2016 2nd International Conference on Next Generation Computing Technologies (NGCT-2016)
Dehradun, India 14-16 October 2016

Capital Market Forecasting By Using Sentimental Analysis

Rohan Srivastava[1], Soumil Agarwal[2], Dhruv Garg[3], Jagdish Chandra Patni[4]

Centre for Information Technology, UPES
Dehradun, India, [1]rohan2970srivastava@gmail.com,
[2]agarwalsoumil8@gmail.com, [3]dhruv.garg05@gmail.com, [4]patnijack@gmail.com

***Abstract*— When a customer invests into a stock market, he wants to attain higher profits in short period of time but with less amount of knowledge he has, this process gets very difficult. This situation often creates more losses to customers rather than profit which is fatal for their wealth. This economically dangerous position can be overcome with proper planning, proper management, apt amount of knowledge of market and useful analysis of previous data. The markets can be studied using their previous years of data and therefore predicting future trends. Market is also affected by market sentiment at a particular moment and if customer is aware of it, then he can take appropriate decisions regarding purchasing and selling of stocks in market. In this paper, we are going to demystify the stock market to help the customers to get better idea of upcoming market trends, polarity of stocks and present market sentiment in regards to the company he is investing.**

Keywords: knowledge; stock market; market sentiment; market trends;

I. INTRODUCTION

The capital stock of a corporation represents the equity stake of the owners of the stock i.e. the net value of the stock is the difference between the value of the assets, interest and the cost of liabilities incurred by the corporation. In order to obtain higher rates of return in a short span of time Individuals get influenced to invest certain portion of their savings into the stock market. The investors need to decide which companies to invest in which they do by analyzing various facts and figures and alas take a decision. However, in most cases people tend to take rash decisions as they lack time to collect sufficient information and they tend to lose more money than to gain. Stock market prediction is an art of trying to forecast the future value of a company's stock price in order to yield higher profits. It can be performed to do shorter span predictions or for longer span prediction of stock market price. There are two popular beliefs when it comes to stock market predictions which are the efficient market hypothesis which proposes that the stock prices are function of information, moreover every newly revealed information regarding a company's prospects gets immediately reflected in its stock price. The second belief is known as random walk which states that stock prices cannot be accurately predicted by looking at price history and each day's value is a deviation from central value which is random and unpredictable. In order to do stock market predictions we can broadly classify three main categories which are fundamental analysis where analysts study about the company by evaluating its past performances and credibility, technical analysis where future prices are predicted from past prices by using statistical techniques like exponential moving average or using k nearest neighbor and support vector machine to predict the polarity of stock i.e. if the price of stock will rise or fall compared to previous values without actually predicting the exact value. Lastly we can use artificial neural networks and data mining where large data sets are filled into systems and useful trends are predicted providing information regarding favorable or unfavorable returns. An investor requires a user friendly platform which he can use in order to predict the future expected price and polarity of stock both for long term and short term spans. Furthermore, the market sentiments are an appropriate source of studying the current trend regarding a company's stock and can prove very useful to predict the future values as well as to bridge the gap between the expected and actual values. In case of technical analysis, it is essential to compare the accuracy of various techniques as well as to determine the suitability of picking the best algorithm which provides results nearest to actual value for which a comparative study is done.

II. RELATED WORKS

Kun-Huang Huarng and Tiffany Hui-Kuang in [1] have applied neural networks in order to implement a new fuzzy time series model to make improvements in forecasting techniques. The difference in their approach is that they have included varying degrees of membership to establish fuzzy relationships to capture the relationships more properly. The use case used in this study is of Taiwan stock market predictions. Apart from this their model can also forecast stock rates even for out of sample observations.

Pawar D.D. and R.K. in [2] have discovered an efficient method for estimation of stock rates for given stocks due to various challenges faced while investing in stocks at a given time so as to get suitable returns. This is accomplished with the aid of artificial neural networks (ANN) in contrast to the traditional approach of time series analysis as neural networks can be used to easily extract useful information from larger data sets. Apart from this the paper also stressed on applications of ANN for stock market predictions.

Lim tan chew and Jing Tao Yao in [3] have utilized artificial neural networks for forecasting, classification and recognition of trends in stock markets. The use of ANN helps in decision making process as it is a good tool for forecasting. This paper also discussed a seven step neural network forecasting model building approach. Apart from this pre and post data analysis, sampling and training criteria were also touched upon.

Bohdan Pavlyshenko in [4] presents a software package for the data mining of Twitter micro blogs to provide an accurate stock market analysis. This package is created with the aid of R language and its appropriate packages and dependencies. The stock market charts have been compared with frequent sets of keywords in Twitter micro blogs messages.

