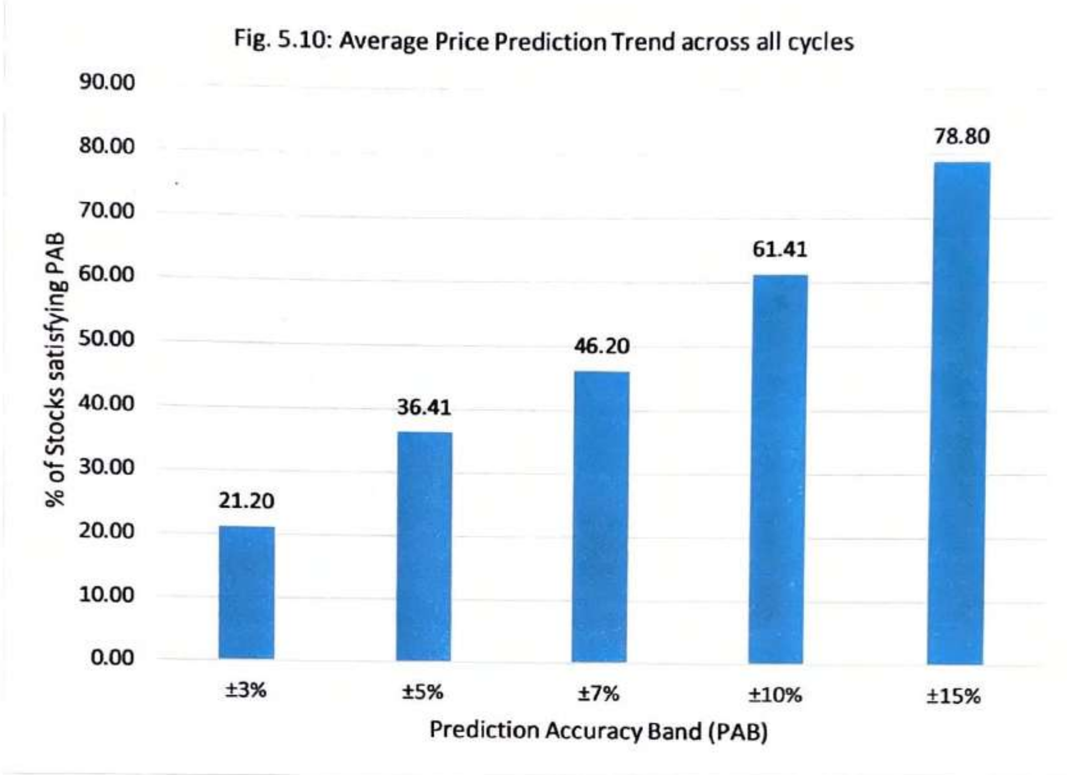


Fig. 5.9: Minimum Price Prediction Trend across all cycles





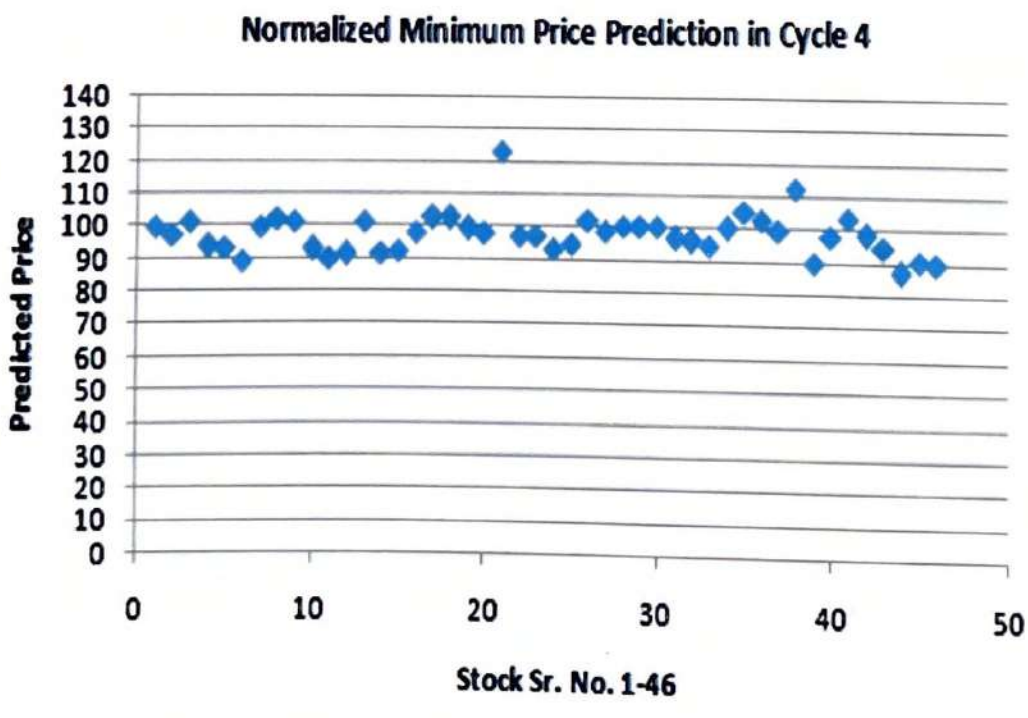


Fig. 5.11: Normalized Minimum Price Prediction in Cycle 4

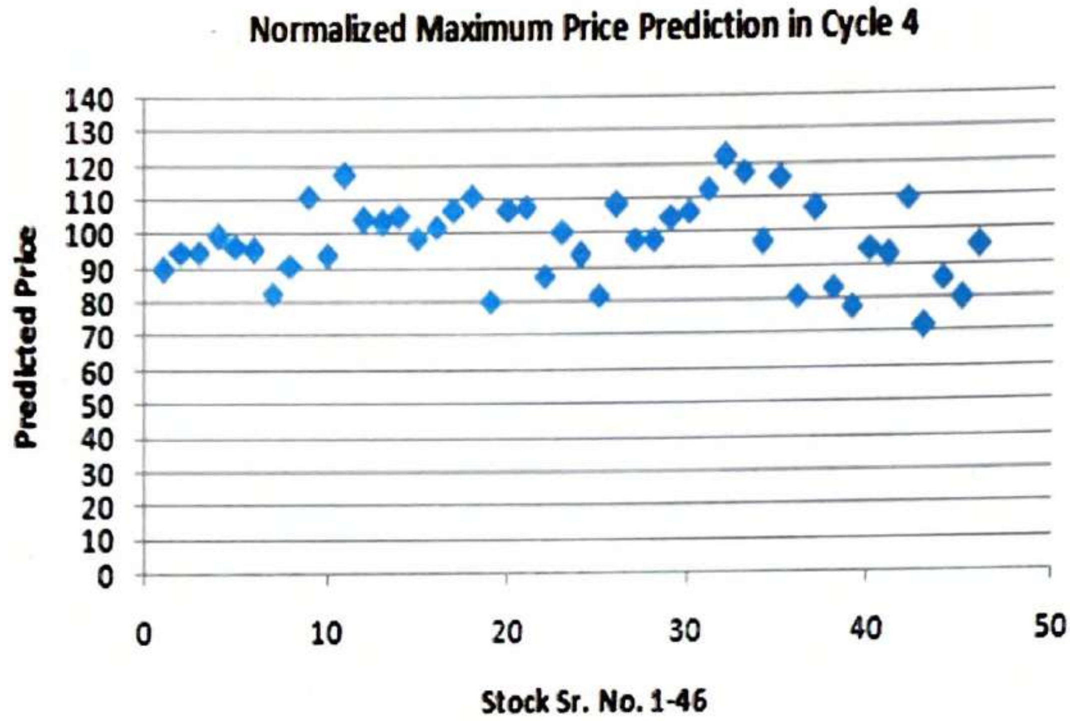


Fig. 5.12: Normalized Maximum Price Prediction in Cycle 4

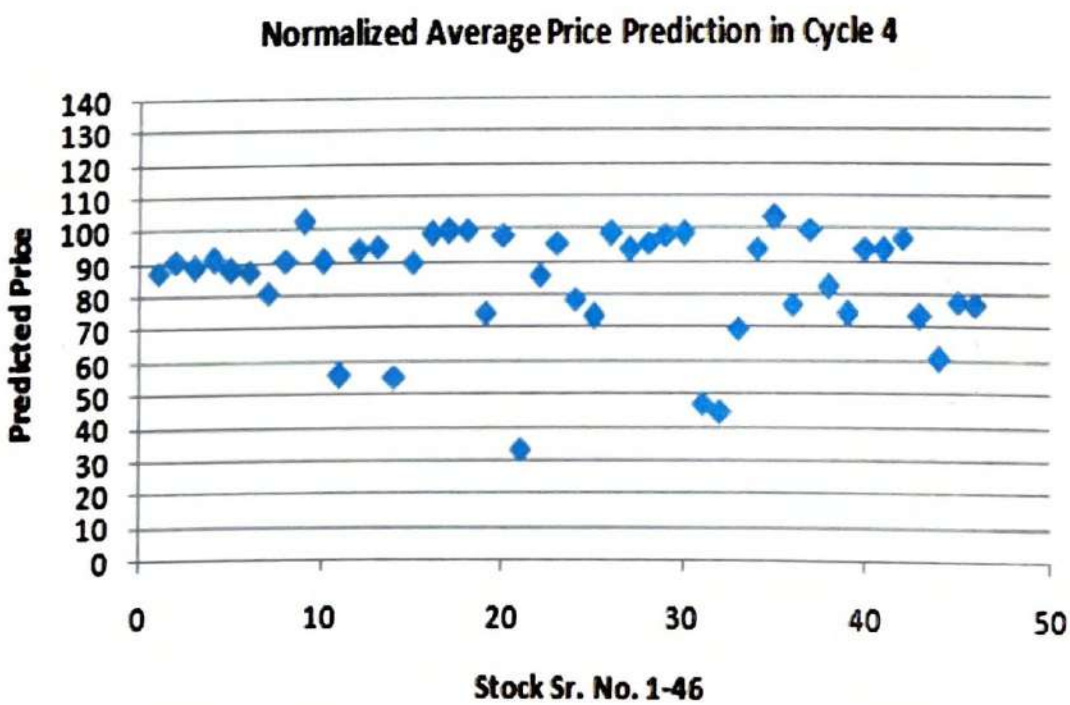


Fig. 5.13: Normalized Average Price Prediction in Cycle 4

5.4 OPPORTUNITY SIGNAL GENERATION

The goal of this system is automated generation of an opportunity signal (indicating the advice to buy/sell/wait) with appropriate rewards. To accomplish the same, following methodology was adopted:

- The range between maximum and minimum prediction was normalized to [0, 100] using a scaling factor. For example, for a stock say X, if the minimum predicted value is a and maximum predicted value is b then the Scaling Factor (S) for normalization is 100/ (b-a). Now, for a current price Cp, the normalized value will be (Cp-a)*S
- This normalized range [0,100] is now fuzzified to three fuzzy sets namely Top, Mid and Bottom as shown in the Fig. 5.14. For a given current price, a situation can be classified into one of the above fuzzy sets.
- Depending upon the fuzzy set(s) and the memberships obtained, the fuzzy sets Top/Mid/Bottom can be classified into sell/wait/buy opportunity call respectively on the basis of reward points.
- The simplest reward point system would be $100*\mu$, where μ is the membership in the fuzzy set. Such a system would create 50 or more reward points as and when μ exceeds 0.5.However, to make the system safer it is advisable to adopt a formula like $100*\mu^k$ with $k>1$ so as to weaken the opportunity call generated at lower memberships (as per the expert advice). Table 5.7 illustrates the mechanism described above for k=2. An increment in the value of k reduces the number of opportunities generated but increases the gain per opportunity thereby maintaining the overall gain reducing the risk factor at the same time. The same has been shown in the next section wherein the opportunities have been explored with value of k from 1 to 7.

