

Automated Stock Market Trading System

Submitted By

Parth Shah

13MCEN34



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2015

Automated Stock Market Trading System

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering
(Networking Technologies)

Submitted By

Parth Shah

(13MCEN34)

Guided By

Prof. Vishal Parikh



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2015

Certificate

This is to certify that the major project entitled “**Automated Stock Market Trading System**” submitted by **Parth Shah (Roll No: 13MCEN34)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering(Networking Technologies) of Institute of Technology, Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Vishal Parikh
Guide & Assistant Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Prof. Gaurang Raval
Associate Professor,
Coordinator M.Tech - CSE(NT)
Institute of Technology,
Nirma University, Ahmedabad

Dr. Sanjay Garg
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. K Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Parth Shah**, Roll. No. **13MCEN34**, give undertaking that the Major Project entitled “**Automated Stock Market Trading System**” submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student

Date:

Place:

Endorsed by
Prof. Vishal Parikh
(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Vishal Parikh**, Assistant Professor, Computer Science Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. K Kotecha**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

See that you acknowledge each one who have helped you in the project directly or indirectly.

- **Parth Shah**
13MCEN34

Abstract

Stock market decision making is a very challenging and difficult task of financial data prediction. Prediction about stock market with high accuracy movement yield profit for investors of the stocks. Because of the complexity of stock market financial data, development of efficient models for prediction decision is very difficult, and it must be accurate. This study attempted to develop models for prediction of the stock market and to decide whether to buy/hold the stock using data mining and machine learning techniques. The machine learning technique like Naive Bayes, k-Nearest Neighbor(k-NN), Support Vector Machine(SVM), Artificial Neural Network(ANN)and Random Forest has been used for developing of prediction model. Technical indicators are calculated from the stock prices based on time-line data and it is used as inputs of the proposed prediction models. Ten years of stock market data has been used for signal prediction of stock. Based on the data set, these models are capable to generate buy/hold signal for stock market as a output. The main goal of this project is to generate output as per users requirement like amount to be invested, time duration for investment, minimum profit, maximum loss using data mining and machine learning techniques.

Contents

Certificate	iii
Statement of Originality	iv
Acknowledgements	v
Abstract	vi
List of Figures	ix
List of Tables	1
1 Introduction	1
1.1 Objective of Project	1
1.2 Scope	2
1.3 Output	2
2 Literature Survey	3
2.1 Fundamental analysis	3
2.2 Technical Analysis	4
2.2.1 Strengths of Technical Analysis	5
2.2.2 Technical Parameter	5
2.3 Data Processing	10
2.3.1 Decision Parameter Generation	10
2.3.2 Feature selection	10
2.3.3 Outlier Detection	11
2.3.4 Discretization	11
2.3.5 Normalization	11
2.3.6 Sampling	11
2.4 Related Work	12
3 Prediction Model	15
3.1 Naive Bayesian Classification	15
3.2 k-Nearest-Neighbor Classifiers	17
3.3 Artificial Neural Networks(ANN)	17
3.4 Support Vector Machine(SVM)	20
3.5 Random Forest Classification	22
3.6 Ensemble Learning	23

4	EXPERIMENTAL RESULTS	24
4.1	Experiment 1:	24
4.1.1	Naive Bayes	25
4.1.2	k-NN	26
4.1.3	ANN	27
4.1.4	SVM Polynomial Kernel	28
4.1.5	SVM Radial Kernel	29
4.1.6	Random Forest	30
4.2	Experiment 2	31
4.2.1	Naive Bayes	31
4.2.2	k-NN	32
4.2.3	ANN	33
4.2.4	SVM Polynomial Kernel	34
4.2.5	SVM Radial Kernel	35
4.2.6	Random Forest	36
5	Conclusion and Future Scope	37
5.1	Conclusion	37
5.2	Future Scope	37
	References	38

List of Figures

2.1	Stock market fundamental analysis[1]	4
3.1	Artificial Neural Networks(ANN) [2]	18
3.2	Structure of ANN for stock market decision generation[3]	19

List of Tables

2.1	Technical Parameter	9
2.2	Techniques used for stock market prediction	14
3.1	ANN Design Parameter	20
3.2	Best ANN Design Parameter Based on Accuracy	20
3.3	SVM Design Parameter	21
3.4	Best SVM Design Parameter based on Accuracy	21
3.5	Random Forest Design Parameter	22
3.6	Best Random Forest Design Parameter based on Accuracy	22
4.1	Naive bayes results of top gainer stock for 10% profit in 30 day	25
4.2	Naive bayes results of top loser stock for 10% profit in 30 day	25
4.3	k-NN results of top gainer stock for 10% profit in 30 day	26
4.4	k-NN results of top loser stock for 10% profit in 30 day	26
4.5	ANN results of top gainer stock for 10% profit in 30 day	27
4.6	ANN results of top loser stock for 10% profit in 30 day	27
4.7	SVM Polynomial Kernel results of top gainer stock for 10% profit in 30 day	28
4.8	SVM Polynomial Kernel results of top loser stock for 10% profit in 30 day	28
4.9	SVM Radial Kernel results of top gainer stock for 10% profit in 30 day	29
4.10	SVM Radial Kernel results of top loser stock for 10% profit in 30 day	29
4.11	Random Forest results of top gainer stock for 10% profit in 30 day	30
4.12	Random Forest results of top loser stock for 10% profit in 30 day	30
4.13	Naive bayes results of top gainer stock for 15% profit in 60 day	31
4.14	Naive bayes results of top loser stock for 15% profit in 60 day	31
4.15	k-NN results of top gainer stock for 15% profit in 60 day	32
4.16	k-NN results of top loser stock for 15% profit in 60 day	32
4.17	ANN results of top gainer stock for 15% profit in 60 day	33
4.18	ANN results of top loser stock for 15% profit in 60 day	33
4.19	SVM Polynomial Kernel results of top gainer stock for 15% profit in 60 day	34
4.20	SVM Polynomial Kernel results of top loser stock for 15% profit in 60 day	34
4.21	SVM Radial Kernel results of top gainer stock for 15% profit in 60 day	35
4.22	SVM Radial Kernel results of top loser stock for 15% profit in 60 day	35
4.23	Random Forest results of top gainer stock for 15% profit in 60 day	36
4.24	Random Forest results of top loser stock for 15% profit in 60 day	36

Chapter 1

Introduction

Stock prediction and automated trading system generates buy/hold signal for investors and traders. Based on the stocks historical data, the system finds the rule for prediction and then generate the signals. One of the advantage of our automated system is to restrict emotion of the traders about stock, hence system trades automatically if certain criteria are satisfied.

1.1 Objective of Project

Automated trading system is also known as an algorithmic trading which analyze the stock data and buy/sell stocks by itself. Based on the analysis, it generates specific rule for each stock and this rules are used for generating the buy/sell signal. This system is directly connected to brokers, who has permissions to buy or sell the stock by itself or it may be permitted by the user privileges.

Stock price time-line data is available for generation the signals. List of technical indicator and it's calculation is available to system for calculate from stock data-set. It find the trading rules from large available data-set. User can also give the restriction for buy/sell the stock like stock name, stock category, time period for investment, minimum profit for investment. User can also chose list of technical indicators that use for finding the rules. If system find the rules and that rule permitted by the user then system takes and action about buy/sell stock. So, ultimately this system is use for the maximize the user's profit of investment in the stock market.

1.2 Scope

Automated Stock market trading system is totally based on prediction using past data. When user start using this system, system asks some data for prediction. The data required to input by the user is amount to be invested, minimum profit, maximum profit, maximum loss, and maximum time duration for investment. From these input parameters and past data set, system designs the strategy for individual stock for individual user. This system generate only buy/hold signal from the generated data. Sell signal is generating from the user's input data such as time duration, minimum profit and maximum loss. So, by this way automated stock market trading system works to make maximum profit by minimum human intervention.

