

## Journal of Industrial Strategic Management

### Forecasting of Tehran Stock Exchange Index by Using Data Mining Approach Based on Artificial Intelligence Algorithms

Mohammad Mahmoodi<sup>\*a</sup>, Akbar Ghasemi<sup>b</sup>

*a Assistant professor, Faculty of management, Accounting department, Islamic Azad University, Firoozkough branch, Iran*

*b Msc., Faculty of management, Islamic Azad University, Firoozkough branch, Iran*

---

**CHRONICLE Abstract**

**Article history:**

Received: 04/07/2017

Received in revised:  
08/18/2017

Accepted: 10/11/2017

**Keywords:**

Data mining, Stock  
Price Changes,  
Decision tree,  
artificial intelligence  
algorithm.

Uncertainty in the capital market means the difference between the expected values and the amounts that actually occur. Designing different analytical and forecasting methods in the capital market is also less likely due to the high amount of this and the need to know future prices with greater certainty or uncertainty. In order to capitalize on the capital market, investors have always sought to find the right share for investment and the right price to buy and sell, and so all the predicted models always seek to answer the three basic questions, i.e., which share, to what extent When and at what price to buy or sell. Before answering the answers given to these questions, you have to answer a more serious question. Including whether forecasting financial markets is possible. Accordingly, in this research, using data mining, we proposed a method for predicting changes in the total stock index of Tehran stock exchanges. The purpose of this research is in the field of applied research. In terms of its implementation, the research is based on a causal research that is carried out using a data collection database. Based on the results obtained from this study on the best decision tree algorithm with respect to the accuracy of 94% of the C & R Tree algorithm, it can be said that this algorithm can be better than predicting stock price changes. Also, using decision tree can also predict changes in the price of the payment.

---

© All rights reserved

## 1. Introduction

Uncertainty in the capital market means the difference between the expected values and the amounts that actually occur. Designing different analytical and forecasting methods in the capital market is also less likely due to the high amount of this and the need to know future prices with greater certainty or uncertainty. In order to capitalize on the capital market, investors have always sought to find the right share for investment and the right price to buy and sell, and so all the predicted models always seek to answer the three basic questions, i.e., which share, to what extent When and at what price to buy or sell. Today, financial managers prefer to have a mechanism that can help them in their decision-making, so paying attention to the methods of forecasting has been very much considered.

In the following, it should be noted that in the event of a predictable capital market, it is necessary to examine the various dimensions of the capital market and the methods presented in each field for prediction. In the following, we need to examine what methods are effective for this prediction and whether the combination of these methods is generally possible. In the following, we can consider the tools used to predict all aspects of the capital market in three broad categories: technical methods, fundamental methods, and mathematical methods, including classical time series methods and regression methods and artificial intelligence methods.

Artificial intelligence algorithms, which are being used rapidly by investors, are a combination of all the predictive methods along with the ability to fit high-level nonlinear curves. These algorithms can work with a large number of variables and find the right relationship between these variables. As mentioned, there are many factors affecting the stock price, and classical models and traditional and empirical analyzes do not account for that number. For short-term forecasts, classic time series models and technical analysis are generally used. Each of these includes many other algorithms that make it difficult to summarize and conclude. Artificial intelligence algorithms have the power to combine all of these analytics with optimal weighting and provide a single, optimal solution. Among the artificial intelligence algorithms, the use of neural networks in the prediction field is very high. This is due to the ability of the neural network to work with a large number of variables, a very accurate fit to the time series, not being affected by outlier's data, a limitation to a certain degree of nonlinearity and flexibility of the network against changes in model parameters.

As given the above, stock price index forecasting and direction of movement are considered as one of the most challenging time series applications. However, much empirical research has been done on the issue of stock price forecasting. But most of the empirical gains are related to developed financial markets, and little research has been done on developing financial markets. Given the high analytical power of data mining technology and its unique processing power, it can be used to analyze numerous real-world problems, including forecasting.

In this study, by examining all of the above and the feasibility of integrating the methods used to predict the price, the answers to the questions will be answered; for the first time, the two prices are foreseen for the upcoming periods; the high price and Low stock price. In this way, speculators can use the method to accurately predict the price and profit through volatility. In this research, by determining the effective variables on the Tehran Stock Exchange index and determining the internal and external factors affecting it over a period of several days, these indicators include such as the price of the dollar, the change in the price of the dollar, the price of gold, the average volume of transactions in Rials and the number , The rate of change in the index in a limited time period, etc., which are determined by reviewing the literature of the research, and in the next step, the change in the total stock index is expected in a one-week period. In this regard, various AI algorithms, such as decision tree algorithms such as C & R TREE, QUEST TREE and CHILD TREE, or algorithms based on vector support rules, including backup vector machine, as well as the use of urban algorithms such as grid Baysin. The final objective of this research is to predict the growth rate of the total stock index of Tehran Stock Exchange using data mining. Therefore, the main question of the research is how to predict the growth of the stock index using Artificial Intelligence?

## 2. Literature Review

The literature of research is examined in three sections; capital market efficiency theory, Data mining, decision trees and decision rules.