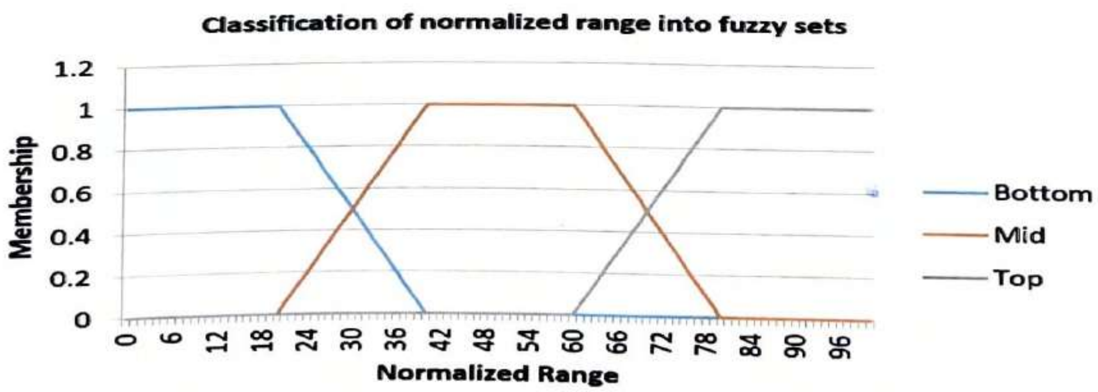


Fig. 5.14: Classification of normalized range into fuzzy sets

Table 5.7: Opportunity Signal Generation illustration

Company	Predicted High price (b)	Predicted low price (a)	Range (b-a)	Normalization factor S= 100/(b-a)	Current Price Cp	Normalized Value	Fuzzy set(s) with membership	Opportunity Signal (Rewards at k=2)
A	18000	16500	1500	0.07	17234	48.93	Mid (1.0)	No action (100)
B	1200	750	450	0.22	903	34.00	Bottom (0.3) Mid (0.7)	Buy (9) No action (49)
C	1100	700	400	0.25	988	72.00	Mid(0.4) Top(0.6)	No Action (16) Sell(36)
D	1500	1250	250	0.4	1310	24.00	Bottom(0.8) Mid(0.2)	Buy (64) No Action (4)
E	1000	800	200	0.50	980	90.00	Top(1)	Sell (100)

Table 5.8: Opportunity Analysis at k=2

	No of Stocks under Consideration = 46						
	CASE 1	**CASE 2**	**CASE 3**	**CASE 4**	**Profit**	**Loss**	**Net Gain**
Cycle 1	18	18	2	8	4250.8	-1576	2674.99
Cycle 2	12	20	4	10	5839.1	-1421	4417.85
Cycle 3	24	7	4	11	2724	-151.5	2604.98
Cycle 4	15	13	7	11	2124.5	-98.26	2026.22
Average Occurrence	17.25	14.5	4.25	10	3734.6	-811.7	2931.01
Success (%)	37.5	31.5	9.24	22	100	-21.78	78.22

Table 5.9: Detailed Computation Scenarios for Cycle 1 at k=2

Co. No.	Ac_MIN	Ac_MAX	Pr_MIN	Pr_MAX	Range	BO POINT	SO POINT	BO status	SO status	RETURNS	Price on 5/Mar/21
1	509.70	752.45	514.48	639.11	124.63	546.71	606.89	Y	Y	60.17	748.95
2	2277.20	2596.65	2597.73	3029.33	431.60	2709.34	2917.72	N	N	0.00	2,387.10
3	632.10	794.00	668.55	786.68	118.13	699.10	756.13	Y	Y	57.03	730.75
4	3798.70	4237.45	3452.04	4002.53	550.50	3594.39	3860.18	N	Y	0.00	3,852.85
5	4734.55	5779.85	4820.55	5583.47	762.93	5017.84	5386.18	Y	Y	368.34	5,444.90
6	8722.75	10537.70	8460.35	9748.72	1288.37	8793.52	9415.55	Y	Y	622.02	9,958.10
7	380.60	469.00	402.60	493.74	91.14	426.17	470.17	Y	N	35.63	461.8
8	532.20	608.85	563.40	688.36	124.96	595.71	656.04	Y	N	-62.91	532.8
9	3331.20	3613.55	3501.86	4040.18	538.32	3641.07	3900.97	N	N	0.00	3,491.75
10	787.05	863.40	804.50	964.08	159.58	845.77	922.82	Y	N	-40.62	805.15
11	125.90	156.10	139.14	176.65	37.51	148.84	166.95	Y	N	2.26	151.1
12	4404.70	5082.40	4927.87	5761.24	833.37	5143.38	5545.73	N	N	0.00	4,492.50
13	2482.35	2972.60	2710.96	3185.99	475.03	2833.80	3063.15	Y	N	-246.05	2,587.75
14	125.30	150.80	131.45	158.41	26.96	138.42	151.44	Y	N	8.73	147.15
15	1004.00	1350.40	996.97	1176.63	179.66	1043.43	1130.17	Y	Y	86.74	1,339.35
16	909.45	985.80	971.82	1156.43	184.61	1019.56	1108.69	N	N	0.00	941.5
17	2377.80	2860.45	2566.76	2964.53	397.77	2669.63	2861.67	Y	N	-125.73	2,543.90
18	3224.00	3584.00	3082.53	3636.13	553.61	3225.69	3492.97	Y	Y	267.28	3,461.70
19	226.30	359.35	251.16	319.65	68.49	268.87	301.94	Y	Y	33.07	337.85
20	2132.05	2409.35	2304.64	2648.63	343.98	2393.60	2559.67	Y	N	-192.50	2,201.10
21	68.84	83.26	73.49	79.31	5.82	74.99	77.80	Y	Y	2.81	82.28
22	522.35	673.95	548.07	647.74	99.67	573.84	621.96	Y	Y	48.12	609.45
23	1239.05	1343.55	1305.89	1526.00	220.11	1362.81	1469.08	N	N	0.00	1,039.90
24	203.25	234.35	214.57	277.16	62.58	230.76	260.97	Y	N	-22.21	208.55
25	366.85	428.15	386.55	481.11	94.56	411.00	456.66	Y	N	-4.60	406.4
26	1712.95	2019.65	1887.49	2144.60	257.11	1953.98	2078.11	Y	N	-31.48	1,922.50
27	1334.70	1566.45	1357.13	1543.97	186.84	1405.45	1495.65	Y	Y	90.21	1,463.55
28	749.60	928.40	770.72	938.61	167.89	814.13	895.19	Y	Y	81.06	840.4
29	6866.15	8048.85	7456.63	8864.92	1408.30	7820.81	8500.74	Y	N	-571.81	7,249.00
30	16096.30	17589.60	17063.90	19737.29	2673.39	17755.24	19045.95	N	N	0.00	16,782.90
31	88.95	112.70	99.07	117.38	18.31	103.81	112.64	Y	Y	8.84	108.7
32	88.30	119.05	98.09	111.95	13.86	101.67	108.37	Y	Y	6.69	114.95
33	184.45	236.50	195.70	256.80	61.10	211.50	241.00	Y	N	-46.31	165.19
34	1841.95	2202.10	1993.22	2305.06	311.85	2073.86	2224.42	Y	N	104.84	2,178.70
35	22773.35	28676.60	22987.02	26739.89	3752.87	23957.51	25769.40	Y	Y	1811.88	27,848.60
36	275.65	415.20	287.62	365.59	77.96	307.79	345.43	Y	Y	37.64	383.65
37	2894.30	3303.10	3058.85	3596.08	537.23	3197.78	3457.15	Y	N	-189.73	3,008.05
38	262.70	348.50	228.45	337.05	108.60	256.54	308.97	N	Y	0.00	325.15
39	601.00	777.15	693.01	798.99	105.98	720.42	771.58	Y	Y	51.17	733.3
40	918.85	1010.60	999.83	1176.61	176.78	1045.54	1130.89	N	N	0.00	959.7
41	1405.25	1563.15	1510.46	1730.34	219.88	1567.32	1673.48	N	N	0.00	1,474.20
42	5327.25	6810.30	5238.84	6063.22	824.38	5452.02	5850.03	Y	Y	398.01	6,810.30
43	534.10	615.45	506.18	672.90	166.71	549.29	629.78	Y	N	42.26	591.55
44	161.25	219.30	164.75	218.60	53.85	178.67	204.67	Y	Y	26.00	219.25
45	410.30	446.45	411.26	536.20	124.93	443.57	503.89	Y	N	-22.72	420.85
46	201.00	249.45	217.94	292.39	74.44	237.19	273.14	Y	N	-19.14	218.05

Table 5.10 Cycle 1

K	CASE 1	CASE 2	CASE 3	CASE 4	Gain	Loss	Net
1	21	14	2	9	3695.63	-1576.36	2119.273
2	18	18	2	8	4250.802	-1575.81	2674.991
3	16	19	3	8	4261.828	-1504.51	2757.321
4	16	20	3	7	4416.822	-1552.58	2864.237
5	16	20	3	7	4513.081	-1526.97	2986.111
6	16	20	3	7	4579.973	-1509.17	3070.803
7	16	20	3	7	4627.287	-1496.58	3130.707

Table 5.11 Cycle 2

K	CASE 1	CASE 2	CASE 3	CASE 4	Gain	Loss	Net
1	15	19	3	9	4911.198	-1655.89	3255.309
2	12	20	4	10	5839.135	-1421.28	4417.852
3	11	22	4	9	6239.678	-1519.58	4720.095
4	9	24	5	8	6407.916	-1605.61	4802.308
5	9	24	5	8	6544.296	-1564.79	4979.511
6	8	23	6	9	6414.57	-1536.42	4878.153
7	8	23	5	10	6479.152	-1516.35	4962.801

Table 5.12 Cycle 3

K	CASE 1	CASE 2	CASE 3	CASE 4	Gain	Loss	Net
1	25	6	3	12	2326.338	-148.893	2204.388
2	24	7	4	11	2723.972	-151.51	2604.981
3	24	8	4	10	2919.586	-147.211	2791.468
4	24	8	4	10	3027.005	-144.849	2902.862
5	24	8	4	10	3093.717	-143.383	2972.043
6	24	8	4	10	3140.76	-143.048	3020.119
7	24	8	4	10	3174.207	-142.983	3054.123

Table 5.13 Cycle 4

K	CASE 1	CASE 2	CASE 3	CASE 4	Gain	Loss	Net
1	15	12	7	12	1707.628	-106.433	1601.195
2	15	13	7	11	2124.479	-98.2612	2026.217
3	15	13	6	12	2309.646	-84.7642	2224.882
4	15	13	6	12	2411.327	-77.3526	2333.975
5	15	13	6	12	2474.5	-72.7726	2401.727
6	15	13	6	12	2519.296	-70.4859	2448.81
7	15	13	6	12	2550.98	-68.8685	2482.112

5.5 VALIDATION OF RESULTS

To validate the results, every cycle was checked for the occurrence of buying / selling opportunity for each of the 46 stocks. If the opportunity occurred, then one number of stock of the company was bought / sold. The data so obtained has been shown in Table 8. Let us represent the occurrence / non-occurrence of the buying / selling opportunity by "Y" / "N" and BO / SO respectively. Now there can be four cases:

Case 1: BO = "Y" and SO = "Y"

In this case both buying and selling opportunities have occurred, during the validation period of 30 working days, leading to profit generation.

Case 2: BO = "Y" and SO = "N"

In this case only buying opportunity has occurred. Since selling opportunity did not occur, there are two possible options: selling can be postponed to subsequent cycle(s) or selling can be done on the last day of the cycle, be it profit or loss. We have gone for the second option in order to finish the task with in the same cycle.

Case 3: BO = "N" and SO = "Y"

Here no buying opportunity has occurred so there is no possibility for selling leading to no transaction.

Case 4: BO = "N" and SO = "N"

Here neither buying nor selling opportunity has occurred, so there is no transaction. Table 5.8 shows the opportunity analysis at k=2. The analysis in Table 5.8 shows that in 37.5% cases the both buying and selling opportunities were generated leading to majority profit. In 31.5% cases, selling opportunity did not occur leading to distress selling on the last day of the cycle which resulted in the loss in many cases but there was overall gain. In the overall scenario, on the average 21.78% profit (809/3713) was wiped out due to distress selling resulting in overall gain of 78.22% (2911/3713). Table 5.9 shows the detailed computation scenario for cycle 1 at k=2.

Table 5.10. 5.11, 5.12 and 5.13 show real net gain obtained in various cycles by varying the value of k from 1 to 7. The results endorse the hypothesis that increase in the value of k, decreases the risk factor without affecting the net gain. Overall gain across all the cycles show that the proposed mechanism is quite trust worthy.

The proposed work is able to make a reasonable stock price band prediction for the upcoming one and a half month with quite significant accuracy. In Indian Stock Market, to prevent the undesirable manipulations of stock prices a circuit of 5% or 10% is imposed on the stock price on the daily basis. Most of the NIFTY50 stocks are in the 10% band. Thus, for a market undergoing a strong trend, whether upward or downward, it is not uncommon to have a change in the range of 30% to 60% in a period of one month and a half. The post pandemic rally, after September 2020, has raised stock price 2 to 3 folds in the period under review (spread across all cycles) for a quite a significant fraction of the popular stocks. 50-60% rise has been seen in majority of stocks. The NSE bench mark index has risen by one and a half times. Under the circumstances, the predictions made by our system are quite appropriate and reliable. Though the scenario undertaken is Indian stock market, yet the work can be utilized to any stock market across the globe. In the literature, the predictions are related to few prominent stocks, in our case almost entire NIFTY50 band of Indian stock market has been taken into consideration. The concern is not only the single day price but the expected price band in the upcoming future to effectively sense buying/wait/selling opportunities. The system has been able to successfully generate the opportunity signal with reasonable net gains. It was observed that larger the value of k, lesser is the risk and more is the gain as well. Thus a larger value of k is desirable but an extremely high value of k (>10) can result in missing of the opportunities to a large extent.