1.3 Output

For development of automated trading system, stock market prediction must be required. There are two ways to predict the stock i) to predict the stock price and ii) to generate the buy or sell signal for the stock. In this study buy/sell signal generation theory has been used for the stock prediction. There are two types of analysis for buy/sell signal generation i)Fundamental Analysis and ii)Technical Analysis. Fundamental analysis is based on company's profile and assets in market, while technical analysis totally depends on company's stock price in market, and volume trade on that particular price. In this study model has been developed based on technical analysis. There are ten technical indicator used to develop the model. These indicators are used as a parameter for the prediction model. Machine learning classification techniques like Naive Bayes, Random Forest, Artificial Neural Network(ANN),Support Vector Machine(SVM),k-Nearest Neighbour(k-NN) has been used to generate the buy/hold signal. The Sell signal is generated by users parameter like minimum profit, maximum profit, maximum loss and time period of investment.

Chapter 2

Literature Survey

Various literature has been studied in order to understand the amount of work done in this field. Since the existence of stock markets, a lot of research had been done in developing models to make predictions on stock price movements. Professional investors favor two dominant schools of thought on investing which are fundamental analysis and technical analysis.

2.1 Fundamental analysis

Fundamental analysis analyze the financial condition or health of particular company on time instance. It also analyze company's condition with respect to it's competitors on same category. Basic criteria that analyze under fundamental analysis are interest rate, production, future contract, employment, government policies, GDP, management, manufacturing etc. Financial analysis evaluated based on the historical as well as current data. The main goal of fundamental analysis is to predict the future of company's stock in the market. Fundamental analysis is performed on historical and present data, but with the goal of making financial forecasts. There are several possible objectives:

- To conduct a company capital stock valuation and predict its probable price evolution.
- To make a projection on its business performance.
- To evaluate its management and make internal business decisions.
- To calculate its credit risk.

Fundamental analysis is also calculate statistics from company’s financial annual report such as balance sheet, profit / loss statement , growth of the company, liquidity of investment are basic fundamental analysis attributes [4].

Text mining approach is used for fundamental analysis. Crawler find company’s fundamental attribute from newspaper and other financial news sources. By using text classifier, company’s news are categorized in to positive or negative news. Then based on historical data relation between news and stock price has been found. Automatic text classification is used to analyze the company’s fundamental statistics. Figure 2.1 from the source [1] is displaying the predictive systems consist of components such as news labeling, classifier input generation and classification.

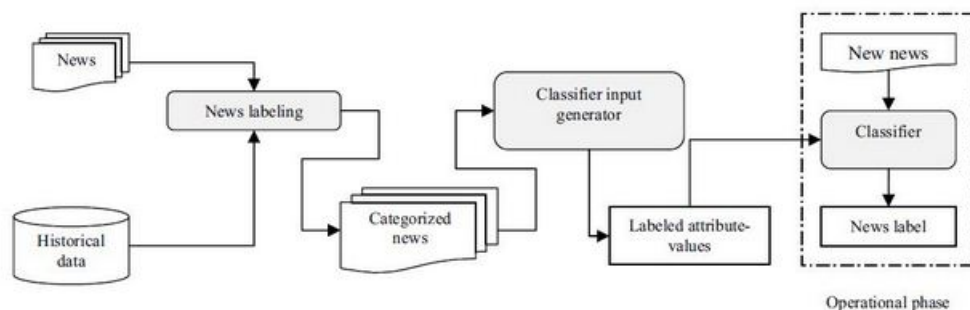


Figure 2.1: Stock market fundamental analysis[1]

There are two ways to assign company’s news label, manually and automated. In manual label generation, financial expert read the news and categorize it. In automated system, label is automatically generated based on the available training data-set. Main goal of the classifier is to classify company’s two separate categories, either good news or bad news about selected stock’s price and company’s status in the market.

2.2 Technical Analysis

Technical Analysis is used to forecast the future financial price movement based on stock historical price movement. Technical parameters do not predict stock price, but based on historical analysis, technical parameters can predict the stock movement (up/down) on current market situation over time. Technical analysis help investor to predict the stock price movement (up/down) in particular time period. Technical analysis uses a wide variety of charts that show price over time.

2.2.1 Strengths of Technical Analysis

Focus on Price and Volume :

Technical indicators are calculated only based on stock price, and volume trade on particular price. Based on the historical data and price movement, technical indicators forecasts about the stock. Even though there are knee-jerks present in the stock market, but technical indicators have enough strength to give hint about the price movement.

Supply, Demand, and Price Action :

Stock prices varies based on the supply and demand of the stock at current time instance in the market. Technical indicators are derived form the stock's high, low, close price and stock trade volume in the market. Thus technical indicators have strength to calculate supply/demand of particular stock in the market.

Support/Resistance :

Based on the supply and demand, technical indicators are able to find it's range. If supply of the stock is above range then it may be decrease in future and if it below range then it may increase in near future[5].

2.2.2 Technical Parameter

Technical indicators are one type of parameter that is based on stock price and trading volume. It has ability to predict stock future price level or stock price direction in market using past patterns. Some basic and most useful technical indicators are as below[6].

Relative Strength Index (RSI) :

The formula for calculating relative strength index is:

$$RSI = 100 - \frac{100}{1 + RS}$$
$$RS = \frac{\text{Average of given periods closes UP}}{\text{Average of given periods closes DOWN}}$$

RSI indicator compare stock gain to losses and forecast about stock is oversold or overbought. RSI return value in range of 0 to 100. In general scenario if RSI is above 70, stock

may be overbought and it indicate sell signal for particular stock and if RSI is bellow 30, stock may be oversold and it indicate buy signal. RSI threshold value for signal may change and it can be found accurately by analyzing stock data.

Moving average convergence divergence (MACD) :

The formula for calculating macd is:

$$MACDLine = 12DAY EMA - 26DAY EMA$$

$$SignalLine = 9DAY EMA of MACD Line$$

Where , EMA(Exponential Moving Average) is just one type of simple moving average(SMA) but in EMA more weight has been assigned for recent value.

when the MACD goes below the signal line,it indicates sell signal and when MACD goes above the signal line it indicate sell signal.

Stochastic Oscillator :

The formula for calculating stochastic Oscillator is:

$$\%K = 100 * \frac{(C - L14)}{(H14 - L14)}$$

where,

C = the most recent closing price

L14 = the low of the 14 previous trading sessions

H14 = the highest price traded during the same 14-day period.

$$\%D = 3 \text{ period moving average of } \%K$$

In general trend id %D is below 20 that indicate oversold means price is increase in near future and is indicate overbought means price is decrease in near future.

Williams %R :

Williams %R is a momentum indicator that is the inverse of the Fast Stochastic Oscillator. Also referred to as %R, Williams %R reflects the level of the close relative to the highest high for the look-back period. Williams %R is calculated as below.

$$\%R = \frac{(H14 - C)}{(H14 - L14)} * (-100)$$

where,

C = the most recent closing price

L14 = the low of the 14 previous trading sessions

H14 = the highest price traded during the same 14-day period.

%R return value between 0 to -100. If %R value above -20 it indicates sell signal and if %R value is below -80 it indicates buy signal for particular stock.

Money Flow Index (MFI) :

The Money Flow Index (MFI) indicator is calculated using stock price and it's volume trade on particular price. MFI is calculated as below.

$$TypicalPrice = \frac{(High + Low + Close)}{3}$$

$$RawMoneyFlow = TypicalPrice * Volume$$

$$MoneyFlowRatio = \frac{(14 \text{ Period Positive Money Flow})}{(14 \text{ Period Negative Money Flow})}$$

$$MoneyFlowIndex(MFI) = 100 - \frac{100}{(1 + MoneyFlowRatio)}$$

MFI is used to indicate overbought and oversold signal. If MFI is less than 20 that means oversold and if MFI greater than 80 that means overbought.