### 2.1 Capital Market Efficiency Theory

For more than a quarter of a century, the attention of the financial and economic professors of the universities has been taken into account the efficiency of the capital market in various countries. For capitalist countries, the efficiency of the market is of great importance, because if the capital market is efficient, the price of securities is determined correctly and fairly, and the optimal allocation of capital, the most important factor of production and economic development, is optimal is done. In the financial world, there are three levels of market efficiency: 1. Information efficiency. 2. Assignment efficiency. 3. Operational efficiency. (Bisoi , Dash 2014)

Advocates of market efficiency theory argue that investors use any irrational trend in stock prices as long as they exist to earn higher-than-average profits. In some cases, such as the January effect (predictable pattern for price changes), high costs in transactions are mostly more profitable than trying to take advantage of this trend. In the real world, markets cannot be completely efficient or completely non-workable, and it's better to imagine markets as a combination of both, so that daily events and decisions do not always immediately affect stock prices. Therefore, the assumption of the predictability of a hypothetical capital market is unreasonable, and it is possible to investigate and expand the methods of forecasting. (Bisoi , Dash 2014)

### 2-2 Data Mining

Data mining is known as the most important use of data in data warehouses. In fact, data mining analyzes existing data to identify possible trends, unrecognizable connections of hidden pattern from massive data. In fact, the purpose of data mining is to create models for decision making. These models predict future behaviors based on past analyzes. In this process, complex mathematical and statistical algorithms are used to transform data into organizational knowledge. (Sinaei et al 2005)

The term data mining is synonymous with one of the expressions of knowledge extraction, information harvesting, data mining, and even data mining, which in fact describes the discovery of knowledge in databases. The discovery of knowledge in databases is aimed at discovering useful information from a large collection of data. Knowledge discovered can be the basis for describing the characteristics of data, patterns that occur appropriately, the clustering of topics within databases, and so on. (Meshkani , Nazemi 2009)

### 3-2 Decision trees and decision rules

Decision trees and decision rules are data mining methodologies that are used in many real-world applications as a powerful solution for categorizing issues. So, first, we briefly describe the basic principles of classification. In general, the classification is a process of learning performance that maps a pen to one of several preset groups. Each classification is based on inductive learning algorithms as a set of examples that include vectors, attribute values (also called feature vectors), and a class and proportional class. The purpose of learning is to create a classifier model known as classifier, which predicts a class for some entities (a given instance) with values of its existing input traits. In other words, the process classification assigns a discrete label value (class or class) to an unlabeled record and in fact predicts a classifier. For example, a simple classification can be done for billing customers in two special categories: those who pay their bills within 30 days and those who pay their bills for more than 30 days. Different classification methods today use some sort of methodology to make a classification, due to the large amount of data that requires process automation. Examples of the classification method used as part of data mining applications include the classification of methods in the financial market and the selection of objects in large image databases. (Alizadeh et al 2008)

#### 4-2. Research background

Fadaei Nejad , Esmail (1995) conducted a study using self-correlation and exodus tests using 50 companies' weekly prices for the period of 1368-1378 and evaluated the efficiency of the Tehran Stock Exchange market at a weak level.

Allah Yari (2008) in a research using the daily price of stock of 95 bourse companies during the period from 1384 to 1387 examined the efficiency of the stock market by statistical methods of correlation analysis and RUN test. The results of the research show that stock price change is not random and has the trend is foreseeable. Therefore, the Tehran Stock Exchange is inefficient

Moshiri and Maravot (2006) predicted the total stock return index with linear and nonlinear models. Using daily and weekly index data from 1377 to 1382 and using different forecasting methods such as GARCH and ARIMA models and networks Neural, predicted the total index. The results indicate that the neural network model has a lower error than the other two models. However, in the meaningful test, it was found that these differences are not meaningful. In other words, the accuracy of the prediction models is not statistically significant

Also, Adel Azar (2006) conducted a prediction of the stock index with three approaches to classical approaches, artificial intelligence approach and a hybrid approach. The results of this research indicate that fuzzy neural networks have superiority over the ARIMA method and have the features of fast and timely uniqueness exceeding the stock price index is appropriate

Charkha (2008) proved that neural networks perform better than predictive value in statistical methods. They designed a prototype based method to forecast daily stock prices and compared the results from neural networks and statistical methods. They have proven that if the neural network is properly trained, properly designed, have the right inputs and outputs, they can well predict the price. In addition to the superiority of the technique introduced by the authors, the model became more complex than statistical methods, and so neural networks can be used as an alternative to predicting stock prices on a daily basis.

Zhi rang et al. (2005). tried to predict the stock market by aggregating neural network capabilities. Researchers are predicting a stock price index based on the ability of the neural networks to admit that the prediction of the global labor cost index is difficult. By examining the classical and modern time series, they concluded that prediction with these models has many challenges and that neural networks are more suitable for this. Neural Networks have the ability to extract useful data from a large amount of data. Researchers in this study, by reviewing the literature on the application of neural networks in predicting stock prices, concluded that neural networks are very useful for predicting stock prices.

Liao et al. (2011) demonstrated, through practical results, that the analysis of integral independent components (ica) based on the filtering map of the existing disturbances among the data along with the neural networks to predict the price, performed better than filtering based on Elliott waves Along with neural networks, nerve networks alone and random stroll patterns. Based on the results, the author concludes that the proposed method can be effective in identifying and eliminating turmoil in stock prices and improving the efficiency of neural networks.

Jakob Kara et al. (2011) investigated the prediction of moving the stock price index of the Istanbul Stock Exchange with neural-fuzzy and SVM models, and the daily data from 1997 to 2007 along with 10 technical indicators as input variables were used. . Neural network-fuzzy 75.74% and support model 52.81% predicted and better performance of neural-fuzzy network compared to the supported car model. It was also the best performance prediction of 2001 .