The price band results obtained in the proposed model can be classified in the good fit category as they are consistent across all the cycles. We tried to extend the work with the inclusion of more features through the inclusion of their component features. This led to the deviation of the results due to overfitting. It will be a good exercise for new researchers if they can include more component features without falling in the trap of overfitting.

CHAPTER-VI

A DYNAMICALLY ADAPTING FRAMEWORK FOR STOCK PRICE PREDICTION

Recent past has seen a flood of research papers relating to stock price predictions [7, 8, 9, 12, 13, 14, 15]. The reasons for this flood include: ease of data availability in electronic form, advances in machine learning, availability of open source APIs for these algorithms. The hypothesis behind all these works is that the machine learning will identify hidden patterns in data, capture trend and there will be a prediction breakthrough [10, 18]. Thus research community has tried to explore different strategies based upon deep learning networks [19, 20, 21], different types of signal/activation functions [39], [74], incorporation of impact of news [54] and social media sentiments etc. [52]. But the success is still far off. The reason for the same is that the share market is not totally governed by mathematical or logical patterns but it is also driven by other factors such as human manipulations, greed, social & economic sentiments and announcement of govt. policies etc. All these factors can affect the different stocks in different manners which may be good/ better for one stock and bad/worse for the other.

We, therefore, are of the view that it is inane to look for a generic, static and universally applicable prediction strategy that shall be appropriate for all the stocks in all the stock markets across the globe. Every stock in the stock market has its own individuality and therefore requires its own distinctive treatment. Finding such treatments for each stock in every stock market is a voluminous and tedious task. If one can design a stock specific customized strategy keeping in view the individuality of the stock, the results will be more accurate. This work is an effort in this direction. The work involves selection of a pair of stock specific prediction strategy from a set of created options. This selected pair is subjected to error correction learning for the further reduction of error thereby minimizing the prediction error.

6.1 Objectives

- To create a Stock specific strategy instead of the general one to be applied to all stocks.

- To improve the input data quality through candidate selection with best proximity.
- To be able to reduce the prediction error further by fine tuning using back propagation.
- To be able to generate multiple predictions instead of relying upon the single one.

6.2 PROPOSED FRAMEWORK

6.2.1 Framework Description

The proposed mechanism involves taking up of the past price data for the stock under consideration from 1st July 2016 to 30 June 2019 on daily basis [17]–[20]. The data so obtained included Stock Price, Sector Index and Nifty 50 index at the close of the day. This data was used for the supervised learning with past n days data as input and the stock close price of n+1th day as the output using sliding window mechanism as shown in the Fig. 6.1.

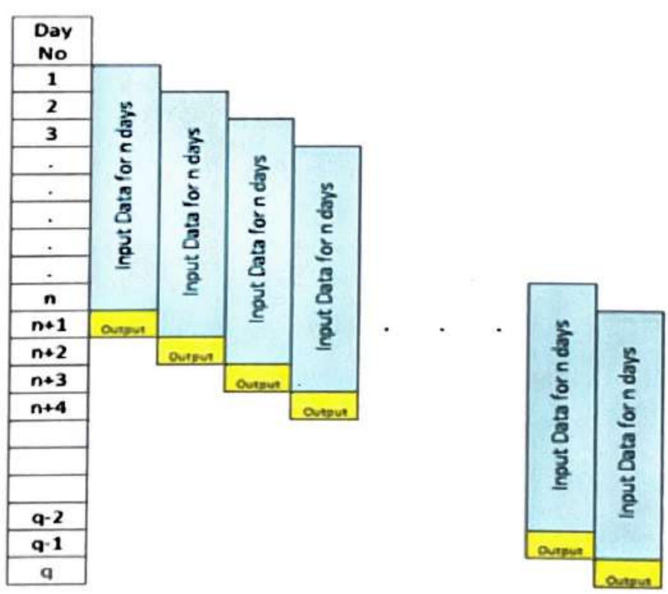


Fig. 6.1 Sliding Window of size n for price prediction

A primary set of predictions Sp is created using a sliding window of 5, 10, 20 and 30 days with LSTM on Python platform. Sp is subjected to the statistical extensions using max, min and mean operations leading to the creation an additional set of pre-diction called mix bag predictions Sm. Union of Sp and Sm is done to create comprehensive set of predictions Sc.

All these predictions were compared with the actual price of the stock in future and coefficient of correlation (CO) and mean absolute error (MAE) is computed. Using this information, two top candidate predictions C1 and C2 are chosen. The basis for the Choice of C1 and C2 is MAE and CO is used to ensure the minimum correlation between the actual price and the candidate prediction. If CO is less than 0.95 the candidate prediction is not used.

Now these candidate predictions also have some error though quite small com-pared to others. Next task is to reduce the error in candidate predictions to the mini-mum level in order to bring it closest to the actual price of the stock. To accomplish this, back propagation algorithm was applied with two candidate predictions as input and the actual close price as the target. This task was performed on WEKA tool using Linear Regression. The results so obtained were tested and it was found that there has been a marked improvement in the prediction accuracy. Fig. 6.2 shows the complete diagram of the design and implementation phase of the experiment conducted by us. Now the details of candidate predictions (c1 and c2) and the back propagation networks (ANN id) details were recorded for the stock under consideration in the data base.

The task is repeated for a set of stocks and for each stock C1,C2 and ANN ID is recorded as shown in Table 6.1.The details of the Table1 were used for making the predictions about the stock in the prediction phase as shown in Fig. 6.3.

Initially the work was performed for next day prediction but the quality of the results encouraged us to extend the experiment for generating predictions for next week, next fortnight and next month. So in general the proposed work uses n days of stock price, stock index and sector index as input and the stock price on (n+k)th day as the output with k=1,7,15,30. The results so obtained have been discussed in the next section.

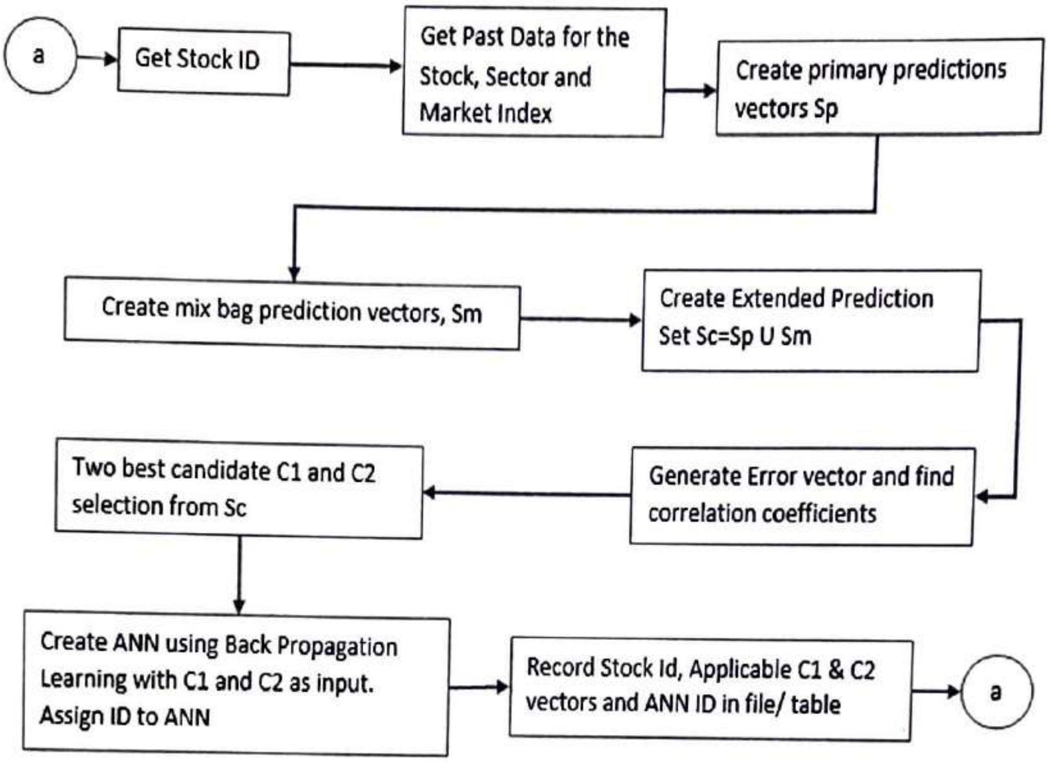


Fig. 6.2 Design and Implementation Phase to be repeated periodically for continuous updation

Table 6.1 Recording of ANN ID

Stock Id	C1	C2	ANN ID
101	Pred 5	Pred Max	A_1
102	Pred10	Pred 30	A_2
103	Pred 30	Pred Avg	A_3
.			
.			
N	Pred 20	Pred 30	A_N

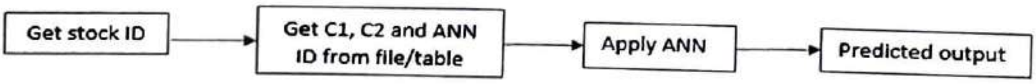


Fig. 6.3: Prediction Phase

6.2.2 Framework Exemplified

We demonstrate the working of the proposed mechanism with the help of the Bajaj Auto Stock. The data was collected for the past three years (from 1st July 2016 to 30 June 2019) [17]–[20] and comprehensive set of prediction of the stock price are created for the test period from 1st July 2019 to 6th November, 2020 using the above mentioned mechanism as shown in Table 6.2. Here Pred5, Pred10, Pred20 and Pred30 indicate the predictions obtained on the basis of past 5,10,20,30 days data. Also we have statistically extended set of predictions:

PredMin = min(Pred5, Pred10, Pred20,Pred30)

PredMax = max(Pred5, Pred10, Pred20,Pred30)

PredAvg = average(Pred5, Pred10, Pred20,Pred30)

Table 6.2: Primary and Statistically extended predictions for Bajaj Auto

Date	Close Price	Pred 5	Pred 10	Pred 20	Pred 30	Pred-Min	Pred-Max	Pred-Avg
7/1/2019	2910.7	2904.18	2922.27	2913.55	2912.27	2904.18	2922.27	2913.07
7/2/2019	2886.25	2886.18	2893.69	2900.07	2906.18	2886.18	2906.18	2896.53
7/3/2019	2895.35	2894.43	2908.39	2904.13	2907.72	2894.43	2908.39	2903.67
7/4/2019	2895.55	2894.88	2905.16	2904.95	2908.21	2894.88	2908.21	2903.3
7/5/2019	2838.2	2837.35	2852.05	2853.78	2858.79	2837.35	2858.79	2850.49
7/8/2019	2779.9	2775.3	2802.36	2795.26	2798.6	2775.3	2802.36	2792.88
7/9/2019	2784.75	2779.55	2808.05	2791.46	2792.5	2779.55	2808.05	2792.89
7/10/2019	2741.9	2739.95	2769.01	2755.99	2760.9	2739.95	2769.01	2756.46
7/11/2019	2711.65	2708.03	2747.74	2723.25	2727.99	2708.03	2747.74	2726.75
7/12/2019	2721.75	2718.77	2759.56	2727.71	2730.99	2718.77	2759.56	2734.26
7/15/2019	2714.25	2715.27	2754.02	2724.67	2729.86	2715.27	2754.02	2730.96
7/16/2019	2732.25	2734.73	2772.76	2739.94	2744.94	2734.73	2772.76	2748.09
7/17/2019	2696.45	2699.18	2738.6	2713.39	2720	2699.18	2738.6	2717.79
7/18/2019	2632.85	2630.59	2681.44	2653.49	2658.91	2630.59	2681.44	2656.11
7/19/2019	2557.55	2550.11	2611.8	2578.24	2581.02	2550.11	2611.8	2580.29

Table 6.3: Error between different predictions for Bajaj Auto

Date	Close price	AE-5	AE-10	AE-20	AE-30	AE-MIN	AE-MAX	AE-AVG
7/1/2019	2910.7	6.52	11.57	2.85	1.57	6.52	11.57	2.37
7/2/2019	2886.25	0.07	7.44	13.82	19.93	0.07	19.93	10.28