Abdulsalam sulaiman olaniyi, adewole, kayoed s, Jimoh R.G in [5] have utilized one of the five methods called Moving Average (MA) method to find hidden pattern relationships which can predict future values of stock prediction variables with the help of time series data. The reason of using MA was

to reduce the fluctuations and to obtain trends with better accuracy.

K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam in [6] have used data mining to forecast the future closing stock prices whether they will increase or decrease and the role of global events on stock markets using numerical and graphical approach. This paper helps the investor to understand hidden patterns from historical data that can help in better future decision making. Here a set of five methods were combined namely Typical price (TP), Bollinger bands, Relative strength index (RSI), CMI and MA to predict the closing stock prices.

G. preethi and B. santhi in [7] had surveyed recent literature associated with the domain of Neural Network, Data Mining, Hidden Markov Model and Neuro-Fuzzy system which have been utilised in order to make predictions for fluctuation in the stock market. Neural Networks and Neuro-Fuzzy systems are the progressive machine learning techniques in stock market index prediction area. NN and Markov Model have been used extensively in the financial markets and for prediction of stock price. This paper proposes a stock market forecasting approach that is much better and accurate than the early generic approaches.

III. PROBLEM STATEMENT

When customers invest into stock market he can lose a quite amount of money if he didn't understands how share prices are fluctuating. In the simplest sense, investors buy shares at a certain price and can then sell the shares to realize capital gains. However, the investor will not realize a gain if the share price drops dramatically; in fact, the investor will lose money. Hence an investor can lose a large amount of money if he has no prior knowledge of how stock prices are changing. So by getting knowledge of previous stock prices customers can save themselves from losses and can even earn profit.

IV. DESIGN

Stock Market Prediction involves requirement of previous years of stock data of a particular company, which we have fulfilled by retrieving it from a finance website. After retrieval, this data is staged in R studio and converted into a dataframe. From the dataframe, appropriate column and rows are selected for running algorithms on it depending on amount of duration of data we want to take for running the algorithm.

For predicting future price of stock by exponentially weighting the data (close price) is done by first selecting the number of days of data to be taken as input then giving the highest weight to the latest data and decreasing the weight to older ones with farthest data getting the lowest weight. The weights depend on algorithm and the data for all the days are first multiplied with their corresponding weights and then performing their average.

For predicting future price of stock by simple averaging the data (closing price) is done by initially selecting the number of days of data to be taken as input and then adding all the data then performing their average.

For predicting next day's polarity of a stock whether the stock price of a company will go up or down classification can be used. We have used KNN algorithm that stores all available cases and classifies new cases based on a similarity measure. It uses majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function.

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

For predicting next day's polarity of a stock we have also used SVM algorithm which uses machine learning and builds a model that assigns new examples into one category or the other, making it a non – probabilistic binary linear classifier.

For predicting market sentiment of the company, we have used sentiment analysis which uses natural language processing and text analysis to identify and extract required information from our source of data.

V. IMPLEMENTATION

For our experimentation purpose we have used some parameters. We retrieved data in .csv format which we have staged in R studio and converted into a dataframe. From the dataframe, appropriate column and rows are selected for running algorithms on it depending on amount of duration of data we want to take for running the algorithm.

For predicting future price of stock by exponentially weighting the data, we took 10 days of data, gave weights 10 to 1 from latest to oldest data correspondingly. We multiplied the close price of data with their weights and took their average.

Figure 1 – Exponential Moving Average Algorithm

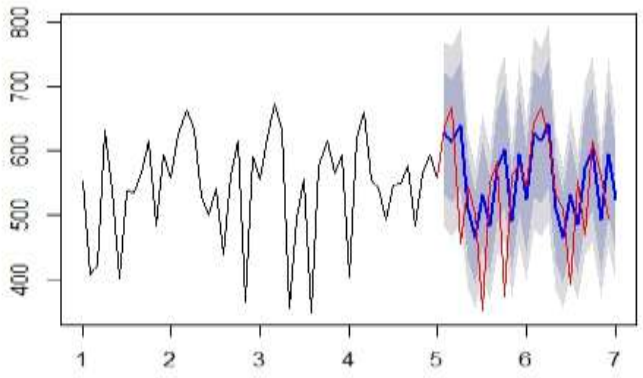


Figure 2 – Forecasts from ETS(M,Md,M)

For predicting future price of stock by simply averaging the data, we took 10 days of data, added the close price of all 10 days and took their average.