Bollinger Bands :

Bollinger Bands is calculated as below.

Middle Band = 20-day simple moving average (SMA)

Upper Band = 20-day SMA + (20-day standard deviation of price * 2)

Lower Band = 20-day SMA - (20-day standard deviation of price * 2)

Where, SMA is Simple Moving Average of particular time period.

When stock close price is above upper band then it indicates overbought signal and if stock close price below lower band then it indicates oversold signal.

Commodity Channel Index (CCI) :

The Commodity Channel Index (CCI) is use to find the recent trends in stock market.

$$CCI = \frac{(Typical\ Price - 20\ Period\ SMA\ of\ TP)}{(0.015 * Mean\ Deviation)}$$

where,

$$Typical\ Price\ (TP) = (High + Low + Close)/3$$

In general trend if CCI above 100 it indicates uptrend and if CCI below -100 it indicates downtrend.

On-Balance Volume (OBV) :

On Balance Volume (OBV) is volume based indicator that is used to find buying and selling trend of stock in stock market. Calculation of OBV is as below.

If the closing price is above the prior close price then:

$$Current\ OBV = Previous\ OBV + Current\ Volume$$

If the closing price is below the prior close price then:

$$Current\ OBV = Previous\ OBV - Current\ Volume$$

If the closing prices equals the prior close price then:

$$Current\ OBV = Previous\ OBV\ (no\ change)$$

Momentum :

Momentum is the measurement of the speed or velocity of price changes.

$$M = V - Vx$$

where,

V is the latest price, and Vx is the closing price of x number of days ago.

Momentum measures the rate of the rise or fall in stock prices. From the standpoint of trending, momentum is a very useful indicator of strength or weakness in the issue's price.

Price Rate Of Change - ROC :

PRoC indicator finds percentage of change in most recent price to the stock price of n period ago. Calculation of PRoC indicator is as below.

$$PRoC = \frac{(Closing\ Price\ Today - Closing\ Price\ n\ Periods\ Ago)}{Closing\ Price\ of\ n\ Periods\ Ago}$$

In general trend value greater than zero to indicate an increase in upward momentum and a value less than zero to indicate an increase in selling pressure.

Technical Parameter Used in Research Papers :

Research Paper	Technical parameter
[3]	Posvol, Negvol, OBV, RSI, MACD, Momentum, %K, %D, Williams %R, Bollinger bands, MA
[5]	RSI, %K, %D
[6]	RSI, MACD, MA
[7]	RSI, %K, %D, Bollinger bands, MA
[8]	OBV, RSI, MACD, Momentum, %K, %D, Williams %R, CCI
[2]	RSI, MACD, Momentum, %K, %D, Williams, MA
[9]	OBV, RSI, MACD, %K, %D, Williams %R, Bollinger bands, CCI, MFI, ATR
[10]	OBV, RSI, MACD, %K, %D, Williams %R, Bollinger bands, MA, EMA
[11]	RSI, MACD, %K, %D
[12]	RSI, MACD, PRoC, MA,

Table 2.1: Technical Parameter

2.3 Data Processing

For generating of the stock decision ten years has been taken from BSE India website (<http://www.bseindia.com/>). In this study daily basis data of Reliance Industry Ltd has been taken from 1st January 2005 to 31st December 2014. Data set attribute that are used to calculate the technical parameter such as open price, close price, high price, low price and volume trade on daily basis. After calculating the technical parameter Decision(class attribute) such as buy/hold has been generated. This buy/hold decision has been generated based on investor parameter such as investment time duration(in days) and desired profit (in percentage). And sell signal is generated if stock price cross boundary of minimum profit and maximum tolerance of loss. If stock price does not cross any of the boundary then sell signal is generated after completion of investment time period. All the technical parameter are used input of and buy/hold signal has been predicted as output[3].

2.3.1 Decision Parameter Generation

This study used to predict buy/hold signal based on user input data. For that we have to calculate actual decision for training data set. Suppose user wants to invest X amount for 30 days time period and he/she wants to earn 10% profit on investment. For training decision calculation if price cross above 10% in next 30 days then, we indicate buy label, else hold label. Sell signal is generated by user parameters like minimum profit, maximum loss and time period. If stock reaches to any parameter boundary then sell signal is generated.

2.3.2 Feature selection

It may happen that all the attributes are not equally important for all the stocks to generate decision using classifier. So, it is require to reduce the attributes for the best result. Weka-API has been used for development this system. `weka.attributeSelection.ClassifierSubsetEval` algorithm has been used to find out best used full attribute for chosen classifier and `weka.attributeSelection.InfoGainAttributeEval` algorithm used to find attribute rank according to their importance.

2.3.3 Outlier Detection

In large data-set it may happen that some values are far away from the mean of the whole data-set, this data is known as outlier. This outlier must be removed for better results. In this study `weka.filters.unsupervised.attribute.InterquartileRange` has been used for detecting the outliers and `weka.filters.unsupervised.instance.RemoveWithValues` has been used for removing the outliers from the data-set.

2.3.4 Discretization

All the feature(technical indicators) has numeric and continuous value. Every prediction model are not compatible with numeric value. Discretization is used to convert numeric and continuous value into district and finite range. In this study `weka.filters.unsupervised.attribute.Discretize` has been used for performing discretization process on our data-set.

2.3.5 Normalization

All the feature(technical indicators) in the data set are not in equal range. Large value feature apply more impact compared to small value. So,it is necessary to place all the feature at same scale. The values of technical indicator are normalized in range of $[-1, 1]$.

2.3.6 Sampling

In this study 20% sample data is to be used for design parameter selection of prediction model. This 20% sample data is generated such a way that, sample data contain same number of instances of each year and ratio of buy and hold decision remains same in whole data set and sample data set. Then this 20% sample data set is further divides into two part. The ratio of buy and hold decision maintain same in each 10% sample data. Purpose of design parameter selection to find optimized output. A 10% sample data is used for model training and another 10% sample data is used for testing. Various experiment has been performed on this sample by changing model's various design parameters. Design parameters are selected by evaluating error rate of of model on test sample data. After finding design parameter, all the prediction model such as Naive bayes, ANN, SVM, k-NN, Random forest has been trained using 80% of entire dataset and performance of all

the prediction model has been evaluated on rest 20% of entire dataset.

2.4 Related Work

Machine learning classification algorithm is successfully used for financial decision generation. Naive Bayes, Artificial Neural Network(ANN), Support Vector Machine(SVM), k-Nearest Neighbour(k-NN) and Random Forest is most widely used classification algorithm. The main contributions of this study is to demonstrate and verify the predictability of financial decision using this machine learning algorithm and technical analysis.

Naive Bayes is very basic, fast and most popular classification algorithm.It is based on bayesian theorem. Naive bayes takes feature vector and respected class label as input for training, and then predict class for unknown feature vector. In naive bayes algorithm all the feature vector is independent to each other. So main advantage of naive bayes algorithm is each feature has capacity to contribute independently to generate decision[10].

Artificial Neural Network(ANN) is a machine learning technique that is developed by simulating the biological nervous systems such as the human brain. It is implemented using network of neurons[12]. The multilayer perceptron is one of the most widely implemented artificial neural network. Two important characteristics of the multilayer perceptron are: its nonlinear processing elements (PEs) and their massive inter connectivity, i.e. any neurons of a layer is connected to all the neurons of the next layer[13].