Samidha D Sharma and Abhishek Gupta,(2014) issued for investigating in academic and financial research. There are various techniques available for the prediction of the stock market value. Here in this paper a survey of all the techniques and schemes for stock market prediction are discussed and analysed.

Vivek Rajput, Sarika Bobde (2016) tried to predict stock price movement using the sentiment analysis from social media, data mining. In this paper they find efficient method which can predict stock movement more accurately. This paper contributes to the field of sentiment analysis, which aims to extract emotions and opinions from text. they examine sentiment expression and polarity

classification within and across various social media streams by building topical datasets within each stream. Different data mining methods are used to predict market more efficiently along with various hybrid approaches. They conclude that stock prediction is very complex task and various factors should be considered for forecasting the market more accurately and efficiently.

Surbhi Sharma and Baijnath Kaushik (2018) investigated Quantitative Analysis of Stock Market Prediction for Accurate Investment Decisions in Future, The objective of this review is to predict the stock market prices in order to make more informed and accurate investment decisions. Various approaches and the results of past years are compared based on methodologies, datasets and efficiency and then it is represented in the form of a Graph. The survey describes different theories and conventional approaches to stock market prediction. Along with it, it discusses recent machine learning techniques along with pros and cons of each technique for effectively predicting the future stock prices followed by various researchers.

Lin Chen and etal. (2018), used high-frequency data to examine the predictive performance of deep learning, and compared three traditional artificial neural networks transaction data of the CSI300 future contract (IF1704) in our empirical analysis, and test three groups of different volume samples. found that the deep learning method of predicting stock index futures out performs the back propagation, the extreme learning machine, and the radial basis function neural network in its fitting degree and directional predictive accuracy.

Malav Shastri, Sudipta Roy and Mamta Mittal(2019), used for pre-processing part and a neural network model with inputs from sentiment analysis and historic data is used to predict the prices. It has been observed from the experiments that the accuracy level reaches above 90% in maximum cases, as well as it also provides the solid base that model will be more accurate if it trained with recent data. The intended combination of sentiment analysis and Neural networks is used to establish a statistical relationship between historic numerical data records of a particular stock and other sentimental factors which can affects the stock prices.

Xiao Zhong and David Enke(2019), presented a comprehensive big data analytics process to predict the daily return direction of the SPDR S&P 500 ETF (ticker symbol: SPY) based on 60 financial and economic features. DNNs and traditional artificial neural networks (ANNs) are then deployed over the entire preprocessed but untransformed dataset, along with two datasets transformed via principal component analysis (PCA), to predict the daily direction of future stock market index returns. While controlling for overfitting, a pattern for the classification accuracy of the DNNs is detected and demonstrated as the number of the hidden layers increases gradually from 12 to 1000. Moreover, a set of hypothesis testing procedures are implemented on the classification, and the simulation results show that the DNNs using two PCA-represented datasets give significantly higher classification accuracy than those using the entire untransformed dataset, as well as several other hybrid machine learning algorithms. In addition, the trading strategies guided by the DNN classification process based on PCA-represented data perform slightly better than the others tested, including in a comparison against two standard benchmarks.

### 3- Methodology

This research is applied in terms of purpose because a set of rules is a valid, reliable and systematic tool for investigating facts, identifying uncertainties, and finding solutions to problems. Since this research is predictive, the research area is well known, and as a field tool for data collection is a survey research.

The time period of research is from July 2018 to December 2018 and the area of this research is Tehran Stock Exchange.

In this research, the statistical sample is the fluctuations and price changes of the total stock index of Tehran Stock Exchange. Also, the sampling method is not used in this paper and all data have been analyzed.

The method of data gathering in this research was in two fields, firstly, using archival studies, the necessary information was gathered about the research, which was based on which data mining models were studied. Field methods are also used to collect the necessary data for different stages of data mining.