7/3/2019	2895.35	0.92	13.04	8.78	12.37	0.92	13.04	8.32
7/4/2019	2895.55	0.67	9.61	9.4	12.66	0.67	12.66	7.75
7/5/2019	2838.2	0.85	13.85	15.58	20.59	0.85	20.59	12.29
7/8/2019	2779.9	4.6	22.46	15.36	18.7	4.6	22.46	12.98
7/9/2019	2784.75	5.2	23.3	6.71	7.75	5.2	23.3	8.14
7/10/2019	2741.9	1.95	27.11	14.09	19	1.95	27.11	14.56
7/11/2019	2711.65	3.62	36.09	11.6	16.34	3.62	36.09	15.1
7/12/2019	2721.75	2.98	37.81	5.96	9.24	2.98	37.81	12.51
7/15/2019	2714.25	1.02	39.77	10.42	15.61	1.02	39.77	16.71
7/16/2019	2732.25	2.48	40.51	7.69	12.69	2.48	40.51	15.84
7/17/2019	2696.45	2.73	42.15	16.94	23.55	2.73	42.15	21.34
7/18/2019	2632.85	2.26	48.59	20.64	26.06	2.26	48.59	23.26
7/19/2019	2557.55	7.44	54.25	20.69	23.47	7.44	54.25	22.74

Table 6.3 shows the error between actual and predicted price. Here AE-5, AE-10, AE-20, AE-30, AE-MIN, AE-MAX and AE-AVG show the absolute error for the predictions Pred5, Pred10, Pred20, Pred30, PredMin, PredMax and PredAvg respectively. Table 6.4 shows the Mean Absolute Error (MAE) for all the predictions in Sc for Bajaj Auto and some other stocks. On the basis of lowest values of MAE-20 and MAE-30, Pred20 and Pred30 become the candidate predictions for the purpose of further error reduction. On the basis of similar argument, Pred20 and PredMin become the candidate predictions for Eicher Motors. Table 6.5 shows the correlation between Close Price Vector and the various Price Prediction Vectors. For example, CO-5 indicates the value of correlation coefficient between the close price and Pred5. Table 6.6 shows the candidate predictions C1 and C2 for the different stocks along with their ANN-ID.

Table 6.4 : Mean Absolute Error of Different Automobile Companies for Next Day Prediction

SECTOR	Stock	MAE-05	MAE-10	MAE-20	MAE-30	MAE-MIN	MAE-MAX	MAE-AVG
Automobile	Bajaj Auto	21.9	37.4	12.4	14.1	14.3	36.2	18
	Eicher Motors	72.7	99	39.3	54.9	39.4	103.8	46.1
	Heromotocorp	26.4	86.9	118.9	49.9	48	121.5	44.1
	Mahindra	11.4	20.3	16.4	17.5	19.3	20.4	6.1
	Maruti	76.7	68.1	189.4	172.4	185.3	173.5	78.9
	Tata Motors	8	22.3	23.5	14.8	8	24.7	13.2

Table 6.5: Correlation of Different Automobile Companies for Next Day Prediction

SECTOR	Stock	CO-05	CO-10	CO-20	CO-30	CO-Min	CO-Max	CO-Avg
Automobile	Bajaj Auto	0.985	0.997	0.999	0.998	0.998	0.991	0.998
	Eicher Motors	0.994	0.996	0.998	0.996	0.998	0.995	0.998
	Heromotocorp	0.997	0.989	0.988	0.993	0.995	0.988	0.997
	Mahindra	0.997	0.999	0.994	0.994	0.996	0.999	0.998
	Maruti	0.994	0.997	0.984	0.977	0.987	0.986	0.992
	Tata Motors	0.993	0.998	0.991	0.985	0.993	0.993	0.995

Table 6.6: Candidate Selection of Automobile Sector for Next Day

Stock Id	C1	C2	ANN ID
Bajaj Auto	Pred20	Pred 30	**BA-1**
Eicher Motors	Pred20	Pred-min	EM-1
Heromotocorp	Pred-5	Pred-ave	HM-1
Mahindra	Pred-5	Pred-ave	MH-1
Maruti	Pred-5	Pred-10	MR-1
Tata Motors	Pred-5	Pred-min	TM-1

After describing the working of the proposed framework through illustration let us take up the actual experiment.

6.3 THE EXPERIMENT AND THE RESULTS

To assess the credibility of the proposed mechanism, the past price data was taken from the period 1/7/2019 to 6/11/2020. This data was collected from various websites [16, 30, 31, 32]. This data was divided into two parts: training part (from 1/7/2019 to 4/8/2020) and the test part (from 5/8/20 to 6/11/2020). The system was trained for the 16 Nifty 50 stocks and their performance was evaluated. Table 6.7 shows the mean absolute error (MAE) for the candidate predictions C1 and C2 and MAE-ATR obtained after the training process for the daily prediction during the testing phase. The last column of the table shows the reduction in error with mean of candidate1 and candidate2 prediction as the base. It can be seen that after applying the ANN, the MAE gets reduced significantly and error reduction up to 77% have been witnessed making the predictions quite accurate. At some places, there is an increase in the error, which is due to due the spikes in the stock prices.

Table 6.7 Reduction in error for daily price prediction after backpropagation learning

SECTOR	Stock	MAE-C1	MAE-C2	MAE-ATR	%Error Reduction
Automobile	Bajaj Auto	9.61	9.3	3.77	60.13
	Eicher Motors	27.53	27.6	11.33	58.90
	Heromotocorp	15.74	19.71	17.83	-0.59
	Mahindra	10.81	5.07	4.81	39.42
	Maruti	79.62	36.11	30.45	47.38
	Tata Motors	8.49	8.49	2.6	69.38
FMCG	Asian Paints	76.78	76.78	107.5	-40.01
	Britannia	101.22	120.39	59.19	46.58
	Hindustan Unilever	54.53	40.11	22.97	51.46
	Nestle	1366.54	1366.54	309.06	77.38
	Titan	13.01	17.95	12.79	17.38
IT	HCL	32.95	43.11	30.38	20.12
	Infosys	79.69	79.69	63.57	20.23
	TCS	104.64	116.12	92.83	15.90
	Tech Mahindra	12.47	12.47	9.35	25.02
	Wipro	8.49	6.23	6.7	8.97

Table 6.7 shows the average reduction in MAE after applying the process. To illustrate the convergence process we are augmenting the convergence graphs in Fig. 6.4 - Fig. 6.12 indicating the reduction in the prediction error on the daily basis for 9 Nifty 50 companies taking 3 companies each from IT, FMCG and Automobile sectors for the period 7/10/2020 to 6/11/2020.