Support Vector Machine(SVM) is a classification algorithm that create set of hyperplane with maximum margin between two class. SVM is a binary classifier but it works for more than two class using one vs all strategies. Linear and nonlinear kernel function is used for creating the hyperplane[2]. SVM has been also successfully applied to predict stock price index and its movements. Nair et al.[4] have used SVM to predict the change of daily stock price direction in the Korea composite stock price index (KOSPI). Jheng-Long Wu et al.[8] have used Support Vector Regression(SVR) technical for intraday stock price prediction with the help of fundamental and technical analysis.

k-Nearest Neighbour(k-NN) is a simple and extremely fast classification algorithm, that classify instance according to the matched training tuples. Teixeira et al.[7] have been predicted stock trend using k-NN classifier and technical analysis. Euclidean distance has been used to find the similarity in training pattern.

Random Forest is ensemble learning algorithm that has ability to built model by create n number of trees using sample data with replacement. And then predict test data by get vote from all the n number of trees. So, it is the hybrid method of bagging and voting. Ash Booth et al.[14] predicted stock market return using random forest regression technique. While Yanru Xu et al.[15] used random forest algorithm to selection feature for trend prediction in stock market.

Table 2.2 describe the classification technique are research paper.

Techniques Used in Literature :

Research Paper	Techniques
[3]	Dimension Reduction, ANN
[1]	Text Mining Approach for fundamental Analysis
[16]	Accuracy Analysis Using Kappa Measure
[17]	Stock market trend Analysis using charts.
[5]	Technical Analysis using Fuzzy Logic
[7]	Stop loss and Stop gain , k-NN
[8]	Technical indices and Sentimental indices, Stepwise Regression Analysis(SRA), SVR model
[2]	Sampling of Data , ANN(3 layered) , SVM
[9]	Naive Bayes Classification
[10]	Naive Bayes Classification , SVM
[11]	Random Forest Theory
[12]	ANN , Rough Set Predictions Model
[18]	Linear Regression and Non-Linear Regression
[19]	Random Forest Classification
[13]	ANN, Dynamic ANN
[14]	Regression using Random Forest Theory
[15]	Features Selection, SVM, Random Forest Classification

Table 2.2: Techniques used for stock market prediction

Chapter 3

Prediction Model

Stock price prediction is the act of trying to determine the future value of a company stock. Researchers try to predict future stock price or future stock trends in the market. Machine Learning algorithms are used for stock market prediction models. There are several machine learning algorithms available for stock market prediction, i.e. Naive Bayes Classification, Artificial Neural Network (ANN), Support Vector Machine (SVM), Support Vector Regression (SVR). Stock market technical parameters have been calculated in the previous section, and these parameters are used as input variables, and the output is the future trend of the stock price [20].

3.1 Naive Bayesian Classification

Naive Bayes classification is based on Bayes' theorem. Bayes' theorem is stated mathematically as follows.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

$P(A)$ and $P(B)$ are the probabilities of A and B independent of each other.

$P(A|B)$ and $P(B|A)$ are conditional probabilities, which are the probability of A given that B is true and the probability of B given that A is true, respectively.

In this study, hypothesis B is the probability of class attribute (decision) buy/hold, and A is the input dataset (technical parameter). $P(B|A)$ is the conditional probability of event B occurring when class attribute A is true. Assume $A_1, A_2, A_3, \dots, A_m$ are the technical parameters, and A is the class attribute, then the probability of each event with respect to class attribute

is calculated as below.

$$P(A_i|B) = (P(B|A_i)P(A_i))/P(B) = P(A_i)P(B_1, B_2, B_3, \dots, B_m|A_i)$$

In Naive Bayes classification this classification method all the attributes values have independent effect on the class attribute. So,

$$P(A_i|B) = P(A_i)P(B_1|A_i)P(B_2|A_i)\dots P(B_m|A_i)$$

Main advantage of this model is each attribute has capacity to contribute individually for decide the class attribute.

In this study all the attribute(technical parameter) has numeric and continuous value. For better accuracy and fast computing this technical parameter has been converted into district value. After calculating each class probability class label of observation B is defined as class label Ci, if following condition is satisfied. $P(A_i)P(B|A_i) > P(A_j)P(B|A_j)$

So, by this way buy/hold decision has been generated from technical parameters using Naive Bayes classification algorithm.

Naive Bayesian Classification for stock market prediction :

The naive Bayesian classifier, or simple Bayesian classifier, works as follows:

1. Let D be a training set of tuples and their associated class labels. As usual, each tuple is represented by an n-dimensional attribute vector, $X = (x_1, x_2, \dots, x_n)$, depicting n measurements made on the tuple from n attributes, respectively, A_1, A_2, \dots, A_n . here $x_1, x_2, x_3 \dots x_n$ is a day wise stock data for n days and A_1, A_2, \dots, A_{n-1} is a technical parameter (i.e. RSI,MACD etc) and A_n is decision buy or hold which is describe in above section.
2. Calculate probability actual value of training data set of buy and hold signal separately.
3. Calculate probability of each technical indicator with actual decision with buy and hold both.
4. Then calculate total probability of buy and total probability of hold separately and generate decision based on this value.

3.2 k-Nearest-Neighbor Classifiers

The k-nearest-neighbor method is widely used in the area of pattern recognition. Nearest-neighbor classifiers compare given test tuples to the training dataset, and find its best similar according to it. The training tuples are described by n features. Each attribute represents a point in an n-dimensional space. In this way, all of the training attributes are saved in an n-dimensional pattern space. When given an unknown attribute, a k-nearest-neighbor classifier searches the pattern space for the k training attributes that are closest to the unknown attributes. Closeness is defined in terms of a distance metric, such as Euclidean distance. The Euclidean distance between two points or tuples, say, $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ and $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$ is $dist(X_1, X_2) = \sqrt{(\sum(x_{1i} - x_{2i})^2)}$. For stock market prediction we can use technical indicator for prediction of decision. k-NN model finds the closest instance for given test set of data[7].

In this study number of neighbors(k) is decided on experimental basis. Two 10% sample data-sets are used as training and testing for tuning the k-NN prediction model. Model tuning is done by applying different values of k such as 1,2,3,...,50. In this study k-NN design parameters(number of neighbour) has been found by performing experiment on Reliance Industry historical data from 01-01-2005 to 01-01-2015. Based on accuracy best value of k that gives minimum error is 10.

3.3 Artificial Neural Networks(ANN)

Artificial Neural Network is a network of interconnected neurons that change their states based on the given input. The weight of the neurons changes as per input and its present weight. Error in the predicted value is minimized using the back propagation technique. ANN should be initialized by the function $f : X \rightarrow Y$ and it changes according to back propagated error[2]. An ANN is typically defined by three types of parameters:

- The interconnection pattern between the different layers of neurons.
- The learning process for updating the weights of the interconnections.
- The activation function that converts a neuron's weighted input to its output activation.

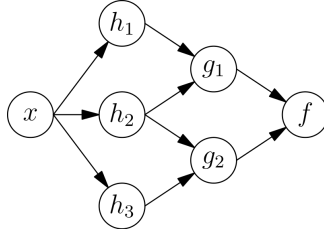


Figure 3.1: Artificial Neural Networks(ANN) [2]

In the Figure 2.2 from source [2], a neuron's network function $f(x)$ is defined as a composition of other functions $g_i(x)$, which can further be defined as a composition of other functions. This figure depicts such a decomposition of f , with dependencies between variables indicated by arrows. These can be interpreted in two ways. i) the input x is changed into a 3-dimensional vector h , which is then changed into a 2-dimensional vector g , which is at long last changed into f , and ii) the irregular variable $F = f(g)$ relies on the arbitrary variable $G = g(h)$, which relies on $H = h(x)$, which relies on the irregular variable X . This perspective is most ordinarily experienced in the context of graphical models. for this specific system network design, the parts of individual layers are free of one another. This characteristic enables a level of parallelism in the usage.