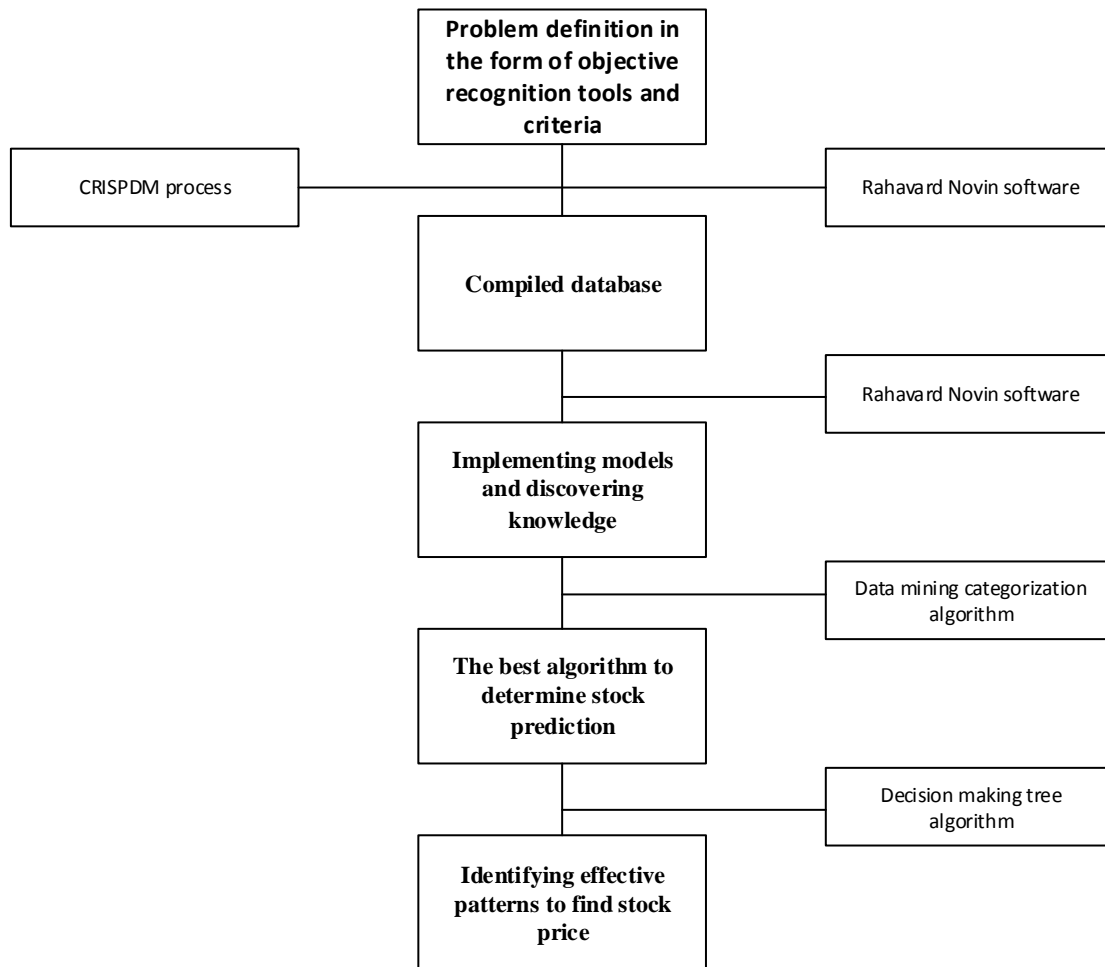


Fig 1 Information Analysis Method

The criteria for predicting stock changes in Table 1 along with the corresponding symbol is shown

Table 1: Criteria with their symbol

Symbol	Criteria
A	No. companies that have been faced with an increase in stock
B	No. Ask queue
C	Changing the exchange rate in comparison with the last 3 days
D	Average increase or decrease in 5 last working days
E	Volume of Trades (millions)
F	No. of Trades
G	Change gold rate
H	Change rate

4. Results

The variables studied includes; The number of companies that face the increase in stock value; Number of shopping queues; Exchange rate change compared to last three days; Average increase or decrease in 5 working days of analysis; Total share transaction (million shares); Number of transactions and Change the gold price as in puts and The amount of changes in the total index in quota for three working days as out puts.

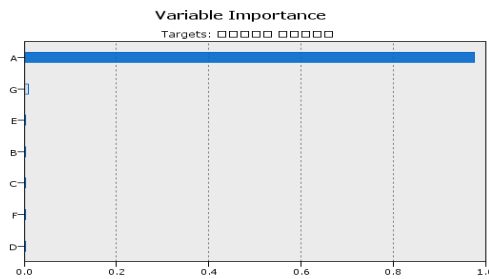


Fig 2 Characterization Ranking Using C & R Tree Algorithm

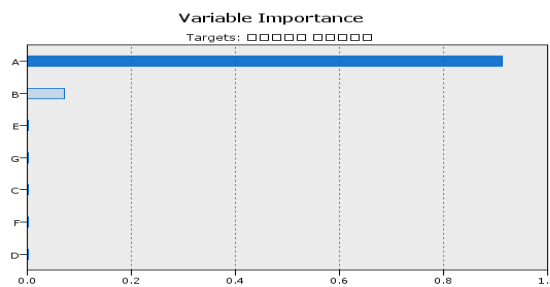


Fig 3: Ranking of attributes using the Quest Tree algorithm

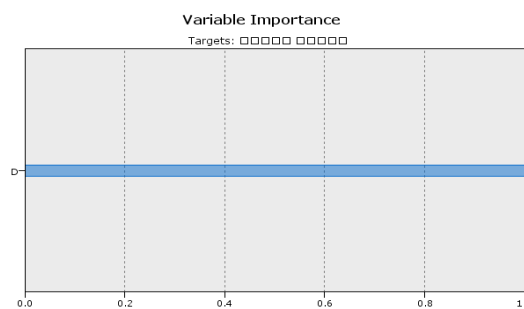


Fig 4 Feature Ratings Using Chaid Tree Algorithm

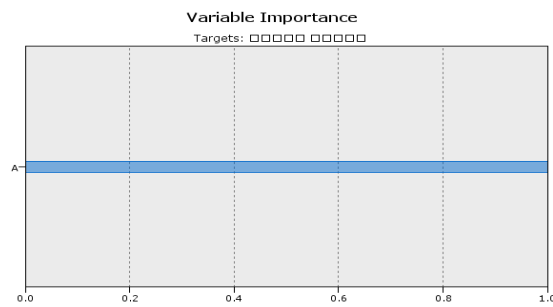


Fig 5 Feature Ratings Using C5 Algorithm

Table 2 the most crucial effects on total indicator growth

Decision trees Algorithm	Factor 1	Factor 2
C&R Tree	The number of companies that have been faced with an increase in stock	Change gold rate
Quest Tree	The number of companies that have been faced with an increase in stock	Ask queue/ bid queue (because of daily price limits)
Chaid Tree	Analysis of average increase or decrease in 5 working days	
C5	The number of companies that have been faced with an increase in stock	

Based on the results obtained from the decision tree algorithms shown in Table 3, the C & R Tree algorithm can provide a good predictor for stock growth.



