Research Article

# An Intelligence prediction of stock market analysis using extreme dynamic data mining techniques Predictions

[1]*V.Rajesh Reddy, [2]Dr A. Gayathri

## ABSTRACT

Stock market, it's the place where everyone got interest in making money. But many people think that stock market predictions as gambling but what actually is, stock exchange movements are depends on capital gains and losses. Most people believes that stock exchange movements are unpredictable. But by using extreme data mining techniques on previous stock exchange data we can predict the future stock exchanges. This paper tries to help the investors in the stock market to decide the better timing for buying or selling stocks based on the knowledge extracted from the previous data of such stocks. In this paper we are using decision tree algorithm which is one of the data mining techniques. To build the proposed model, the CRISP-DM methodology is used over real previous data of two major companies.

*Keywords:* Data mining, Decision tree, data mining techniques, CRISP-DM methodology

## 1. INTRODUCTION

The stock market is essentially a non-linear, non-parametric system that is extremely hard to model with any reasonable accuracy. Now-a-days investors are trying to find a right path to predict stock prices and right timing to buy or sell the shares.

Recently, artificial intelligence and data mining techniques like decision trees and artificial neural networks have been applied to this area that means stock exchange area. Data mining refers to extracting or mining knowledge from large data sets. Some functionalities of data mining are as follows

a) Class description

b) Associations and Correlations

c) Classification

[1]*UG Scholar, Dept. of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India. rajeshreddy1525@gmail.com
[2]Associate Professor, Dept. of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India. gayathribala.sse@saveetha.com

d) Prediction

e) Clustering

f) Outlier and Deviation analysis.etc

Data classification can be done in many different methods, one of those methods is the classification by using Decision Tree. The main aim of this paper is analyze the previous data using decision tree technique in order to help the investors to know when to buy or when to sell their stocks.

## 2. LITERATURE REVIEW

Over the past two decades many changes have taken place in the field of financial market. Forecasting stock return is an important financial subject that has attracted researchers attention for many years

For that reason, several researchers have focused on technical analysis and using advanced math and science. Extensive attention has been dedicated to the field of artificial intelligence and data mining techniques. Some models have been proposed and implemented using the above mentioned techniques, the authors made an empirical study on building a stock buying/selling alert system using back propagation neural networks (BPNN), their NN was codenamed NN5. The system was trained and tested with past price data from Hong Kong and Shanghai Banking Corporation Holdings over the period from January 2004 to December 2005. The empirical results showed that the implemented system was able to predict short-term price movement directions with accuracy about 74%.

El-Baky proposed a new approach for fast forecasting of stock market prices. The proposed approach uses new high speed time delay neural networks (HSTDNNs). The authors used the MATLAB tool to simulate results to confirm the theoretical computations of the approach.

## 3. PROPOSED METHOD

Data mining methodology is designed to ensure a stable model that successfully addresses the problem it solve. To build the model that analyses the stock trends using the decision tree technique, the CRISP-DM (Cross Industry Standard Process for data mining) is used. This methodology was proposed by an European consortium of companies .This model consists of the following six steps

- Understanding the reason and objective of mining the stock prices.
- Understanding the collected data and how it is structured.
- Preparing the data that is used in the classification model.
- Selecting the technique to build the model.
-  Evaluating the model by using one of the well known evaluation methods.
- Deploying the model in the stock market to predict either selling or buying the stocks.
- Understanding the reason and objective of building the model

The main aim of this paper is to try to help the investors in the stock market to decide timing for buying or selling stocks based on the knowledge extracted from the previous prices of such stocks.

## Understanding the collected data

The Oracle database of Amman Stock Exchange (ASE) contains the historical prices of the 230 companies listed in the exchange from the year 2000. As the amount of such data is very large and complicated, so here we choose two companies listed in the exchange. We selected these two companies based on the following five criteria which represent the

- Companies' size and liquidity:
- Market capitalization,
- Days traded,
- Turnover ratio,
- Value traded and
- The number of shares traded.