For generate decision about stock ANN architecture has been structured as shown in Figure 2.3.

As shown in Figure 2.3 three layered architecture has been designed for generating stock decision. ANN model consist input layer, hidden layer and output layer. All technical parameter is applied as a input to input layer of ANN. And this model generate buy/hold decision from a output layer. All the neurons in a layer are completely connected with all the neighbor layer neurons. The four design parameters are to be used to build ANN.

- Number of neurons : It is define as number of the neurons that are used in hidden layer. Number of input layer neurons is same as number of input(technical indicators) and number of output layer neurons are same as number of output(buy/sell). So, number of neurons in the hidden layer can be change as per the application.
- Epochs : An epoch is a measure of the number of times all of the training data are used once to update the weights.
- Momentum constant It is the momentum that applied to the weight at the starting

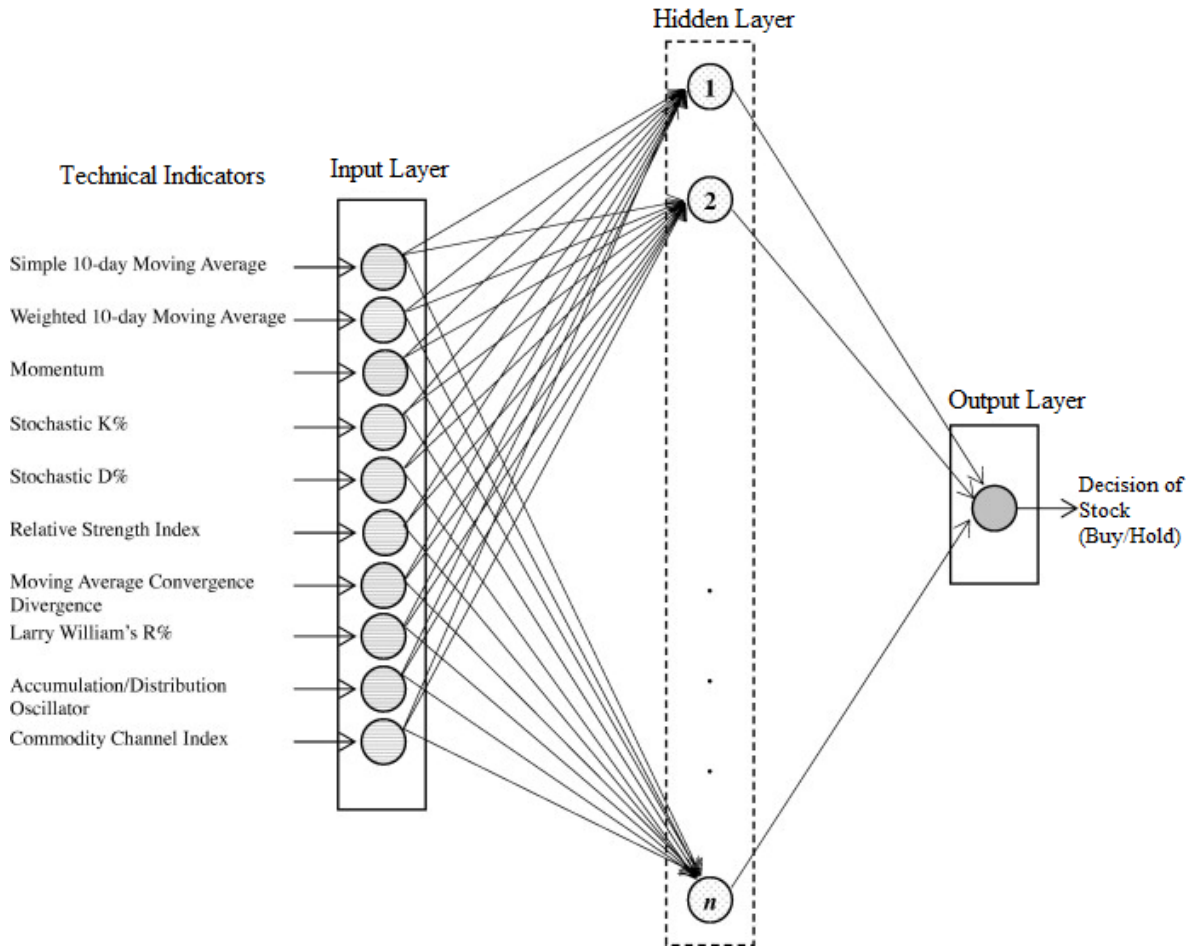


Figure 3.2: Structure of ANN for stock market decision generation[3]

point of training of neural network model.

- **Learning rate** It is the amount the weights are updated after each iteration of neural network.

For design parameter there is no thumb of rule to decide parameter value. So, It has been used all the parameter combination as shown in Table 3.1. And from all the combinations(900) best combination is chosen for the training purpose.

All 900 combination has been applied on on ANN prediction model and then find accuracy on test data-set. Best combinations are chosen and used for prediction model. In this study ANN design parameters are found by perform experiment on Reliance Industry historical data from 01-01-2005 to 01-01-2015. Best combination based on accuracy is as below.

Parameters	Values
Number of neurons (n)	10, 20,....., 100
Epochs (ep)	1000, 2000,....., 10000
Momentum constant (mc)	0.1, 0.2,....., 0.9
Learning rate (lr)	0.1

Table 3.1: ANN Design Parameter

Parameters	Values
Number of neurons (n)	100
Epochs (ep)	6000
Momentum constant (mc)	0.5
Learning rate (lr)	0.1

Table 3.2: Best ANN Design Parameter Based on Accuracy

3.4 Support Vector Machine(SVM)

support vector machines is supervised learning model that use to recognize pattern in data. Based on the training data set svm is capable to classify data into two or more categories. SVM construct the hyper plain to classification of data. Hyper plain should be either linear or nonlinear. SVM also capable to classify data in two or more then two dimension. There are infinite no of separation is possible in each dimension.SVM also use for regression. Using Support vector regression the next n value of stock can be found based on the training dataset.We can use technical indicator as input predict stock buy/sell decision[10]. Classification result using svm is depends on below function and quadratic programming is used to solve the function.

$$f(x) = sgn(\sum_{i=1}^N y_i \alpha_i . K(x, x_i) + b)$$

where x represents input data attribute(technical parameter) and y represents class attribute(buy/hold decision). alpha and b are constant. c is a regularization parameter which can be changed according to misclassification error.

There are two types of kernel function are used in support vector machine. i)Polynomial

Kernel ii) Radial Basis Kernel.

$$\text{PolynomialFunction} : K(x_i, x_j) = ((x^T * y) + \gamma)^d$$

$$\text{RadialBasisFunction} : K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

where γ is a constant and d is a degree of a function. So, Gamma constant (gamma), polynomial degree (d) and cost function (c) is used as design parameter in polynomial kernel function. Gamma constant (gamma) and cost function (c) is used as design parameter in radial basis kernel function. So, by changing these values prediction model can be configured according to requirement. Value of design parameter can be decided on experimental basis only. Two 10% sample data-set is used as training and testing for tuning the prediction model. Model tuning is done by applying different combination of design parameter which are as in table 3.3.

Parameters	Polynomial Kernel	Radial Basis Kernel
Degree(d)	1, 2, 3, 4	-
Gamma(γ)	0, 0.1, 0.2, ..., 5.0	0, 0.1, 0.2, ..., 5.0
Regularization parameter (c)	1, 10, 100	1, 10, 100

Table 3.3: SVM Design Parameter

There is no thumb of rule to decide the svm model parameter. From all 765 combination has been applied on prediction model and then find accuracy on test data-set. Best combinations are chosen and used for prediction model. In this study SVM design parameters are found by perform experiment on Reliance Industry historical data from 01-01-2005 to 01-01-2015. Best combination based on accuracy is table 3.4.