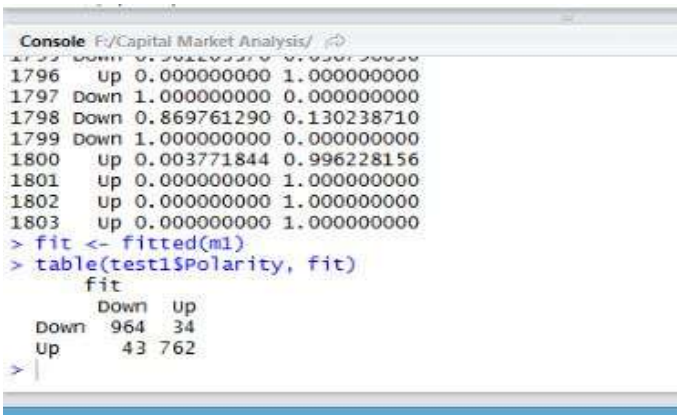
```
> library(forecast)
> data1=read.csv("table.csv")
> data2 = data1[1:10,5]
> weighted = 100
> sumer = 0
> emasum = 0
> for(i in seq(1,10,1)){
+   emasum = emasum + data2[i]
+ }
> emasum
[1] 1060.15
> result = emasum/10
> result
[1] 106.015
>
```

Figure 3 – Simple Moving Average Algorithm

For KNN, determined the value of parameter K = number of nearest neighbours beforehand. This value of k is all up to person to person. Then calculated the distance between the query and all the training data values. One can apply any distance algorithm. Then sorted the distances for each of the training data values and calculated the nearest neighbour based on the K (minimum distance). Finally used the majority of nearest neighbours as the prediction value.

```
library(caret)
library(kknn)
rp = read.csv("apple.csv")

Train <- createDataPartition(y= rp$Number.Of.Trades,p=0.25,list=FALSE)

test1 <- rp[Train,]
train1 <- rp[-Train,]
dim(train1)

m1 <- kknn(Polarity~.,train1,test1,k=13)
summary(m1)
fit <- fitted(m1)
table(test1$Polarity, fit)
```

Figure 4 – K Nearest Neighbor Algorithm



```
Console F:/Capital Market Analysis/
1796    Up 0.000000000 1.000000000
1797  Down 1.000000000 0.000000000
1798  Down 0.869761290 0.130238710
1799  Down 1.000000000 0.000000000
1800    Up 0.003771844 0.996228156
1801    Up 0.000000000 1.000000000
1802    Up 0.000000000 1.000000000
1803    Up 0.000000000 1.000000000
> fit <- fitted(m1)
> table(test1$Polarity, fit)
      fit
       Down  Up
  Down  964  34
  Up     43 762
>
```

Figure 5 – Prediction using KNN Algorithm

For SVM, segregated the data into two parts i.e. test set and training set. Then created the SVM classification object. Then trained the model using training sets and calculated our score.

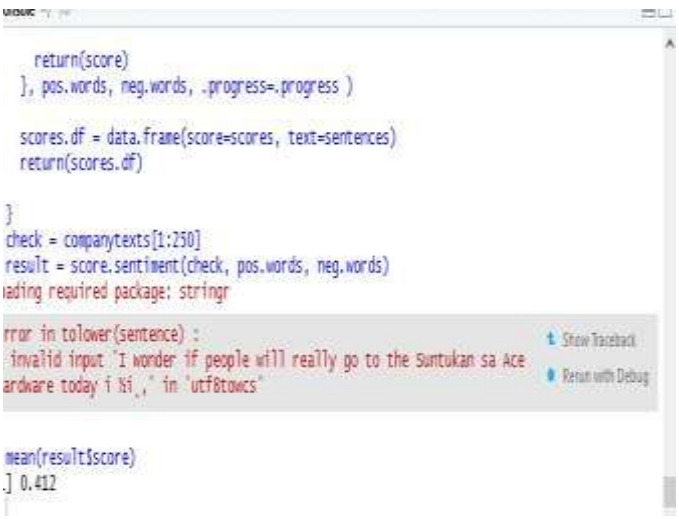
```
dataset = read.csv("table1.csv")

index <- 1:nrow(dataset)

testindex <- sample(index, trunc(length(index)*30/100))

testset <- dataset[testindex,]

trainset <- dataset[-testindex,]


model <- svm(Polarity~., data = trainset)
prediction <- predict(model, testset[,-9])
summary(model)
tab <- table(pred = prediction, true = testset[,9])
```