Fig. 6.4 : Error Reduction in Bajaj Auto

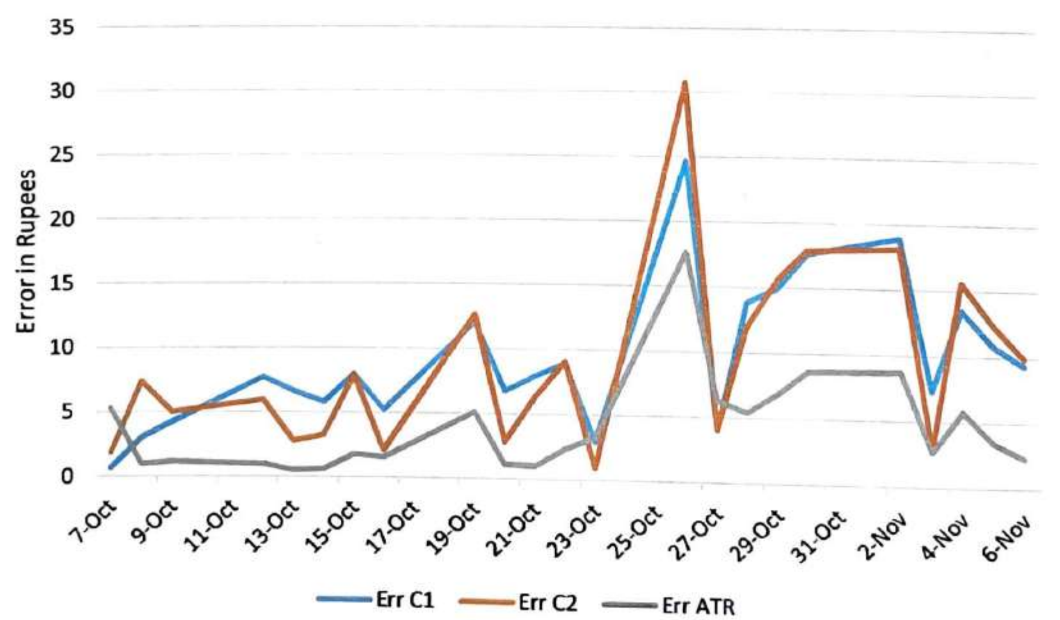


Fig. 6.5: Error Reduction in Mahindra & Mahindra

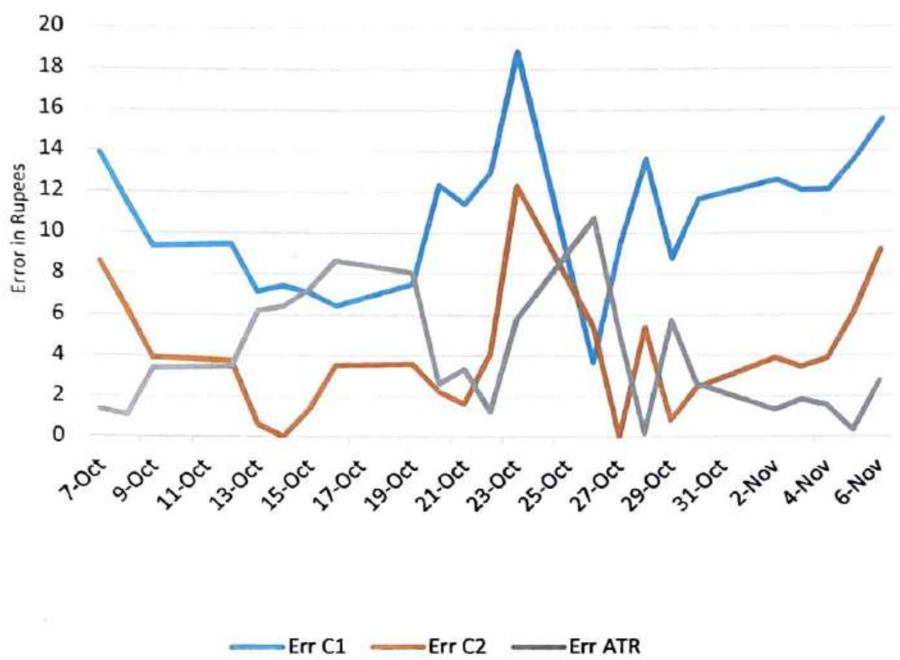


Fig. 6.6: Error Reduction in Maruti

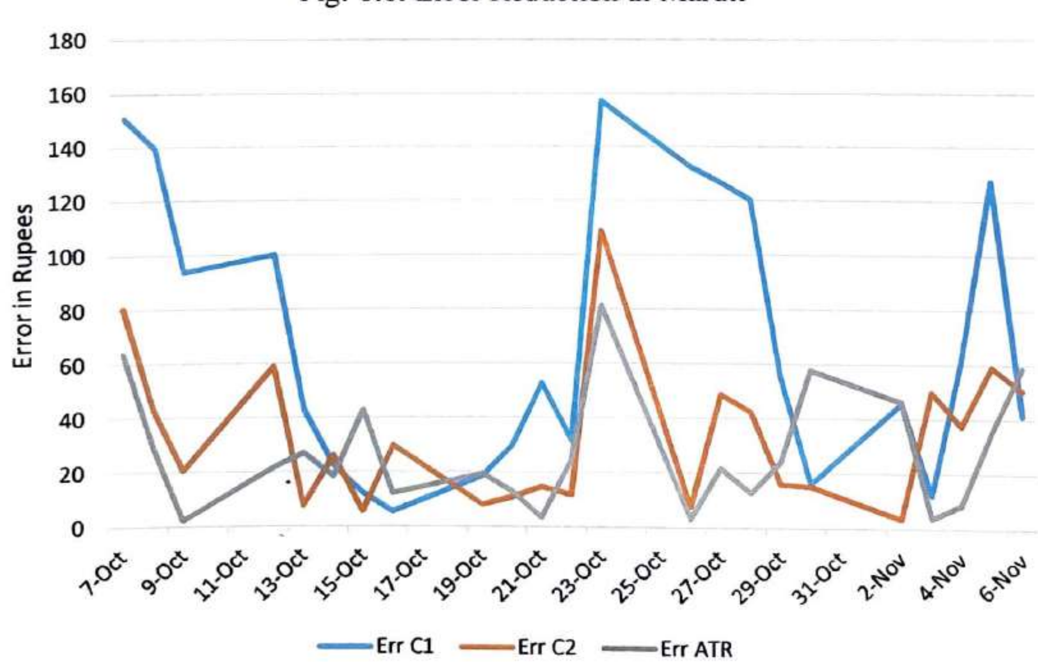


Fig.6.7 : Error Reduction in Britannia

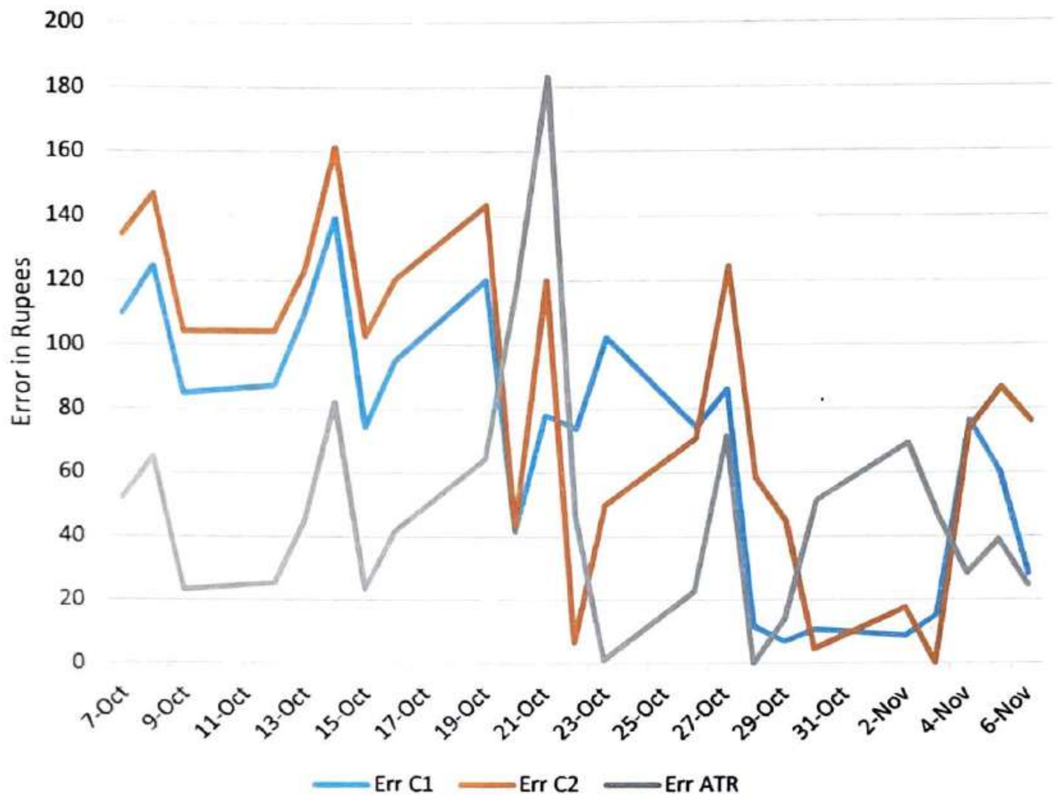


Fig. 6.8 : Error Reduction in Hindustan Unilever Limited

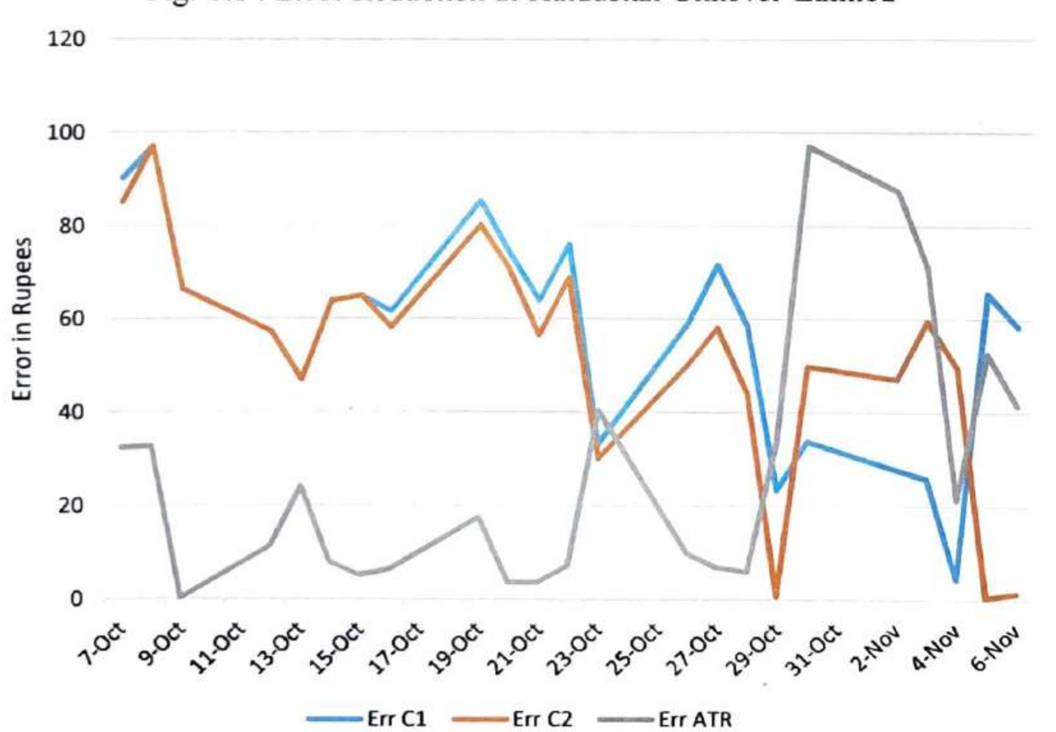


Fig. 6.9 : Error Reduction in Titan

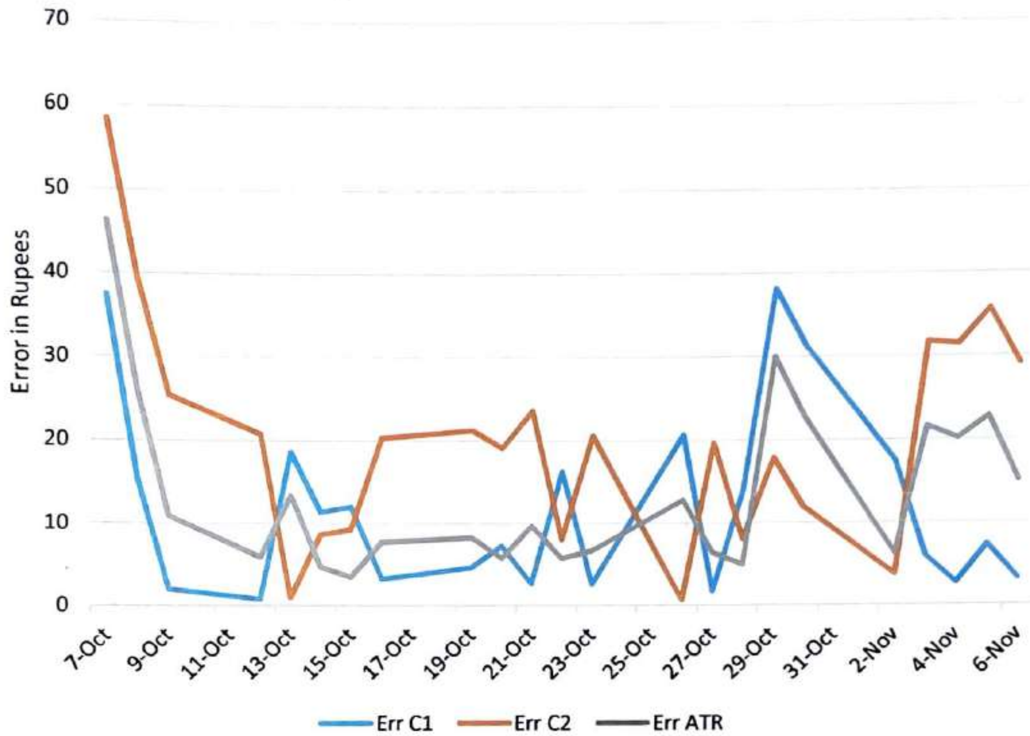


Fig. 6.10: Error Reduction in HCL Technologies

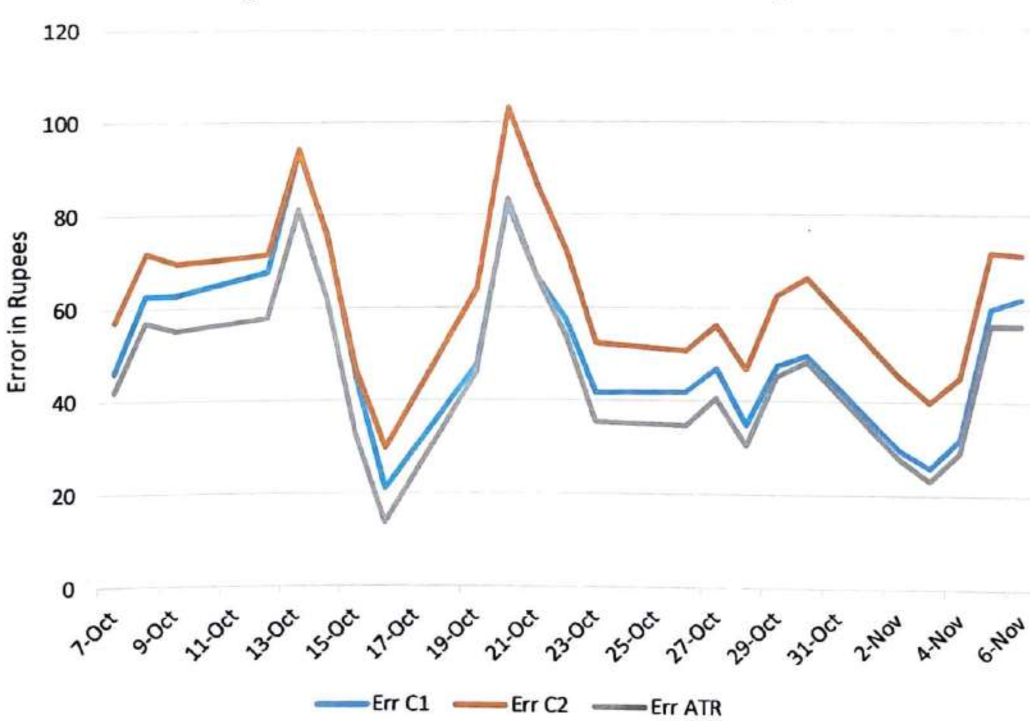


Fig.6.11 : Error Reduction in Tech Mahindra

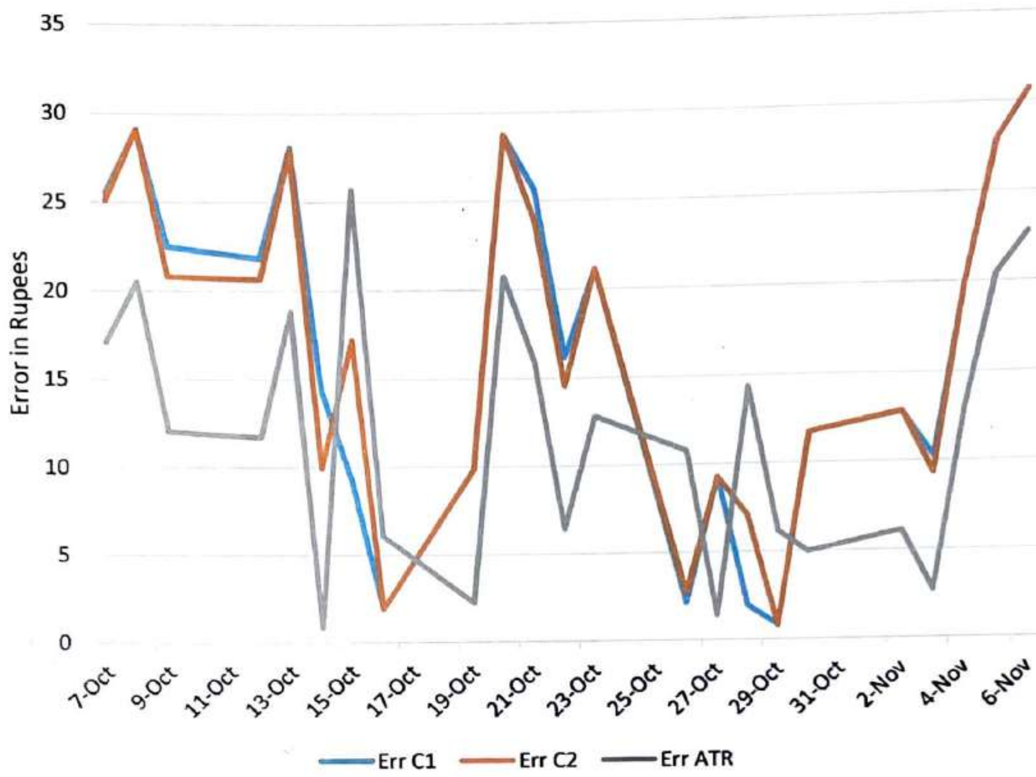
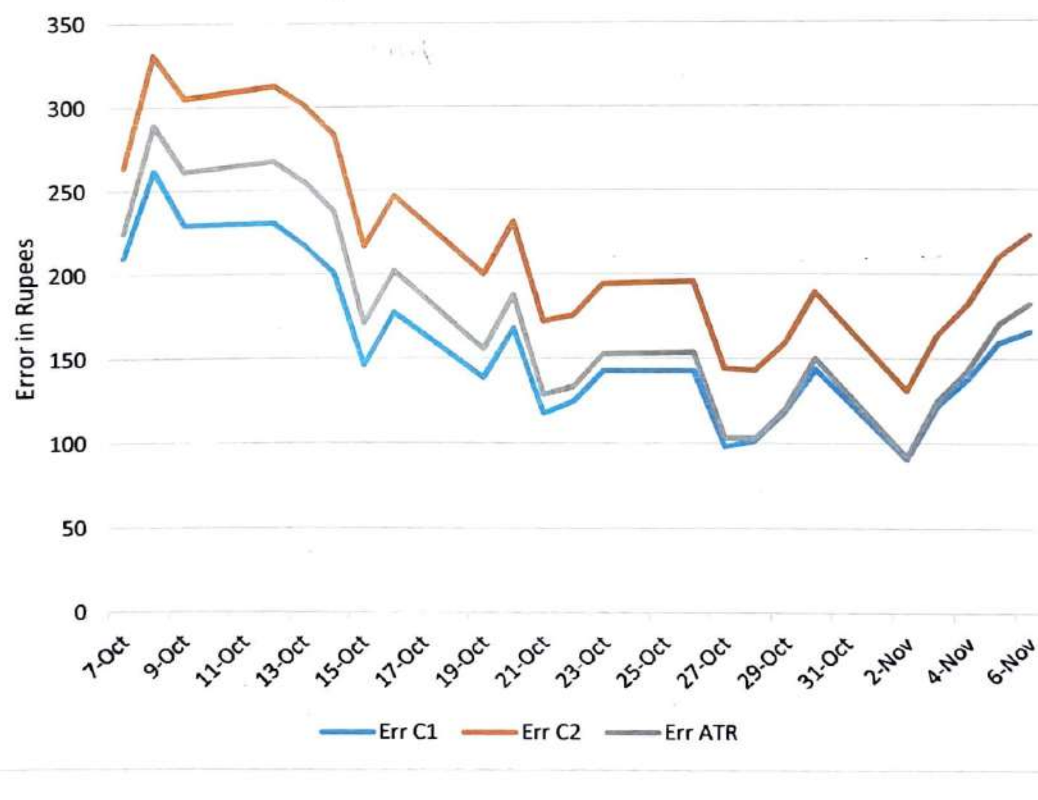


Fig.6.12 : Error Reduction in TCS



The results for the daily predictions were so enthusiastic that we were motivated to extend our experiment to the weekly, fortnightly and monthly predictions as well. The results were quite encouraging and have been shown in Table 6.8, 6.9 and 6.10 in the order. It can be seen that there has been a significant reduction in error.

Table 6.8 Extension of proposed work from Daily predictions to Weekly Predictions

Period	SECTOR	Stock	MAE-C1	MAE-C2	MAE-ATR	%Error Reduction
Weekly	Automobile	Bajaj Auto	61.64	61.46	61.71	-0.26
		Eicher Motors	63.27	63.48	62.19	1.87
		Heromotocorp	14.4	16.56	8.38	45.87
		Mahindra	7.09	9.08	2.72	66.36
		Maruti	80.81	36.65	30.92	47.35
		Tata Motors	5.55	3.48	2.55	43.52
	FMCG	Asian Paints	75.07	75.07	71.2	5.16
		Britannia	167.74	178.23	58.45	66.21
		Hindustan Unilever	54.53	74.84	26.03	59.76
		Nestle	2001.95	1940.65	244.68	87.59
		Titan	35.99	17.95	14.1	47.72
	IT	HCL	30.32	30.32	27.64	8.84
		Infosys	112.28	106.86	54.87	49.92
		TCS	48.05	48.05	51.73	-7.66
		Tech Mahindra	10.72	9.16	10.41	-4.73
		Wipro	3.69	4.68	2.63	37.16

Table 6.9 Extension of proposed work from Daily predictions to Fortnightly Predictions

Period	SECTOR	Stock	MAE-C1	MAE-C2	MAE-ATR	%Error Reduction
Fortnightly	Automobile	Bajaj Auto	9.61	9.3	3.79	59.92
		Eicher Motors	63.27	59.94	63.02	-2.30
		Heromotocorp	15.98	20	12.48	30.63
		Mahindra	16.73	14.55	15.54	0.64
		Maruti	80.81	36.65	30.98	47.25
		Tata Motors	5.99	8.12	6.88	2.48
	FMCG	Asian Paints	75.15	75.15	106.45	-41.65
		Britannia	151.45	101.79	29.47	76.73
		Hindustan Unilever	33.83	33.83	30.46	9.96
		Nestle	1366.13	1366.13	309.3	77.36
		Titan	12.99	17.89	12.42	19.56
	IT	HCL	30.32	30.32	27.72	8.58
		Infosys	112.28	106.86	54.49	50.27
		TCS	48.05	48.05	49.87	-3.79
		Tech Mahindra	10.72	9.16	10.21	-2.72
		Wipro	3.69	4.68	5.02	-19.95

Table 6.10 Extension of proposed work from Daily predictions to Monthly Predictions

Period	SECTOR	Stock	MAE-C1	MAE-C2	MAE-ATR	%Error Reduction
Monthly	Automobile	Bajaj Auto	4.19	4.01	3.87	5.61
		Eicher Motors	16.4	17.45	8.3	50.96
		Heromotocorp	28.48	29.06	25.34	11.92
		Mahindra	5.56	11.1	5.73	31.21
		Maruti	69.04	39.48	29.99	44.73
		Tata Motors	4.38	4	4.56	-8.83

FMCG	Asian Paints	64.73	64.73	92.95	-43.60
	Britannia	94.34	93.22	30.78	67.18
	Hindustan Unilever	35.16	35.16	35.07	0.26
	Nestle	1322.57	1322.57	347.78	73.70
	Titan	13.17	17.41	11.45	25.11
IT	HCL	65.57	63.16	60.95	5.31
	Infosys	109.28	109.28	56.97	47.87
	TCS	50.17	50.17	49.85	0.64
	Tech Mahindra	10.4	9.02	10.28	-5.87
	Wipro	3.75	4.92	5.76	-32.87

This work presents a stock price prediction strategy which is both simplistic and effective. The hallmark of the strategy is that it is not general but stock specific. The localization of the precdiction mechanism helps in generating the quite accurate results that may not be feasible in a generic proposal. The prediction accuracy has not only been quite high but is scalable as well, which is indicated by the weekly, fortnightly and monthly predictions. Though the work shows results for three sectors of the Nifty50, for the purpose of illustration but in actual practice, it was carried out for many other sectors as well and the results were quite encouraging. We hope that the strategy will be useful for the software developers and the end user investers. The results confirm efficacy of the proposed mechanism. Some critics may object that repeating the similar customized process for each stock is not worthwhile. In this regard, it is stated that with the availability of parallel processing techniques, huge memory spaces, and superfast processor, this is quite possible. With money being the most motivating factor, all these efforts are really worth.

CHPATER VII

CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