Parameters	Polynomial Kernel	Radial Basis Kernel
Degree(d)	5	-
Gamma(γ)	0.5	2.4
Regularization parameter (c)	10	10

Table 3.4: Best SVM Design Parameter based on Accuracy

3.5 Random Forest Classification

Random forest is one of the most popular classification techniques for stock market prediction. It is based on tree based learning. It is more efficient and accurate compare to other classification techniques. Random forest is ensemble learning technique that based on multidimensional decision tree for training.

Single decision is not able to predict accurately, so ensemble learning technique is used. It create n no of trees for the learning purpose that gain more accuracy and overcome on overfitting issue. Three design parameter such as number of tree(n), number of feature(nf) and maximum depth(d) of each tree have been used in random forest prediction model. Random forest algorithm randomly select nf number of feature for each n no of tree and each tree has maximum depth of d.

Parameters	Values
Number of tree(n)	10,20,30...,200
Number of feature (nf)	3,4,5,...,10
maximum depth(d)	3,4,5,...,10

Table 3.5: Random Forest Design Parameter

All 1280 combination have been applied on Random forest prediction model and then find accuracy on test data-set. Best combinations are chosen and used for prediction model. In this study random forest design parameters are found by perform experiment on Reliance Industry historical data from 01-01-2005 to 01-01-2015. Best combination based on accuracy is describe in table 3.6.

Parameters	Values
Number of tree(n)	190
Number of feature (nf)	7
maximum depth(d)	6

Table 3.6: Best Random Forest Design Parameter based on Accuracy

3.6 Ensemble Learning

Ensemble learning is a machine learning paradigm where multiple learners are trained to solve the same problem. In contrast to ordinary machine learning approaches which try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them to use.

Finally, if several learning schemes are available, it may be advantageous not to choose the best-performing one for dataset (using cross-validation) but to use them all and combine the results. Several machine learning techniques do this by learning an ensemble of models and using them in combination: Prominent among these are schemes called bagging, boosting, and stacking.

Chapter 4

EXPERIMENTAL RESULTS

For validate this system experiment is performed on stock based in BSE(India). Data of BSE-200 top ten gainer and top ten loser of the year 2014 has been from 1st January 2005 to 1st January 2015. Prediction model parameters are calculated as describe in chapter 3. Two different experiment has been performed and then find the accuracy of respected stock prediction. 80% data of each stock is used as training and remaining 20% data of each stock is used for testing.

4.1 Experiment 1:

User parameters has been added such as user want to invest X amount for 30 days and he wanted minimum 10% profit for his investment. Predefined model design parameters have been used to configure the models. Then model accuracy has been measured on top ten gainer and top ten loser stock of the year 2014 as below.

4.1.1 Naive Bayes

Stock Name	Accuracy	Precision	Recall
TVS Motor	70.1014	0.748	0.681
Aurobindo Pharma	70.489	0.689	0.699
Ashok Leyland	75.2108	0.743	0.781
Bharat Forge	73.5245	0.72	0.813
Gujarat Pipavav Port	79.9043	0.821	0.881
Eicher Motors	67.9595	0.722	0.725
Apollo Tyres	73.6931	0.714	0.692
IRB Infra.&Developer	75.0751	0.768	0.715
AIA Engineering	75.1693	0.789	0.728
HPCL	72.6813	0.778	0.734

Table 4.1: Naive bayes results of top gainer stock for 10% profit in 30 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	76.7285	0.691	0.742
JP Associate	68.5934	0.728	0.663
Jindal Steel & Power	72.6813	0.726	0.752
JP Power Ventures	70.8595	0.713	0.612
Reliance Comm	64.3519	0.657	0.668
Cairn India	81.5385	0.875	0.723
Mcleod Russel	80.0866	0.799	0.743
Reliance Power	72.5373	0.639	0.838
Sun TV Network	70.892	0.657	0.752
GMR Infrastructure	75.5501	0.777	0.855

Table 4.2: Naive bayes results of top loser stock for 10% profit in 30 day

4.1.2 k-NN

Stock Name	Accuracy	Precision	Recall
TVS Motor	82.6014	0.825	0.864
Aurobindo Pharma	89.0388	0.905	0.915
Ashok Leyland	85.6661	0.843	0.86
Bharat Forge	88.8702	0.874	0.905
Gujarat Pipavav Port	92.823	0.955	0.933
Eicher Motors	83.1366	0.838	0.818
Apollo Tyres	86.6779	0.888	0.896
IRB Infra.&Developer	93.0931	0.965	0.96
AIA Engineering	88.0361	0.896	0.864
HPCL	91.3997	0.896	0.929

Table 4.3: k-NN results of top gainer stock for 10% profit in 30 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	87.6897	0.857	0.899
JP Associate	81.3102	0.816	0.842
Jindal Steel & Power	82.4621	0.823	0.837
JP Power Ventures	88.6792	0.895	0.854
Reliance Comm	86.5741	0.859	0.889
Cairn India	93.0769	0.962	0.916
Mcleod Russel	92.2078	0.943	0.933
Reliance Power	91.9403	0.883	0.935
Sun TV Network	89.4366	0.873	0.907
GMR Infrastructure	79.4621	0.763	0.806

Table 4.4: k-NN results of top loser stock for 10% profit in 30 day

4.1.3 ANN

Stock Name	Accuracy	Precision	Recall
TVS Motor	74.6622	0.791	0.728
Aurobindo Pharma	81.9562	0.815	0.827
Ashok Leyland	77.9089	0.779	0.816
Bharat Forge	83.9798	0.843	0.89
Gujarat Pipavav Port	86.6029	0.874	0.926
Eicher Motors	71.6695	0.747	0.736
Apollo Tyres	76.054	0.773	0.789
IRB Infra.&Developer	88.8889	0.91	0.894
AIA Engineering	81.4898	0.824	0.759
HPCL	82.1248	0.755	0.815

Table 4.5: ANN results of top gainer stock for 10% profit in 30 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	81.2816	0.775	0.838
JP Associate	74.1811	0.7	0.91
Jindal Steel & Power	77.2344	0.755	0.761
JP Power Ventures	82.1803	0.832	0.767
Reliance Comm	79.8611	0.781	0.854
Cairn India	88.4615	0.912	0.798
Mcleod Russel	84.1991	0.84	0.8
Reliance Power	82.6866	0.814	0.917
Sun TV Network	78.4038	0.775	0.854
GMR Infrastructure	78.2396	0.754	0.802

Table 4.6: ANN results of top loser stock for 10% profit in 30 day

4.1.4 SVM Polynomial Kernel

Stock Name	Accuracy	Precision	Recall
TVS Motor	63.5135	0.643	0.743
Aurobindo Pharma	68.2968	0.722	0.797
Ashok Leyland	63.2378	0.614	0.676
Bharat Forge	73.3558	0.747	0.852
Gujarat Pipavav Port	77.9904	0.773	0.933
Eicher Motors	61.0455	0.609	0.525
Apollo Tyres	63.9123	0.643	0.666
IRB Infra.&Developer	76.5766	0.748	0.649
AIA Engineering	71.7833	0.714	0.555
HPCL	64.5868	0.637	0.884

Table 4.7: SVM Polynomial Kernel results of top gainer stock for 10% profit in 30 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	70.8263	0.615	0.635
JP Associate	62.0424	0.661	0.602
Jindal Steel & Power	71.8381	0.687	0.673
JP Power Ventures	68.5535	0.738	0.511
Reliance Comm	61.8056	0.62	0.699
Cairn India	81.5385	0.884	0.748
Mcleod Russel	71.2121	0.696	0.562
Reliance Power	74.0299	0.686	0.875
Sun TV Network	71.1268	0.705	0.833
GMR Infrastructure	67.2372	0.824	0.943