The data collected contain 6 attributes. Below table shows the 6 attributes selected with their descriptions and their possible values .

| ATTRIBUTE | DESCRIPTION | POSSIBLE VALUES |
|---|---|---|
| Previous | Previous day close price of the stock | Positive, Negative, Equal |
| Open | Current day open price of the stock | Positive, Negative, Equal |
| Min | Current day minimum price of the stock | Positive, Negative, Equal |
| Max | Current day maximum price of the stock | Positive, Negative, Equal |
| Last | Current day close price of the stock | Positive, Negative, Equal |
| Action | The action taken by the investor on this stock | Buy,Sell |

## Preparing the data

At the beginning, when the data was collected, all the values of the attributes were numeric values. We can transform the numeric data to a higher-level concept so now all the values will become discrete.
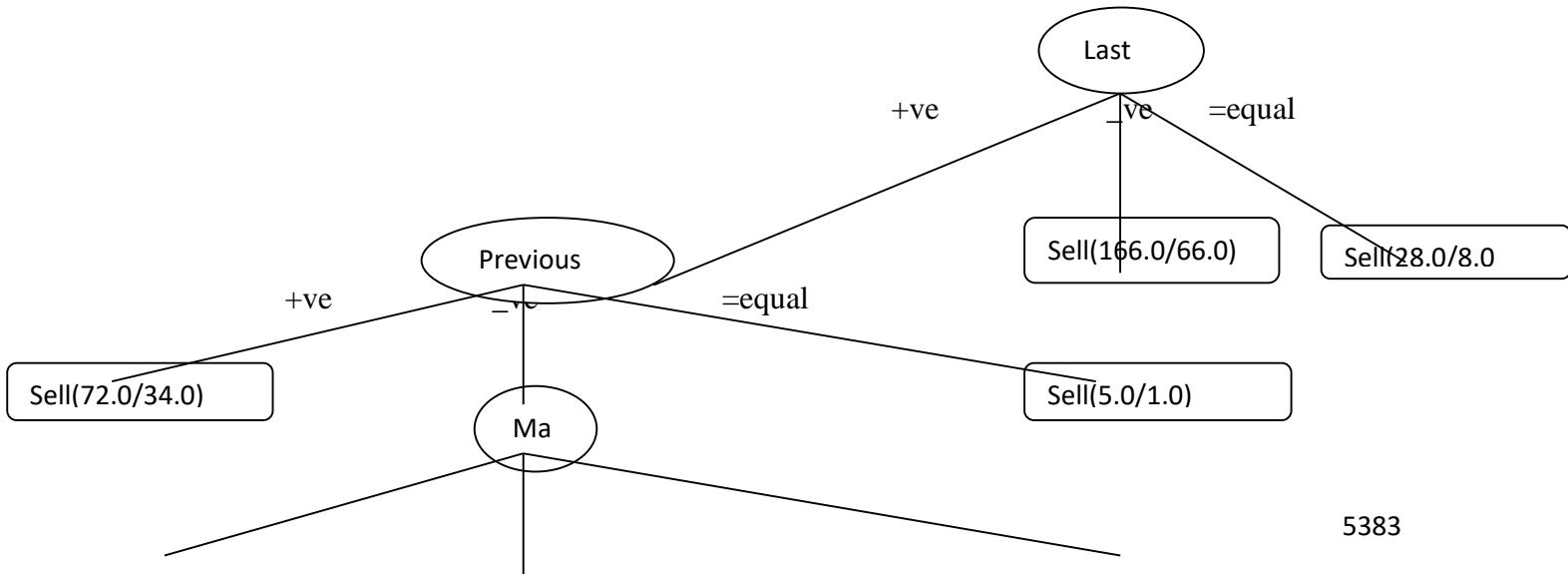
If the values of the attributes open, min, max, last were greater than the value of attribute previous for the same trading day, the numeric values of the attributes were replaced by the value Positive. If the values of the attributes mentioned above were less than the value of the attribute previous, the numeric values of the attributes were replaced by Negative. If the values of those attributes were equal to the value of the attribute previous, the values were replaced by the value Equal .Below tables show us a sample of the continuos numeric values of the data before selecting the 6 attributes and before generalizing them to descrete values ,while another table shows us the sample after selecting the 6 attributes and after transforming thme to discrete values.