When it comes to the predictive modeling a domain for which the data is easily available and the results are verifiable can be an ideal candidate. Moreover, if this predictive analytics can directly support the general public then it would be an added advantage. Keeping this aspect in view, the work in this thesis has taken up the stock market domain. Upon studying the literature, it was found that almost all the research work, in the domain, was related to short term price prediction for few prominent stocks. None of the paper had taken up the stock market scenario in the holistic manner. The major drawback identified was that the predictions was related to the single day where normally there is not much change unless something exceptional happens. Another deficiency identified was that the researchers were banking upon the raw stock price data for applying the machine learning algorithms, which contains random fluctuations and destructive patterns. These factors, in our view, were the hinderance in making long term predictions. The aim of the thesis is to overcome all this limitations and drawbacks of the available work and to explore the various strategies that can be helpful in making the rational decision for making long term continuous investment in the stock market.

The contribution of the thesis can be attibuted to the following factors:

- for developing various strategies and mechanisms the stock market scenario has been taken as a whole instead of few prominent stocks. The domain under consideration has been entire NIFTY50 spectrum that contains the top 50 companies of the Indian Stock Market.
- While developing the strategies and designing the mechanisms, we have used the gist of the data instead of raw data leading to the following advantages: avoidance of the random fluctuations, destructive patterns, decreased volume, improved input data quality, better visualization etc.
- Instead of depending merely on the stock price data, other factors related to companies' have also been considered and a Stock Health Index (SHI) has been

developed that helps the investor in identifying the healthy companies to invest with.

- Since the stock markets are not merely governed by the mathematical laws, but are also manipulated by vested interests and affected by different political and economic scenarios. Therefore we have not gone for pure machine learning approach, rather machine learning has been a vital component of the complete scenarios that create reinforcement signals or generates macro features for creating the advice.

All the strategies proposed in this thesis have been properly validated by partitioning the available data into training, testing anf validation parts. The results have been quite encouraging.

7.2 FUTURE SCOPE

The startegies proposed in this work have been tested through the processes which were somewhat manual such as downloading of the data, pre-processing in the excel sheets, implementation of the machine learning scenarios, etc. All the procces have to be combined to create a fully automated scenario that can be used for the development of a commercial website for any stock market across the globe. We will look forward for the same in the future.

REFERENCES

[1] M. Biswas, T. Dhoom, and S. Barua, "Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data," *Int. J. Comput. Appl.*, vol. 182, no. 34, pp. 20–24, Dec. 2018, doi: 10.5120/ijca2018918265.

[2] A. J. Iseh and T. Y. Woma, "Weather Forecasting Models, Methods and Applications," *Int. J. Eng. Res.*, vol. 2, no. 12, p. 12, 2013.

[3] K. U. Jaseena and B. C. Kovoor, "Deterministic weather forecasting models based on intelligent predictors: A survey," *J. King Saud Univ. - Comput. Inf. Sci.*, p. S1319157820304729, Sep. 2020, doi: 10.1016/j.jksuci.2020.09.009.

[4] A. Muneer and S. M. Fati, "Automated Health Monitoring System Using Advanced Technology:," *J. Inf. Technol. Res.*, vol. 12, no. 3, pp. 104–132, Jul. 2019, doi: 10.4018/JITR.2019070107.

[5] H. Mshali, T. Lemlouma, M. Moloney, and D. Magoni, "A survey on health monitoring systems for health smart homes," *Int. J. Ind. Ergon.*, vol. 66, pp. 26–56, Jul. 2018, doi: 10.1016/j.ergon.2018.02.002.

[6] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques – Part II: Soft computing methods," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5932–5941, Apr. 2009, doi: 10.1016/j.eswa.2008.07.006.

[7] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega, and A. L. I. Oliveira, "Computational Intelligence and Financial Markets: A Survey and Future Directions," *Expert Syst. Appl.*, vol. 55, pp. 194–211, Aug. 2016, doi: 10.1016/j.eswa.2016.02.006.

[8] N. A. C. Azhar and M. A. Mansor, "Development of plant maintenance management system (pmms): a case study," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 50, p. 012058, Dec. 2013, doi: 10.1088/1757-899X/50/1/012058.

[9] A. Das, N. Gandhewar, D. Singh Nehra, M. Baraskar, S. Gurjar, and M. Khan, "Survey on Vehicle Tracking Services," *J. Inf. Technol. Softw. Eng.*, vol. 08, no. 01, 2017, doi: 10.4172/2165-7866.1000224.

[10] Department of CSE VLITS Guntur, India and S. K. R. Mallidi, "IOT BASED SMART VEHICLE MONITORING SYSTEM," *Int. J. Adv. Res. Comput. Sci.*, vol. 9, no. 2, pp. 738–741, Feb. 2018, doi: 10.26483/ijarcs.v9i2.5870.