Table 4.8: SVM Polynomial Kernel results of top loser stock for 10% profit in 30 day

4.1.5 SVM Radial Kernel

Stock Name	Accuracy	Precision	Recall
TVS Motor	75.6757	0.808	0.728
Aurobindo Pharma	77.7403	0.822	0.859
Ashok Leyland	73.6931	0.723	0.759
Bharat Forge	80.7757	0.832	0.893
Gujarat Pipavav Port	79.4258	0.788	0.933
Eicher Motors	72.6813	0.727	0.675
Apollo Tyres	81.4503	0.836	0.849
IRB Infra.&Developer	80.7808	0.807	0.755
AIA Engineering	78.5553	0.774	0.66
HPCL	78.4148	0.784	0.878

Table 4.9: SVM Radial Kernel results of top gainer stock for 10% profit in 30 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	79.5953	0.738	0.797
JP Associate	73.6031	0.75	0.763
Jindal Steel & Power	74.3676	0.724	0.729
JP Power Ventures	79.6646	0.851	0.676
Reliance Comm	73.8426	0.734	0.783
Cairn India	79.4872	0.84	0.622
Mcleod Russel	81.1688	0.789	0.714
Reliance Power	76.7164	0.736	0.894
Sun TV Network	75.1174	0.71	0.793
GMR Infrastructure	73.1051	0.786	0.881

Table 4.10: SVM Radial Kernel results of top loser stock for 10% profit in 30 day

4.1.6 Random Forest

Stock Name	Accuracy	Precision	Recall
TVS Motor	93.9189	0.939	0.95
Aurobindo Pharma	94.4351	0.954	0.958
Ashok Leyland	92.5801	0.927	0.937
Bharat Forge	94.0978	0.947	0.961
Gujarat Pipavav Port	94.7368	0.943	0.978
Eicher Motors	93.2546	0.931	0.921
Apollo Tyres	92.5801	0.934	0.936
IRB Infra.&Developer	92.1922	0.938	0.927
AIA Engineering	93.4537	0.941	0.921
HPCL	92.9174	0.913	0.94

Table 4.11: Random Forest results of top gainer stock for 10% profit in 30 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	93.4233	0.937	0.957
JP Associate	92.8709	0.92	0.95
Jindal Steel & Power	94.6037	0.935	0.938
JP Power Ventures	93.0818	0.939	0.909
Reliance Comm	89.5833	0.882	0.925
Cairn India	93.3333	0.945	0.874
Mcleod Russel	94.8052	0.949	0.938
Reliance Power	93.7313	0.922	0.958
Sun TV Network	95.3052	0.965	0.976
GMR Infrastructure	90.709	0.909	0.93

Table 4.12: Random Forest results of top loser stock for 10% profit in 30 day

4.2 Experiment 2

User parameters has been added such as user want to invest X amount for 60 days and he wanted minimum 15% profit for his investment. Predefined model design parameters have been used to configure the models. Then model accuracy has been measured on top ten gainer and top ten loser stock of the year 2014 as below.

4.2.1 Naive Bayes

Stock Name	Accuracy	Precision	Recall
TVS Motor	72.069	0.769	0.737
Aurobindo Pharma	80.5508	0.827	0.791
Ashok Leyland	81.4114	0.858	0.78
Bharat Forge	81.5835	0.808	0.863
Gujarat Pipavav Port	90.8629	0.917	0.838
Eicher Motors	81.4114	0.86	0.82
Apollo Tyres	82.9604	0.817	0.781
IRB Infra.&Developer	80.3738	0.735	0.856
AIA Engineering	82.5986	0.799	0.767
HPCL	76.0757	0.771	0.69

Table 4.13: Naive bayes results of top gainer stock for 15% profit in 60 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	82.6162	0.798	0.877
JP Associate	73.9645	0.833	0.77
Jindal Steel & Power	78.4854	0.839	0.745
JP Power Ventures	83.6559	0.828	0.728
Reliance Comm	77.8571	0.787	0.765
Cairn India	83.0688	0.719	0.836
McLeod Russel	84	0.818	0.811
Reliance Power	75.8514	0.781	0.638
Sun TV Network	75.3623	0.729	0.751
GMR Infrastructure	79.597	0.817	0.707

Table 4.14: Naive bayes results of top loser stock for 15% profit in 60 day

4.2.2 k-NN

Stock Name	Accuracy	Precision	Recall
TVS Motor	86.7241	0.888	0.88
Aurobindo Pharma	92.9432	0.942	0.921
Ashok Leyland	90.0172	0.906	0.906
Bharat Forge	91.3941	0.905	0.939
Gujarat Pipavav Port	93.9086	0.961	0.926
Eicher Motors	88.296	0.879	0.93
Apollo Tyres	91.222	0.91	0.898
IRB Infra.&Developer	93.7695	0.928	0.928
AIA Engineering	89.0951	0.872	0.857
HPCL	93.9759	0.95	0.937

Table 4.15: k-NN results of top gainer stock for 15% profit in 60 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	92.4269	0.934	0.915
JP Associate	86.9822	0.896	0.837
Jindal Steel & Power	87.6076	0.873	0.901
JP Power Ventures	92.4731	0.928	0.895
Reliance Comm	89.2857	0.903	0.895
Cairn India	94.1799	0.896	0.944
Mcleod Russel	92	0.908	0.908
Reliance Power	92.2601	0.909	0.85
Sun TV Network	90.5797	0.901	0.896
GMR Infrastructure	85.6423	0.87	0.796

Table 4.16: k-NN results of top loser stock for 15% profit in 60 day

4.2.3 ANN

Stock Name	Accuracy	Precision	Recall
TVS Motor	80.8621	0.868	0.787
Aurobindo Pharma	87.2633	0.875	0.881
Ashok Leyland	82.7883	0.836	0.841
Bharat Forge	87.7797	0.878	0.898
Gujarat Pipavav Port	92.8934	0.946	0.897
Eicher Motors	83.3046	0.848	0.875
Apollo Tyres	87.6076	0.864	0.839
IRB Infra.&Developer	91.2773	0.868	0.942
AIA Engineering	87.239	0.846	0.824
HPCL	87.6076	0.894	0.867

Table 4.17: ANN results of top gainer stock for 15% profit in 60 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	90.3614	0.894	0.918
JP Associate	83.8264	0.852	0.755
Jindal Steel & Power	83.3046	0.816	0.892
JP Power Ventures	91.3978	0.927	0.895
Reliance Comm	85.9524	0.868	0.855
Cairn India	90.7407	0.855	0.924
Mcleod Russel	86.4444	0.874	0.882
Reliance Power	88.8545	0.86	0.756
Sun TV Network	87.9227	0.859	0.886
GMR Infrastructure	86.9018	0.885	0.822

Table 4.18: ANN results of top loser stock for 15% profit in 60 day

4.2.4 SVM Polynomial Kernel

Stock Name	Accuracy	Precision	Recall
TVS Motor	62.7586	0.629	0.862
Aurobindo Pharma	68.3305	0.74	0.603
Ashok Leyland	69.191	0.667	0.841
Bharat Forge	79.1738	0.758	0.901
Gujarat Pipavav Port	85.7868	0.826	0.603
Eicher Motors	69.3632	0.717	0.8
Apollo Tyres	66.2651	0.631	0.573
IRB Infra.&Developer	75.7009	0.723	0.712
AIA Engineering	77.9582	0.723	0.629
HPCL	70.9122	0.737	0.709