| Previous | Open | Max | Min | Last | Action |
|---|---|---|---|---|---|
| 25.83 | 26.00 | 26.01 | 25.42 | 25.68 | Sell |
| 25.68 | 25.69 | 25.69 | 25.3 | 25.4 | Buy |
| 25.4 | 24.9 | 25.4 | 24.42 | 24.91 | Buy |
| 24.91 | 24.8 | 24.91 | 24.31 | 24.88 | Sell |

| 24.88 | 24.88 | 25.56 | 24.86 | 25.31 | Buy |
| 25.31 | 25.26 | 26.01 | 25.26 | 25.83 | Buy |
| 25.83 | 26.0 | 26.41 | 26.0 | 26.31 | Buy |
| 26.31 | 26.31 | 26.71 | 26.01 | 26.03 | Buy |
| 26.03 | 26.10 | 26.26 | 25.56 | 25.64 | Sell |

| Previous | Open | Max | Min | Last | Action |
|----------|----------|----------|----------|----------|--------|
| Positive | Positive | Positive | Negative | Negative | Sell |
| Negative | Positive | Positive | Negative | Negative | Buy |
| Negative | Negative | Equal | Negative | Negative | Buy |
| Negative | Negative | Equal | Negative | Negative | Sell |
| Negative | Equal | Positive | Negative | Positive | Buy |
| Positive | Negative | Positive | Negative | Positive | Buy |
| Positive | Positive | Positive | Positive | Positive | Buy |
| Positive | Equal | Positive | Negative | Negative | Buy |
| Negative | Positive | Positive | Negative | Negative | Sell |

The gain Ratio method is used to place the attributes and to build the decision tree where we can place the attributes according to its gain ratio. For the company MECE, Last was the root for the decision tree. For the company ARBK, root attribute for creating decision tree was Open. As the process of building the decision tree goes on, all the remaining attributes were used to continue with process. After building the complete decision tree, the set of classification rules were generated by following all the paths of the tree. The maximum number of attributes that were used in some of the classification rules generated were 4 attributes, while some classification rules used only 1 attribute. Both the ID3 and C4.5 algorithms were used in building the decision trees and the pruning technique was used in the C4.5 algorithm in order to reduce the size of the produced decision trees. Below table gives a summary about the numbers of the classification rules that resulted after building the decision trees for each company using the C4.5 algorithm. The graphs of the resulting decision trees using the C4.5 algorithm with pruning technique is presented in the below figures 1 and figures 2 for the two companies under study.



5383

**Predictions**

+ve      _ve      =equal

Buy(19.0/9.0)

Open

Sell(10.0/3.0)

+ve      _ve      =equal

Sell(9.0/4.0)

Sell(19.0/9.0)

Buy(8.0/2.0)

Open

+ve      _ve      =equal

Sell(188.0/90.0)

Last

Buy(47.0/19.0)

+v      _ve      =equal

Sell(27.0/10.0)

Buy(69.0/32.0)