[11] A. Nugroho and B. H. Simanjuntak, "ARMA (Autoregressive Moving Average) Model for Prediction of Rainfall in Regency of Semarang - Central Java - Republic of Indonesia," vol. 11, no. 3, p. 6, 2014.

[12] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction," *J. Appl. Math.*, vol. 2014, pp. 1–7, 2014, doi: 10.1155/2014/614342.

[13] A. A. Adebiyi, C. K. Ayo, A. O. Adewumi, and S. O. Otokiti, "Stock Price Prediction using Neural Network with Hybridized Market Indicators," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 1, no. 1, pp. 1–9, Jan. 2012.

[14] S. Degiannakis and E. Xekalaki, "Autoregressive Conditional Heteroscedasticity (ARCH) Models: A Review," *Qual. Technol. Quant. Manag.*, vol. 1, no. 2, pp. 271–324, Jan. 2004, doi: 10.1080/16843703.2004.11673078.

[15] S. Mohamadi, H. Amindavar, and S. M. A. Tayaranian Hosseini, "ARIMA-GARCH modeling for epileptic seizure prediction," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, Mar. 2017, pp. 994–998. doi: 10.1109/ICASSP.2017.7952305.

[16] K. Singh *et al.*, "Implementation of Exponential Smoothing for Forecasting Time Series Data," vol. 8, no. 1, p. 5.

[17] "Rediff Money," May 25, 2021. https://www.money.rediff.com

[18] "Yahoo Finance," May 25, 2021. https://in.finance.yahoo.com

[19] "Money Control," May 25, 2021. https://www.moneycontrol.com

[20] "National Stock Exchange," May 25, 2021. https://www.nseindia.com

[21] J. Sirignano and R. Cont, "Universal Features of Price Formation in Financial Markets: Perspectives From Deep Learning," *SSRN Electron. J.*, 2018, doi: 10.2139/ssrn.3141294.

[22] L.-P. Ni, Z.-W. Ni, and Y.-Z. Gao, "Stock trend prediction based on fractal feature selection and support vector machine," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5569–5576, May 2011, doi: 10.1016/j.eswa.2010.10.079.

[23] S. McNally, J. Roche, and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning," in *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, Cambridge, Mar. 2018, pp. 339–343. doi: 10.1109/PDP2018.2018.00060.

[24] Y. Kara, M. Acar Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5311–5319, May 2011, doi: 10.1016/j.eswa.2010.10.027.

[25] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018, doi: 10.1016/j.ejor.2017.11.054.

[26] W. Long, Z. Lu, and L. Cui, "Deep learning-based feature engineering for stock price movement prediction," *Knowl.-Based Syst.*, vol. 164, pp. 163–173, Jan. 2019, doi: 10.1016/j.knosys.2018.10.034.

[27] Jingyi Shen and M. O. Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system," *J. Big Data*, vol. 7, no. 1, p. 66, Dec. 2020, doi: 10.1186/s40537-020-00333-6.

[28] A. Kong, H. Zhu, and R. Azencott, "Predicting intraday jumps in stock prices using liquidity measures and technical indicators," *ArXiv191207165 Q-Fin*, Dec. 2019, Accessed: Oct. 19, 2021. [Online]. Available: http://arxiv.org/abs/1912.07165

[29] L. Lei, "Wavelet Neural Network Prediction Method of Stock Price Trend Based on Rough Set Attribute Reduction," *Appl. Soft Comput.*, vol. 62, pp. 923–932, Jan. 2018, doi: 10.1016/j.asoc.2017.09.029.

[30] A. Pimenta, C. A. L. Nametala, F. G. Guimarães, and E. G. Carrano, "An Automated Investing Method for Stock Market Based on Multiobjective Genetic Programming," *Comput. Econ.*, vol. 52, no. 1, pp. 125–144, Jun. 2018, doi: 10.1007/s10614-017-9665-9.

[31] S. M. Idrees, M. A. Alam, and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," *IEEE Access*, vol. 7, pp. 17287–17298, 2019, doi: 10.1109/ACCESS.2019.2895252.

[32] B. Weng, L. Lu, X. Wang, F. M. Megahed, and W. Martinez, "Predicting short-term stock prices using ensemble methods and online data sources," *Expert Syst. Appl.*, vol. 112, pp. 258–273, Dec. 2018, doi: 10.1016/j.eswa.2018.06.016.

[33] H. NekoeiQachkanloo, B. Ghojogh, A. S. Pasand, and M. Crowley, "Artificial Counselor System for Stock Investment," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 9558–9564, Jul. 2019, doi: 10.1609/aaai.v33i01.33019558.

[34] S. Jeon, B. Hong, and V. Chang, "Pattern graph tracking-based stock price prediction using big data," *Future Gener. Comput. Syst.*, vol. 80, pp. 171–187, Mar. 2018, doi: 10.1016/j.future.2017.02.010.

[35] Amit K. Sinha, "The reliability of geometric Brownian motion forecasts of S&P500 index values," *J. Forecast.*, vol. 40, no. 8, pp. 1444–1462, Dec. 2021, doi: 10.1002/for.2775.

[36] M. He, X. Hao, Y. Zhang, and F. Meng, "Forecasting stock return volatility using a robust regression model," *J. Forecast.*, p. for.2779, Apr. 2021, doi: 10.1002/for.2779.

[37] T. M. Cover, "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Trans. Electron. Comput.*, vol. EC-14, no. 3, pp. 326–334, Jun. 1965, doi: 10.1109/PGEC.1965.264137.

[38] A. Nayak, M. M. M. Pai, and R. M. Pai, "Prediction Models for Indian Stock Market," *Procedia Comput. Sci.*, vol. 89, pp. 441–449, 2016, doi: 10.1016/j.procs.2016.06.096.

[39] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," *PLOS ONE*, vol. 14, no. 2, p. e0212320, Feb. 2019, doi: 10.1371/journal.pone.0212320.

[40] C. Stoean, W. Paja, R. Stoean, and A. Sandita, "Deep architectures for long-term stock price prediction with a heuristic-based strategy for trading simulations,"

PLOS ONE, vol. 14, no. 10, p. e0223593, Oct. 2019, doi: 10.1371/journal.pone.0223593.

[41] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLOS ONE*, vol. 12, no. 7, p. e0180944, Jul. 2017, doi: 10.1371/journal.pone.0180944.

[42] W.-C. Chiang, D. Enke, T. Wu, and R. Wang, "An adaptive stock index trading decision support system," *Expert Syst. Appl.*, vol. 59, pp. 195–207, Oct. 2016, doi: 10.1016/j.eswa.2016.04.025.

[43] M. Göçken, M. Özçalıcı, A. Boru, and A. T. Dosdoğru, "Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction," *Expert Syst. Appl.*, vol. 44, pp. 320–331, Feb. 2016, doi: 10.1016/j.eswa.2015.09.029.

[44] M. Göçken, M. Özçalıcı, A. Boru, and A. T. Dosdoğru, "Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection," *Neural Comput. Appl.*, vol. 31, no. 2, pp. 577–592, Feb. 2019, doi: 10.1007/s00521-017-3089-2.

[45] M. Qiu and Y. Song, "Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model," *PLOS ONE*, vol. 11, no. 5, p. e0155133, May 2016, doi: 10.1371/journal.pone.0155133.

[46] Y. Shynkevich, T. M. McGinnity, S. A. Coleman, A. Belatreche, and Y. Li, "Forecasting price movements using technical indicators: Investigating the impact of varying input window length," *Neurocomputing*, vol. 264, pp. 71–88, Nov. 2017, doi: 10.1016/j.neucom.2016.11.095.

[47] J. Zahedi and M. M. Rounaghi, "Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange," *Phys. Stat. Mech. Its Appl.*, vol. 438, pp. 178–187, Nov. 2015, doi: 10.1016/j.physa.2015.06.033.

[48] X. Zhong and D. Enke, "Forecasting daily stock market return using dimensionality reduction," *Expert Syst. Appl.*, vol. 67, pp. 126–139, Jan. 2017, doi: 10.1016/j.eswa.2016.09.027.

[49] R. Dash and P. K. Dash, "A hybrid stock trading framework integrating technical analysis with machine learning techniques," *J. Finance Data Sci.*, vol. 2, no. 1, pp. 42–57, Mar. 2016, doi: 10.1016/j.jfds.2016.03.002.

[50] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 259–268, Jan. 2015, doi: 10.1016/j.eswa.2014.07.040.

[51] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Syst. Appl.*, vol. 42, no. 4, pp. 2162–2172, Mar. 2015, doi: 10.1016/j.eswa.2014.10.031.

[52] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," *J. Ambient Intell. Humaniz. Comput.*, Mar. 2020, doi: 10.1007/s12652-020-01839-w.

[53] H. S. Sim, H. I. Kim, and J. J. Ahn, "Is Deep Learning for Image Recognition Applicable to Stock Market Prediction?," *Complexity*, vol. 2019, pp. 1–10, Feb. 2019, doi: 10.1155/2019/4324878.

[54] X. Jiawei and T. Murata, "Stock Market Trend Prediction with Sentiment Analysis based on LSTM Neural Network," *Hong Kong*, p. 5, 2019.

[55] X. Zhong and D. Enke, "Predicting the daily return direction of the stock market using hybrid machine learning algorithms," *Financ. Innov.*, vol. 5, no. 1, p. 24, Dec. 2019, doi: 10.1186/s40854-019-0138-0.

[56] D. Lien Minh, A. Sadeghi-Niaraki, H. D. Huy, K. Min, and H. Moon, "Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network," *IEEE Access*, vol. 6, pp. 55392–55404, 2018, doi: 10.1109/ACCESS.2018.2868970.

[57] c T. Ngo and T. T. H. Truong, "Forecasting time series data using moving-window swarm intelligence-optimised machine learning regression," *Int. J. Intell. Eng. Inform.*, vol. 7, no. 5, p. 422, 2019, doi: 10.1504/IJIEI.2019.103625.

[58] I. Arora and A. Saha, "Software fault prediction using firefly algorithm," *Int. J. Intell. Eng. Inform.*, vol. 6, no. 3/4, p. 356, 2018, doi: 10.1504/IJIEI.2018.091870.

[59] R. Ouhibi, S. Bouslama, and K. Laabidi, "Fault detection and isolation of asynchronous machine based on the probabilistic neural network," *Int. J. Intell. Eng. Inform.*, vol. 6, no. 3/4, p. 378, 2018, doi: 10.1504/IJIEI.2018.091879.

[60] S. S. Kumar and H. H. Inbarani, "Covering rough set-based classification for cardiac arrhythmia," *Int. J. Intell. Eng. Inform.*, vol. 5, no. 2, p. 101, 2017, doi: 10.1504/IJIEI.2017.084167.

[61] S. Hanafi, R. Benmansour, and S. Khalouli, "Ant colony optimisation combined with variable neighbourhood search for scheduling preventive railway maintenance activities," *Int. J. Intell. Eng. Inform.*, vol. 6, no. 1/2, p. 78, 2018, doi: 10.1504/IJIEI.2018.10012067.

[62] Z. Zhao and N. Yu, "An improved quantum particle swarm optimisation and its application on hand kinematics tracking," *Int. J. Intell. Eng. Inform.*, vol. 6, no. 3/4, p. 266, 2018, doi: 10.1504/IJIEI.2018.091872.

[63] R. A. Meese and K. Rogoff, "Empirical exchange rate models of the seventies," *J. Int. Econ.*, vol. 14, no. 1–2, pp. 3–24, Feb. 1983, doi: 10.1016/0022-1996(83)90017-X.

[64] G. R. Duffee, "Term Premia and Interest Rate Forecasts in Affine Models," *J. Finance*, vol. 57, no. 1, pp. 405–443, Feb. 2002, doi: 10.1111/1540-6261.00426.

[65] G. R. Duffee, "Handbook of Economic Forecasting," in *Handbook of Economic Forecasting*, vol. 2, Elsevier, 2013, p. iii. doi: 10.1016/B978-0-444-53683-9.00023-2.

[66] T. S. Eicher, D. J. Kuenzel, C. Papageorgiou, and C. Christofides, "Forecasts in times of crises," *Int. J. Forecast.*, vol. 35, no. 3, pp. 1143–1159, Jul. 2019, doi: 10.1016/j.ijforecast.2019.04.001.

[67] I. Welch and A. Goyal, "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Rev. Financ. Stud.*, vol. 21, no. 4, pp. 1455–1508, Jul. 2008, doi: 10.1093/rfs/hhm014.

[68] M. Nikou, G. Mansourfar, and J. Bagherzadeh, "Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms," *Intell. Syst. Account. Finance Manag.*, vol. 26, no. 4, pp. 164–174, Oct. 2019, doi: 10.1002/isaf.1459.