Table 4.19: SVM Polynomial Kernel results of top gainer stock for 15% profit in 60 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	74.8709	0.711	0.846
JP Associate	69.6252	0.813	0.76
Jindal Steel & Power	76.42	0.783	0.78
JP Power Ventures	74.8387	0.792	0.707
Reliance Comm	71.4286	0.702	0.63
Cairn India	84.9206	0.775	0.884
Mcleod Russel	77.3333	0.765	0.768
Reliance Power	78.9474	0.776	0.591
Sun TV Network	75.6039	0.742	0.731
GMR Infrastructure	75.5668	0.751	0.548

Table 4.20: SVM Polynomial Kernel results of top loser stock for 15% profit in 60 day

4.2.5 SVM Radial Kernel

Stock Name	Accuracy	Precision	Recall
TVS Motor	81.7241	0.85	0.829
Aurobindo Pharma	79.346	0.858	0.722
Ashok Leyland	78.1411	0.799	0.786
Bharat Forge	80.3787	0.767	0.914
Gujarat Pipavav Port	84.264	0.818	0.588
Eicher Motors	79.5181	0.791	0.89
Apollo Tyres	81.4114	0.831	0.814
IRB Infra.&Developer	85.6698	0.866	0.791
AIA Engineering	79.5824	0.749	0.681
HPCL	83.3046	0.824	0.757

Table 4.21: SVM Radial Kernel results of top gainer stock for 15% profit in 60 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	76.7642	0.734	0.846
JP Associate	79.4872	0.846	0.765
Jindal Steel & Power	76.5921	0.778	0.793
JP Power Ventures	82.3656	0.827	0.733
Reliance Comm	79.0476	0.782	0.745
Cairn India	81.746	0.757	0.888
Mcleod Russel	81.1111	0.804	0.807
Reliance Power	78.9474	0.764	0.551
Sun TV Network	76.57	0.773	0.705
GMR Infrastructure	81.3602	0.799	0.643

Table 4.22: SVM Radial Kernel results of top loser stock for 15% profit in 60 day

4.2.6 Random Forest

Stock Name	Accuracy	Precision	Recall
TVS Motor	95.8621	0.97	0.958
Aurobindo Pharma	97.074	0.977	0.967
Ashok Leyland	96.3855	0.968	0.964
Bharat Forge	96.2134	0.945	0.987
Gujarat Pipavav Port	97.4619	0.977	0.956
Eicher Motors	96.0413	0.965	0.968
Apollo Tyres	95.0086	0.963	0.96
IRB Infra.&Developer	95.0156	0.924	0.964
AIA Engineering	94.4316	0.95	0.948
HPCL	95.1807	0.949	0.933

Table 4.23: Random Forest results of top gainer stock for 15% profit in 60 day

Stock Name	Accuracy	Precision	Recall
Bhushan Steel	96.0413	0.956	0.966
JP Associate	95.069	0.955	0.929
Jindal Steel & Power	95.8692	0.962	0.962
JP Power Ventures	96.129	0.957	0.937
Reliance Comm	95	0.942	0.935
Cairn India	96.0317	0.938	0.968
Mcleod Russel	95.3333	0.935	0.934
Reliance Power	94.4272	0.936	0.898
Sun TV Network	95.4106	0.944	0.959
GMR Infrastructure	95.7179	0.955	0.93

Table 4.24: Random Forest results of top loser stock for 15% profit in 60 day

Chapter 5

Conclusion and Future Scope

5.1 Conclusion

This study shows how stock market decision is predicted using technical analysis. It also presents how machine learning and data mining techniques has been used for generate stock signal(buy/hold/sell) with the use of technical analysis. In this study we have just predicted buy/hold signal for stocks and it based on user input perimeter like time duration for investment and minimum profit that user want. Various data mining technique like feature selection, outlier detection, discretization, normalization is use for data preprocessing. In this study we also shows the result of top ten loser and top ten gainer of BSE-200 for the year 2014 using classification technique like naive Naves, k-Nearest Neighbour(k-NN), Artificial Neural Network(ANN), Support Vector Machine(SVM) and Random forest classification technique. Random Forest classification algorithm gives better results compare to all other algorithm. So, by this way automated trading system works by predicting the stock using data analysis.

5.2 Future Scope

After generating the decision using different classification algorithm individually it can be combine in to single decision using ensemble learning. After generating decision for individual stock it can be applied for create a portfolio. Risk management feature can also be implement with portfolio.

References

- [1] S. M. Azadeh Nikfarjam, Ehsan Emadzadeh, “Text mining approaches for stock market prediction,” *IEEE Computer and Automation Engineering*, 2010.
- [2] O. K. B. Yakup Kara, Melek Acar Boyacioglu, “Predicting direction of stock price index movement using artificial neural networks and support vector machines,” *Elsevier Expert Systems with Applications*, 2011.
- [3] S. B. V. M. Binoy B. Nair, M. Minuvarthini, “Stock market prediction using a hybrid neuro-fuzzy system,” *IEEE International Conference on Advances in Recent Technologies in Communication and Computing*, 2010.
- [4] V. M. Binoy B. Nair, N.Mohana Dharini, “A stock market trend prediction system using a hybrid decision tree-neuro-fuzzy system,” *IEEE International Conference on Advances in Recent Technologies in Communication and Computing*, 2010.
- [5] J. T. Simon Fong, Yain-Whar Si, “Trend following algorithms in automated derivatives market trading,” *Elsevier Expert Systems with Applications*, 2012.
- [6] H. K. Izumi, F.Toriumi, “Evaluationofautomated-tradingstrategiesusinganartificialmarket,” *Elsevier Neurocomputing*, 2009.
- [7] A. L. I. d. O. Lamartine Almeida Teixeira, “A method for automatic stock trading combining technical analysis and nearest neighbor classification,” *Elsevier Expert Systems with Applications*, 2010.
- [8] P.-C. C. Jheng-Long Wu, Liang-Chih Yu, “An intelligent stock trading system using comprehensive features,” *Elsevier Applied Soft Computing*, 2014.
- [9] M. S. R. Sheikh Shaugat Abdullah, “Stock market prediction model using tpws and association rules mining,” *IEEE Computer and Information Technology*, 2012.

- [10] J. K. A. Hyun Joon Jung, "A binary stock event model for stock trends forecasting," IEEE International Conference on Intelligent Systems Design and Applications, 2011.
- [11] Y. Ye, "The information content of technical trading rules: Evidence from us stock markets," IEEE Business Management and Electronic Information, 2011.
- [12] A. K. K. Shipra Banik and M. Anwer, "Dhaka stock market timing decisions by hybrid machine learning technique," IEEE Computer and Information Technology, 2012.
- [13] T. U. D. Erkam Guresen, Gulgun Kayakutlu, "Using artificial neural network models in stock market index prediction," Elsevier Expert Systems with Applications, 2011.
- [14] F. M. Ash Booth, Enrico Gerding, "Automated trading with performance weighted random forests and seasonality," Elsevier Expert Systems with Applications, 2014.
- [15] L. L. Yanru Xu, Zhengui Li, "A study on feature selection for the trend prediction of stock trading price," IEEE International Conference on Computational and Information Sciences, 2013.
- [16] S. S. Rahul Gupta, Nidhi Garg, "Stock market prediction accuracy analysis using kappa measure," IEEE International Conference on Communication Systems and Network Technologies, 2013.
- [17] J. R. Aseel Hmood, "Analyzing and predicting software quality trends using financial patterns," IEEE Computer Software and Applications Conference Workshops, 2013.
- [18] M. J. N. Han Lock Siew, "Regression techniques for the prediction of stock price trend," IEEE Statistics in Science, Business, and Engineering, 2012.
- [19] F. M. Ash Booth, Enrico Gerding, "Predicting equity market price impact with performance weighted ensembles of random forests," Computational Intelligence for Financial Engineering and Economics (CIFEr), 2014.
- [20] B. D. Aditya Gupta, "Stock market prediction using hidden markov models," IEEE Engineering and Systems, 2013.