Sell(4.0)
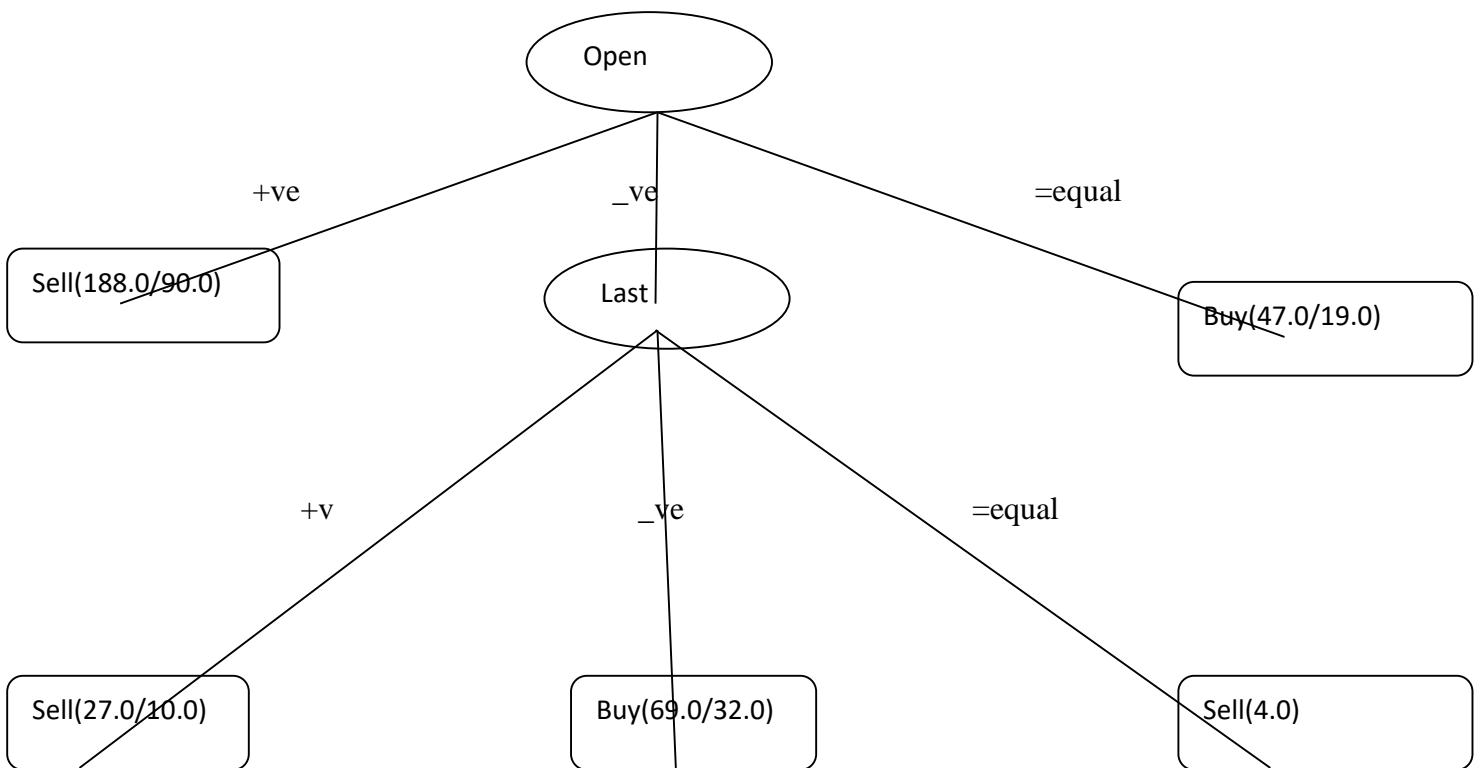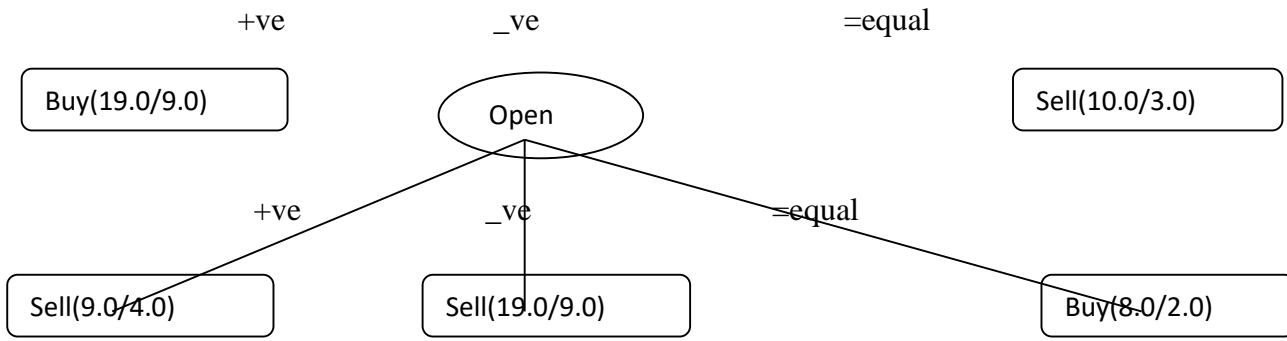
## Deploying the model

The classification rules that were generated from the tree model can be used to predict the best timing for the investors either to buy or sell the stocks.