Table 3 compares the accuracy of decision tree algorithms

<b>Decision trees Algorithm</b>	<b>The overall accuracy of the algorithm (Percentage)</b>	<b>The number of wrong predictiona</b>	<b>The number of right predictiona</b>
C&R Tree	94	24	376
Quest Tree	92	32	368
Chaid Tree	92	32	368
C5	92	32	368

If the decisions of trees are examined in the form of trees, then the rules that lead to the change in the price of the total stock index can be determined.

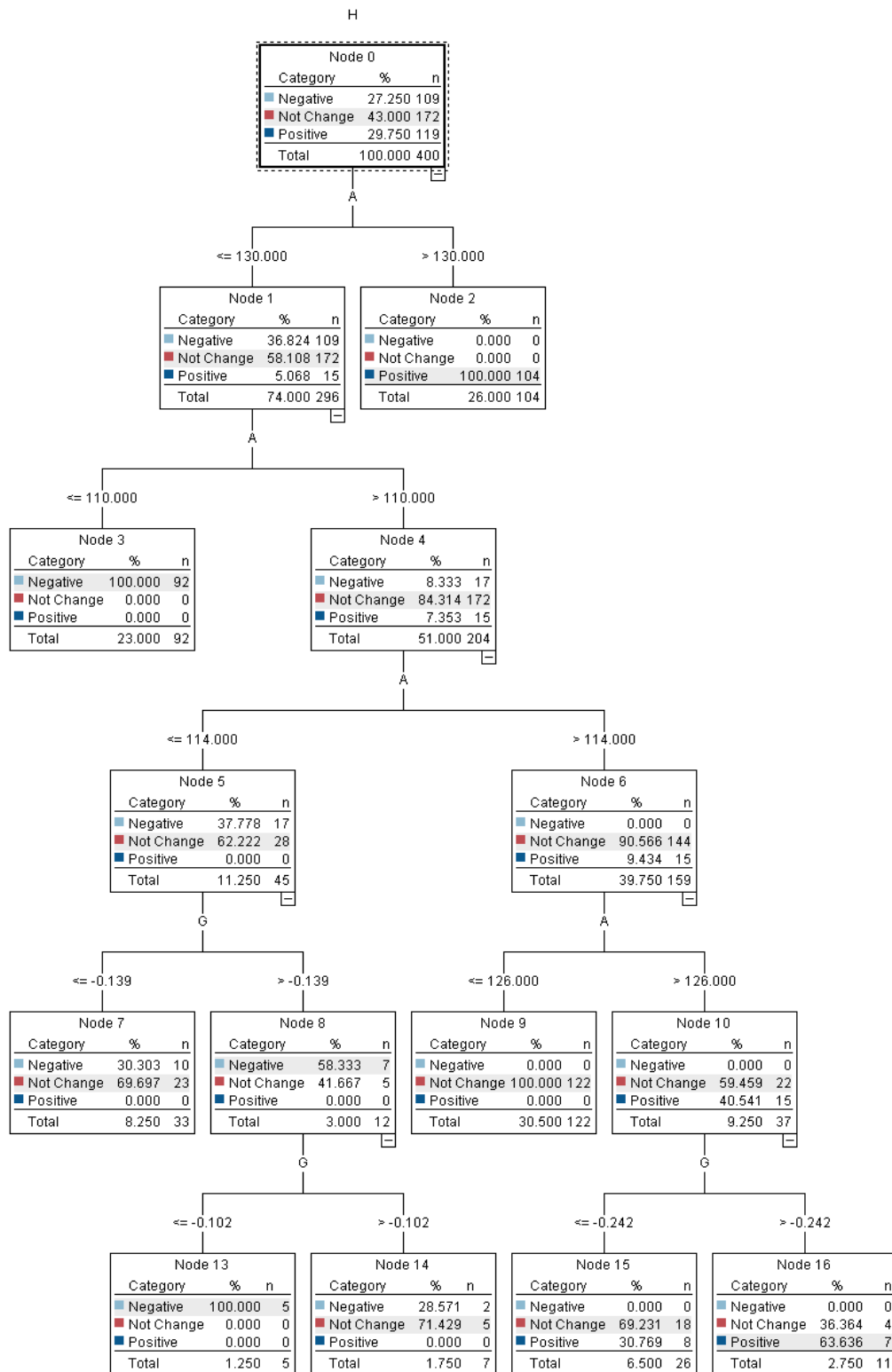


Fig 6 total tree C&R Tree

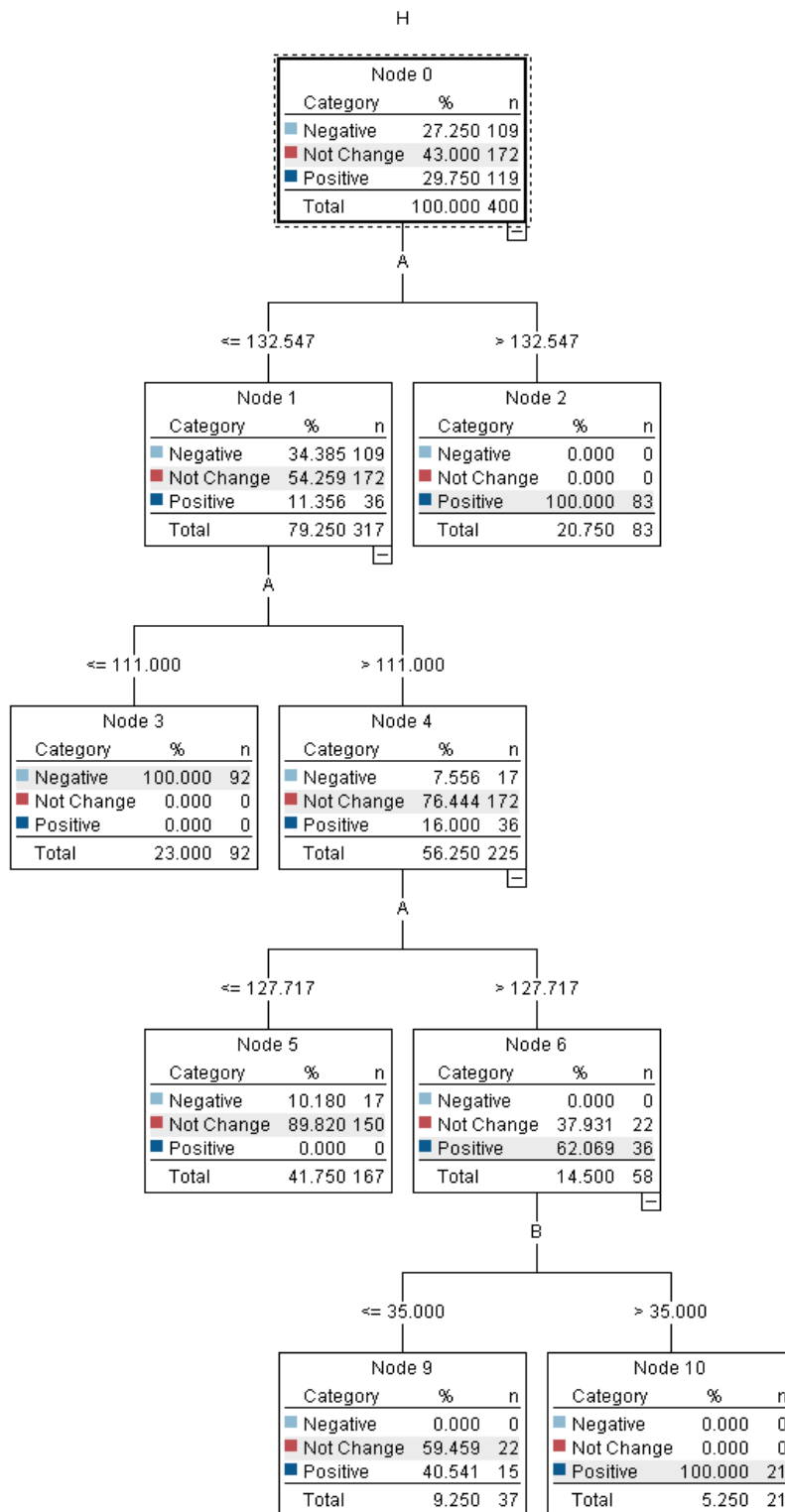


Fig 7 total tree algorithm Quest Tree

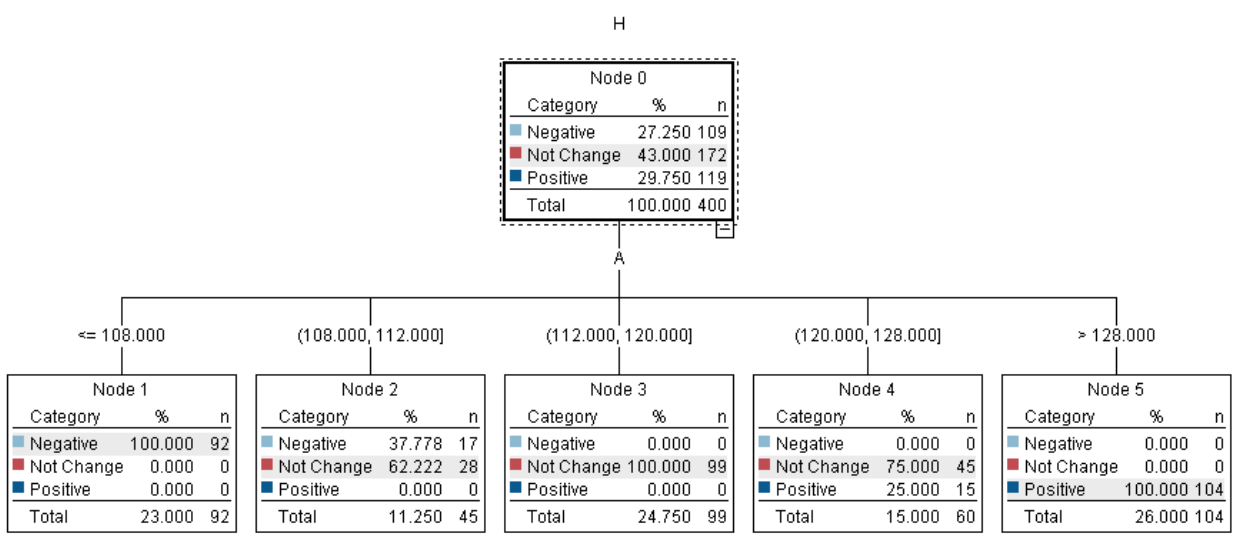


Fig 8 total tree algorithm Chaid Tree

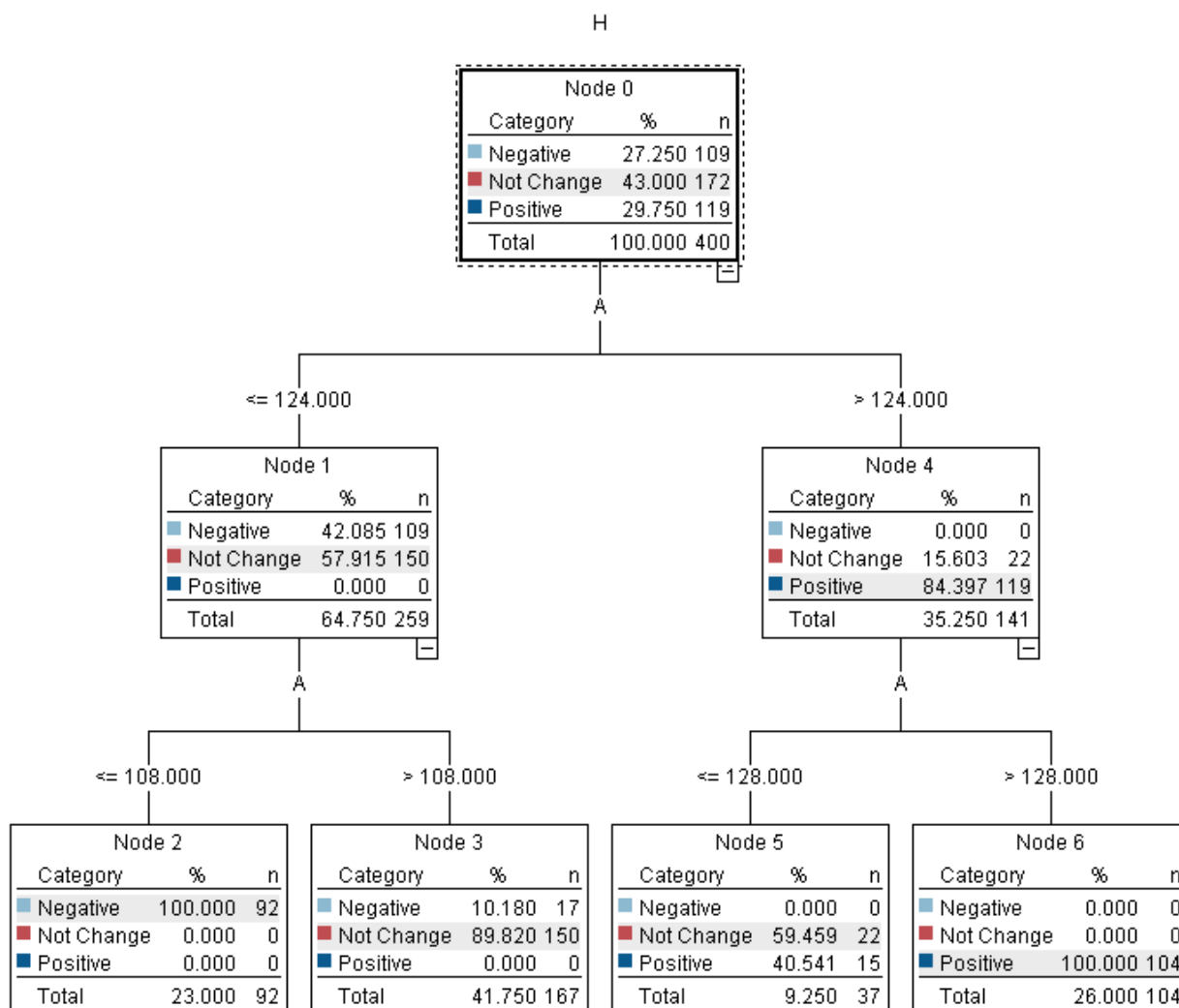


Fig 9 total tree algorithm C5

### 5. Discussion and Conclusion

In this study, stock market basics are discussed and then the need for predicting the future stock market prices. Few of the approaches which may be used for stock market predict are elaborated. also, In order to enhance the capitalized value on the capital market, investors have always sought to find the right share to invest and the right price to do a business deal, and so all the predicted models always try to answer the three basic questions, i.e., which share, to what extent When and at How long to buy or sell. Before scrutiny to the answer of these questions, crucially you should answer to more serious issues. Including whether forecasting financial markets is possible. In the following, it should be noted that in the event of a predictable capital market, it is necessary to examine the various dimensions of the capital market and the methods presented in each field for prediction.

Although there are various techniques implemented for the prediction of stock market. Here in this paper we survey of the stock market prediction techniques based on artificial intelligence. On the

basis of existing artificial techniques used for the future prediction of stock market C & R Tree, Quest Treet, Chaid Tree, and C5 decision tree algorithm techniques for the prediction is proposed which provides close prediction of stock market.

Our prediction models expands our ability to analyze financial market behavior. Because there are complex relationships among stock futures prices and such factors as the economy, politics, the environment, and culture, future research could apply complex theory to key input variables that influence stock prices and returns. That would allow the construction of that would facilitate better predictive performance.