[69] Y. Baek and H. Y. Kim, "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module," *Expert Syst. Appl.*, vol. 113, pp. 457–480, Dec. 2018, doi: 10.1016/j.eswa.2018.07.019.

[70] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Syst. Appl.*, vol. 103, pp. 25–37, Aug. 2018, doi: 10.1016/j.eswa.2018.03.002.

[71] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," *J. Supercomput.*, vol. 76, no. 3, pp. 2098–2118, Mar. 2020, doi: 10.1007/s11227-017-2228-y.

[72] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, and T.-S. Chua, "Temporal Relational Ranking for Stock Prediction," *ACM Trans. Inf. Syst.*, vol. 37, no. 2, pp. 1–30, Mar. 2019, doi: 10.1145/3309547.

[73] E. Guresen, G. Kayakutlu, and T. U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10389–10397, Aug. 2011, doi: 10.1016/j.eswa.2011.02.068.

[74] M. Seo, S. Lee, and G. Kim, "Forecasting the Volatility of Stock Market Index Using the Hybrid Models with Google Domestic Trends," *Fluct. Noise Lett.*, vol. 18, no. 01, p. 1950006, Mar. 2019, doi: 10.1142/S0219477519500068.

[75] J. Wang, J. Wang, W. Fang, and H. Niu, "Financial Time Series Prediction Using Elman Recurrent Random Neural Networks," *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–14, 2016, doi: 10.1155/2016/4742515.

[76] F. Yang, Z. Chen, J. Li, and L. Tang, "A novel hybrid stock selection method with stock prediction," *Appl. Soft Comput.*, vol. 80, pp. 820–831, Jul. 2019, doi: 10.1016/j.asoc.2019.03.028.

[77] D. Shah, H. Isah, and F. Zulkernine, "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques," *Int. J. Financ. Stud.*, vol. 7, no. 2, p. 26, May 2019, doi: 10.3390/ijfs7020026.

[78] D. A. Kumar and S. Murugan, "Performance analysis of Indian stock market index using neural network time series model," in *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, Salem, Feb. 2013, pp. 72–78. doi: 10.1109/ICPRIME.2013.6496450.

[79] A. V. Metcalfe and P. S. P. Cowpertwait, *Introductory Time Series with R*. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-88698-5.

[80] Y.-W. Si and J. Yin, "OBST-based segmentation approach to financial time series," *Eng. Appl. Artif. Intell.*, vol. 26, no. 10, pp. 2581–2596, Nov. 2013, doi: 10.1016/j.engappai.2013.08.015.

[81] A. M. Rather, A. Agarwal, and V. N. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3234–3241, Apr. 2015, doi: 10.1016/j.eswa.2014.12.003.

[82] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," in *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, Cambridge, United Kingdom, Mar. 2014, pp. 106–112. doi: 10.1109/UKSim.2014.67.

[83] E. Ahmadi, M. Jasemi, L. Monplaisir, M. A. Nabavi, A. Mahmoodi, and P. Amini Jam, "New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the Support Vector Machine and Heuristic Algorithms of Imperialist Competition and Genetic," *Expert Syst. Appl.*, vol. 94, pp. 21–31, Mar. 2018, doi: 10.1016/j.eswa.2017.10.023.

[84] G. S. Atsalakis and K. P. Valavanis, "SURVEYING STOCK MARKET FORECASTING TECHNIQUES - PART I: CONVENTIONAL METHODS," p. 57.

[85] L.-C. Cheng, Y.-H. Huang, and M.-E. Wu, "Applied attention-based LSTM neural networks in stock prediction," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, Dec. 2018, pp. 4716–4718. doi: 10.1109/BigData.2018.8622541.

[86] S. Chopra, D. Yadav, and A. N. Chopra, "Artificial Neural Networks Based Indian Stock Market Price Prediction: Before and After Demonetization," no. 174, p. 7.

[87] M. R. Senapati, S. Das, and S. Mishra, "A Novel Model for Stock Price Prediction Using Hybrid Neural Network," *J. Inst. Eng. India Ser. B*, vol. 99, no. 6, pp. 555–563, Dec. 2018, doi: 10.1007/s40031-018-0343-7.

[88] J. Wang and J. Wang, "Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks," *Neurocomputing*, vol. 156, pp. 68–78, May 2015, doi: 10.1016/j.neucom.2014.12.084.

[89] J. Wang and J. Wang, "Forecasting stochastic neural network based on financial empirical mode decomposition," *Neural Netw.*, vol. 90, pp. 8–20, Jun. 2017, doi: 10.1016/j.neunet.2017.03.004.

[90] O. Cagcag Yolcu and H.-K. Lam, "A combined robust fuzzy time series method for prediction of time series," *Neurocomputing*, vol. 247, pp. 87–101, Jul. 2017, doi: 10.1016/j.neucom.2017.03.037.

[91] "NIFTY Indices," May 25, 2021. https://www.niftyindices.com/reports/historical-data

[92] H. Kim and W. Shi, "Forecasting financial vulnerability in the USA: A factor model approach," *J. Forecast.*, vol. 40, no. 3, pp. 439–457, Apr. 2021, doi: 10.1002/for.2724.

[93] T. A. Bohn, "Improving Long Term Stock Market Prediction with Text Analysis," p. 91.

[94] X. Li, C. Wang, J. Dong, F. Wang, X. Deng, and S. Zhu, "Improving Stock Market Prediction by Integrating Both Market News and Stock Prices," in *Database and Expert Systems Applications*, vol. 6861, A. Hameurlain, S. W. Liddle, K.-D. Schewe, and X. Zhou, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 279–293. doi: 10.1007/978-3-642-23091-2_24.

[95] "BSE India." [Online]. Available: https://www.bseindia.com/

[96] S. S. Izquierdo and L. R. Izquierdo, "Mamdani fuzzy systems for modelling and simulation: A critical assessment," p. 18.

[97] M. Risse, "Combining wavelet decomposition with machine learning to forecast gold returns," *Int. J. Forecast.*, vol. 35, no. 2, pp. 601–615, Apr. 2019, doi: 10.1016/j.ijforecast.2018.11.008.

[98] J. quinn, "An overview of the four main approaches to predictive analytics," Feb. 27, 2020. An overview of the four main approaches to predictive analytics

[99] H. Chung and K. Shin, "Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction," *Sustainability*, vol. 10, no. 10, p. 3765, Oct. 2018, doi: 10.3390/su10103765.

[100] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9713–9729, Jul. 2020, doi: 10.1007/s00521-019-04504-2.

[101] J. Long, Z. Chen, W. He, T. Wu, and J. Ren, "An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market," *Appl. Soft Comput.*, vol. 91, p. 106205, Jun. 2020, doi: 10.1016/j.asoc.2020.106205.

[102] M.-Y. Chen, C.-H. Liao, and R.-P. Hsieh, "Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach," *Comput. Hum. Behav.*, vol. 101, pp. 402–408, Dec. 2019, doi: 10.1016/j.chb.2019.03.021.

[103] H. Li, Y. Shen, and Y. Zhu, "Stock Price Prediction Using Attention-based Multi-Input LSTM," p. 16.

[104] D. Shah, W. Campbell, and F. H. Zulkernine, "A Comparative Study of LSTM and DNN for Stock Market Forecasting," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, Dec. 2018, pp. 4148–4155. doi: 10.1109/BigData.2018.8622462.

[105] B.-S. Lin, W.-T. Chu, and C.-M. Wang, "Application of Stock Analysis Using Deep Learning," in *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)*, Yonago, Japan, Jul. 2018, pp. 612–617. doi: 10.1109/IIAI-AAI.2018.00128.

[106] T. Skehin, M. Crane, and M. Bezbradica, "Day Ahead Forecasting of FAANG Stocks Using ARIMA, LSTM Networks and Wavelets.," p. 12.

[107] H. Niu, K. Xu, and W. Wang, "A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network," *Appl. Intell.*, vol. 50, no. 12, pp. 4296–4309, Dec. 2020, doi: 10.1007/s10489-020-01814-0.

[108] Q. Wang, W. Xu, X. Huang, and K. Yang, "Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning," *Neurocomputing*, vol. 347, pp. 46–58, Jun. 2019, doi: 10.1016/j.neucom.2019.03.006.

[109] W. Xu, H. Peng, X. Zeng, F. Zhou, X. Tian, and X. Peng, "A hybrid modelling method for time series forecasting based on a linear regression model and deep learning," *Appl. Intell.*, vol. 49, no. 8, pp. 3002–3015, Aug. 2019, doi: 10.1007/s10489-019-01426-3.

[110] L. D. Persio and H. Oleksandr, "Recurrent neural networks approach to the financial forecast of Google assets," *Int. J. Math. Comput. Simul.*, vol. 11, 2017.

[111] W. Leigh, C. J. Frohlich, S. Hornik, R. L. Purvis, and T. L. Roberts, "Trading With a Stock Chart Heuristic," *IEEE Trans. Syst. Man Cybern. - Part Syst. Hum.*, vol. 38, no. 1, pp. 93–104, Jan. 2008, doi: 10.1109/TSMCA.2007.909508.

[112] J. L. Ticknor, "A Bayesian regularized artificial neural network for stock market forecasting," *Expert Syst. Appl.*, vol. 40, no. 14, pp. 5501–5506, Oct. 2013, doi: 10.1016/j.eswa.2013.04.013.

[113] C.-Y. Yeh, C.-W. Huang, and S.-J. Lee, "A multiple-kernel support vector regression approach for stock market price forecasting," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2177–2186, Mar. 2011, doi: 10.1016/j.eswa.2010.08.004.

[114] U. Fayyad, D. Haussler, and P. E. Stolorz, "KDD for Science Data Analysis: Issues and Examples," p. 7.

[115] A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, "Stock market index prediction using artificial neural network," *J. Econ. Finance Adm. Sci.*, vol. 21, no. 41, pp. 89–93, Dec. 2016, doi: 10.1016/j.jefas.2016.07.002.

[116] D. Venugopal, S. Kannan Kaliyaperumal, and S. Muthu Niraikulathan, "Stock Market Trend Prediction Using Hidden Markov Model," in *Forecasting in Mathematics*

- *Recent Advances, New Perspectives and Applications*, A. Abou Jaoude, Ed. IntechOpen, 2021. doi: 10.5772/intechopen.93988.

[117] J. H. Wright, "Some observations on forecasting and policy," *Int. J. Forecast.*, vol. 35, no. 3, pp. 1186–1192, Jul. 2019, doi: 10.1016/j.ijforecast.2019.04.003.

[118] A. Carriero, A. B. Galvão, and G. Kapetanios, "A comprehensive evaluation of macroeconomic forecasting methods," *Int. J. Forecast.*, vol. 35, no. 4, pp. 1226–1239, Oct. 2019, doi: 10.1016/j.ijforecast.2019.02.007.

RESEARCH SCHOLAR'S PROFILE

Shruti Mittal is an Assistant Professor in the Department of Computer Engineering at J C Bose University of Science & Technology, YMCA, Faridabad, since 2007. She received her M.Tech. (Computer Engineering) in 2006. Her research work spans both data analytics and machine learning.

LIST OF PUBLICATIONS

List of Published Papers

Sr. No.	Title of the Paper	Name of The Journal	ISSN	Volume & Issue	Year	Pages
1	Data Fusion Reference Models: Basic Concept, Architecture and Comparison	International Journal of Emerging Technologies and Innovative Research (UGC Approved)	2349-5162	Vol.6, Issue 6	June-2019	241-245
2	Reinforcement Learning Based Predictive Analytics Framework for Survival in Stock Market	International Journal of Intelligent Engineering Informatics (indexed in ESCI)	1758-8723	Vol. 9, Issue 3	October, 2021	294-327
3	Predicting a reliable stock for mid and long term investment	Journal of King Saud University - Computer and Information Sciences (indexed in SCI E with IF 13.47)	1319-1578	https://doi.org/10.1016/j.jksuci.2021.08.022	August , 2021	9 pages
4	Stock Market Prediction by Incorporating News Sentiments Using Bert	ICCCMLA 2021 (Scopus Indexed)	Will be published as a book chapter			
5	Impact of Technical indicators in Stock Price Prediction	ICACCT, 2021 (Scopus Indexed)	Will be published as a book chapter			
6	A Dynamically Adapting Framework for Stock Price Prediction	ICSOFTCOMP, 2021 (Scopus Indexed)	Will be published as a book chapter			
7	URBAN COMPUTING: Key Challenges and Issues of Traffic Management System	RSTTMI, 2016 (National Conference)				

List of Communicated Papers

Sr.No.	Title of the Paper	Name of the Journal
1	A Predictive Analytics Framework for Opportunity Sensing in Stock Market	Kuwait Journal of Science (SCI E indexed)