Figure 6 – SVM Classification

```
> prediction <- predict(model, testset[,-9])
Warning messages:
1: In rep.int(c(1, numeric(n)), n - 1L) :
  Reached total allocation of 3875Mb: see help(memory.size)
2: In rep.int(c(1, numeric(n)), n - 1L) :
  Reached total allocation of 3875Mb: see help(memory.size)
> tab <- table(pred = prediction, true = testset[,9])
> tab
      true
pred   Down   Up
  Down 1224  839
  Up     16   81
> View(trainset)
> View(trainset)
> classAgreement(tab)
$diag
[1] 0.6041667
```

Figure 7 – Prediction using SVM Algorithm

For our experimentation purpose we have retrieved about 250 tweets of the related company (xyz) using twitter API. Of all 250 tweets, every word of each tweet is passed through a couple of files, one contains positive words and other contains negative words. If the word exists in the file of positive words then the sentiment score is incremented by 1 and if the word exists in file of negative words then the sentiment score is decremented by 1. By checking all the words of a particular tweet a sentiment score of entire tweet is calculated and finally a mean sentiment score for all 250 tweets is calculated indicating market sentiment of the company.



```
    return(score)
  }, pos.words, neg.words, .progress=.progress )

  scores.df = data.frame(score=scores, text=sentences)
  return(scores.df)

}
check = companytexts[1:250]
result = score.sentiment(check, pos.words, neg.words)
ading required package: stringr

rror in tolower(sentence) :
 invalid input 'I wonder if people will really go to the Suntukan sa Ace
ardware today i %i ,' in 'utf8towcs'

mean(result$score)
] 0.412
```

Figure 8-Mean Sentiment Score

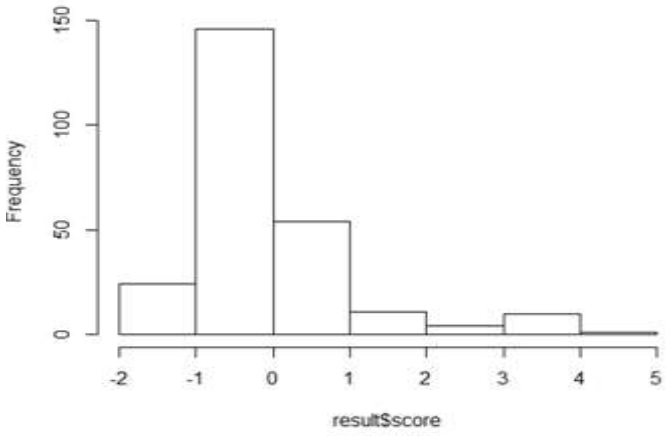


Figure 9-Histogram of Sentiment Score

VI. RESULT

For instance, took the stock price for a company whose actual share price on a particular day was **Rs. 105.80.** Then we tested the algorithms on previous days of stock data and calculated results which are:

•For predicting future price of stock by exponentially weighting the data, our algorithm result was **Rs. 106.47.**
•For predicting future price of stock by simply averaging the data, our algorithm result was **Rs. 106.015.**
•Using KNN algorithm, our predicted result was a confusion matrix with over 90% accuracy and following values:

	fit	
	Down	Up
Down	964	34
Up	43	762

•Using SVM algorithm, our predicted result was a confusion matrix with close to 60% accuracy and following values:

	true	
pred	Down	Up
Down	1224	839
Up	16	81

•The sentiment score was found to be **0.412**, which is a positive market sentiment of the company.

Hence, we predicted the company's next day close price, its polarity and the market sentiment of the company. All three of these information can be used by the investor to get a better insight of next day's trends before buying or selling of company's stocks.

REFERENCES

[1] Tiffany Hui-Kuang yu and Kun-Huang Huarng, "A Neural network-based fuzzy time series model to improve forecasting", Elsevier, 2010, p: 3366-3372.

[2] Das R.K. and Pawar D.D., "Application of Artificial Neural Network for stock market predictions: A review of literature" International Journal of Machine Intelligence, ISSN: 0975–2927, Volume 2, Issue 2, 2010, p-14-17.

[3] JingTao YAO and Chew Lim TAN,"Guidelines for Financial Prediction with Artificial neural networks".

[4] Bohdan Pavlyshenko , "Tweets Miner for Stock Market Analysis survey "

[5] Abdulsalam sulaiman olaniyi, adewole, kayoed, Jimoh R.G, "Stock Trend Prediction using Regression Analysis – A Data Mining Approach", AJSS journal, ISSN 2222-9833, 2010.

[6] K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam, "Financial stock market forecast using data mining Techniques", 2010, Proceedings of the international multiconference of engineers and computer scientists.

[7] G. preethi and B. santhi, "Stock market forecasting techniques: A survey ", Journal of Theoretical and Applied Information Technology, 15th December 2012, Vol. 46 No.1