The main objective of this research was to develop and expand the techniques used in financial modeling with the aid of artificial intelligence (AI) approaches, and the approach taken for this research was to predict the future price of the stock market. The vast amount of work done in this area with this techniques has focused more on predicting the direction of the future price with accurate prediction result to optimize forecasting behavior in the market. Most of the authors have used methodologies in artificial intelligence to achieve accuracy and performance But still there is a need to improve the parameters accuracy and performance. The results of the proposed models are compatible with researches that state that there is a strong relation between stock news and changes in stock prices.

In this research, using data mining, we proposed a method for predicting changes in the total index of Tehran Stock Exchange shares. The methodology used was to initially predict the primary measures that were effective on profits: the number of companies faced with the increase in stock value, the number of queues, the exchange rate change compared to the previous three days, the average increase or decrease In 5 working days of analysis, the total transaction volume (million shares), the number of transactions and the change in gold rates were identified, and in the next step, the most important criteria and rules were obtained using four C & R Tree, Quest Treet, Chaid Tree, and C5 decision tree algorithms. It has been reported that, based on the results obtained, the four criteria of the number of companies faced with the increase in stock values are male Gold, the number of sales queues, and the average increase or decrease in 5 business days of analysis are identified as important criteria by algorithms. Regarding the best decision tree algorithm, considering the accuracy of 94% of the C & R Tree algorithm, it can be said that this algorithm can better predict stock price changes. Also, using decision tree can also be used to predict changes in the price of payment, which can be seen in figers 6 to 9.

In this regard, it is suggested to other researchers that:

- Combine variables related to fundamental analysis in the method used in this research as inputs of decision tree models.
- Time series data mining at the entire market level to enhance the accuracy of identifying the most similar time series to the target time series.
- Combine the neural network and the decision tree, in such a way that the two methods are combined and compared.

## 9. Resources

- 1) Adel, Azar, Afsar, Amir (2006), Comparison of Classical and Artificial Intelligence Methods in Stock Expected Stock Indicators and Designing a Combined Model, Quarterly Journal of Humanities, No. 4.
- 2) Alhayari, Ibrahim (2008) Investigating the weak form of capital market efficiency in Tehran Stock Exchange. Stock Exchange Quarterly. Number 4
- 3) Alizadeh S., Timurpour B, Ghazanfari M.,(2008) , Data mining and knowledge discovery, Iran University of Science and Technology.
- 5) Bisoi, R., Dash, P. K., (2014). A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter. Journal of Applied Soft Computing, 3(19), pp. 41-56

- 6) Charkha, P. R., (2008) . Stock Price Prediction and Trend Prediction Using Neural Networks. Journal of First International Conference on Emerging Trends in Engineering and Technology, 3(7), pp. 592-594.
- 7) Fedaye Nejad, Mohammad Esmail (1995), Reviewing the Effectiveness of Tehran Stock Exchange, PhD dissertation, Tehran University, Faculty of Management.
- 8) Kara, Y., Boyacioglu, M. A., Baykan, O. K., (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Journal of Expert Systems with Applications, 38(5), pp. 5311-5319.
- 9) Lien Chen, Zhilin Qiao , Minggang wang, chao wang, ruijin du, and harry eugen stanely (2018), Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market, special section on big data learning and discovery, Digital Object Identifier 10.1109/access.2018.2859809.
- 11) Liao, SH. L., Chu, P. H., You, Y. L.,(2011). Mining the co-movement between foreign exchange rates and category stock indexes in the Taiwan financial capital market. Expert Systems with Applications, 38(4), pp. 324-331.
- 12) Malav Shastri<sup>1</sup>, Sudipta Roy<sup>2</sup> and Mamta Mittal(2019), Stock Price Prediction using Artificial Neural Model: An Application of Big Data, EAI Endorsed Transactions on Scalable Information Systems.
- 13) Meshkani, A. and A.Nazemi. 2009. Introduction to Data Mining, Ferdowsi University of Mashhad Press, Mashhad.
- 14) Moshiri, Saeed, Moravot, Habib (2006), Estimation of total stock returns of Tehran stock exchanges using linear and nonlinear models, Journal of Commercial Law Research, No. 41.
- 15) Samidha ,Sharma and Abhishek Gupta,(2014) A Survey on Stock Market Prediction Using Various Algorithms Abhishek Gupta et al, Int.J.Computer Technology & Applications,Vol 5 (2),530-533.
- 16) Seyed Saman Emami (2018), Predicting Trend of Stock Prices by Developing Data Mining Techniques with the Aim of Gaining Profit, Journal of Accounting & Marketing.
- 17) Sinaei, Hassanali, Mortazavi, Saeedeh ... and Teimouri Asl, Yaser (2005), Tehran Stock Exchange Stock Exchange (TSE) forecasting using Artificial Neural Networks, Accounting and Auditing, Vol. 12, No. 41, pp. 83-59.
- 19) Surbhi Sharma and Baijnath Kaushik (2018), Quantitative Analysis of Stock Market Prediction for Accurate Investment Decisions in Future, Journal of Artificial Intelligence.
- 20) Vivek Rajput<sup>1</sup> , Sarika Bobde (2016), Stock market forecasting techniques: literature survey, International Journal of Computer Science and Mobile Computing.
- 21) Xiao Zhong<sup>1</sup> and David Enke(2019), Predicting the daily return direction of the stock market using hybrid machine learning algorithms, Zhong and Enke Financial Innovation.
- 22) Zhi rang, Zhang, Z.Y., et al. (2005). Stock time series forecasting using support vector machines employeing analyst recommendations, Springer-Verlag Berlin Heidelberg.