[1]*V.Rajesh Reddy, [2]Dr A. Gayathri

# 4. RESULT AND ANALYSIS

This study presents a proposal to use the decision tree classifier on the historical prices of the stocks to create decision rules that give buy or sell recommendations in the stock market. Such proposed model can be a helpful tool for the investors to take the right decision regarding their stocks based on the analysis of the historical prices of stocks in order to extract any predictive information from that historical data. The results for the proposed model were not perfect because many factors including but not limited to political events, general economic conditions, and investors' expectations influence stock market.

| Company | Classification Method | 10-CV | | | Hold Out66% | | |
|---------|-----------------------|-------|-------|-------|-------|-------|-------|
| | | Total Instances | Correctly classified | Accuracy % | Total Instances | Correctly classified | Accuracy % |
| MECE | Id3 | 502 | 255 | 50.797 | 171 | 84 | 49.123 |
| | C45 | | 265 | 53.586 | | 91 | 53.216 |
| UAIC | Id3 | 502 | 269 | 53.586 | 171 | 88 | 51.462 |
| | C45 | | 264 | 52.590 | | 94 | 54.971 |

Classification accuracy using ID3 & C4.5 classification methods and using 10-CV & Holdout evaluation methods

# 5. CONCLUSION

In order to evaluate the model, the WEKA software was used to calculate the accuracy of the classification model. Two evaluation methods were used, the K-Fold Cross Validation (K-CV) where K=10 folds and the percentage split method where 66% of the data was used for training and the remainder for testing. Both evaluation methods were used on the ID3 and C4.5 decision tree classification methods. Below table shows the accuracy of all the classifiers generated using both classification methods and both evaluation methods

# REFERENCE

1. Wang, J.L., Chan, S.H. (2006) "Stock market trading rule discovery using two-layer bias decision tree", Expert Systems with Applications, 30(4), pp. 605-611.
2. Lin, C. H. (2004) Profitability of a filter trading rule on the Taiwan stock exchange market. Master thesis, Department of Industrial Engineering and Management, National Chiao Tung University.

3. Cao, Q., Leggio, K.B., and Schniederjans, M.J., (2005) "A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market", Computers & Operations Research, 32, pp. 2499-2512.
4. Fama, E.F., French, K.R., (1993) "Common risk factors in the returns on stocks and bonds", The Journal of Finance, 33, pp. 3-56.
5. Fama, E.F., French, K.R., (1992) "The cross-section of expected stock returns", The Journal of Finance, 47, pp. 427-465.
6. Al-Haddad W. Alzurqan S. and Al_Sufy S, The Effect of Corporate Governance on the Performance of Jordanian Industrial Companies: An empirical study on Amman Stock Exchange. International Journal of Humanities and Social Science, Vol. 1 No. 4; April 2011
7. Hajizadeh E., Ardakani H., and Shahrabi J., Application of data mining techniques in stock markets: A survey, Journal of Economics and International Finance Vol. 2(7), pp. 109-118, July 2010.
8. Soni S., Applications of ANNs in Stock Market Prediction: A Survey, International Journal of Computer Science & Engineering Technology (IJCSET), pp 71-83, Vol. 2 No. 3, 2011.
9. Hazem M. El-Bakry, and Wael A. Awad, Fast Forecasting of Stock Market Prices by using New High Speed Time Delay Neural Networks, International Journal of Computer and Information Engineering 4:2 2010. Pp 138-144.
10. Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., and Wirth R., (2000). "CRISPDM 1.0: Step-by-step data mining guide".