JOURNAL OF CURRENT RESEARCH IN SCIENCE (ISSN 2322-5009)
CODEN (USA): JCRSDJ
2015, Vol. 3, No. 2, pp: 150-161
Available at www.jcrs010.com

ORIGINAL ARTICLE

Investigation the Effectiveness of Artificial Intelligence on the Forecasting of Tehran Stock Exchange Index

Setila Rostami[1*], Marziyeh Bayat[2], Darush Javid[3], Mansur Esmaeilpuor[4]

1 . Master Of Accounting, Hamedan Branch, Islamic Azad University, Hamedan, Iran.
2 . Accounting Dept, Hamedan Branch, Islamic Azad University, Hamedan, Iran.
3 . Accounting Dept, Hamedan Branch, Islamic Azad University, Hamedan, Iran.
4. Computer Engineering Dept, Hamedan Branch, Islamic Azad University, Hamedan, Iran.

Corresponding Author email: setilarostami@yahoo.com

ABSTRACT: In this study, we investigated the effectiveness of artificial intelligence on the forecasting of Tehran stock exchange index. Since the forecast is mainly used to reduce risk and increase profits, the accuracy in prediction is a very important issue. Therefore, the aim of the present study is to evaluate the efficiency of artificial intelligence algorithms in the forecasting of Tehran's Stock Exchange index. In this study, we used gather data from Tehran's Stock Exchange organization through the RAHAVARD NOVIN software. The population and statistical samples of the study, is the Tehran Stock Exchange organization during 2007 to 2015. We forecasted Tehran Stock Exchange index using three models of decision-tree and Rough Set and Logistic Regression that are the subset are artificial intelligence algorithm, and using ROSSETA and WEKA software. Subsequently, we compared the values predicted by these models with the actual values using the paired sample t-test in SPSS software; and finally we compared the superior performance of models using the ANOVA test. Since the hypothesis is confirmed in the study, the artificial intelligence algorithm (for example: this tree models) can be used as a reliable method to predict the Tehran Stock Exchange index.

Key words: Forecast, index, Artificial Intelligence, Logistic Regression, Decision Tree, Rough set

INTRODUCTION

As we know, capital and labor force are the main pillars of production. The supply of these factors and their optimal diagnosis is essential for economic growth. This allocation requires the presence of markets and the optimal performance of market forces. With regards to the capital, the stock market can do this important task. The most important task of the stock market is to attract outspread capitals and direct them towards investment activities through an optimal allocation process. Investors, with the motivation to receive income, enter the field of investment from two channels, from the profits of the company whose shares they have purchased, and also from selling these shares again. The fluctuation of shares in all stock markets is a natural and normal issue; however, with a prediction of the price and index of stocks, a desirable combination of them can be chosen, and fluctuations can be reduced, and in this way the information individuals have can be increased. It seems that the increase in the information in the market will lead to its better performance. The prediction of what might happen in the future and planning on that basis are very important. It is clear that the characteristic of uncertainty is an undesirable issue; however, this characteristic in unavoidable for investors who have selected the stock market as a place to invest. Therefore, normally all efforts from the investor is to reduce uncertainty, and making predictions in the Stock market is a tool to reduce uncertainty. Forecasting of important Stock market indices can be helpful in increasing the information and making it transparent. Forecasts of the stock market or the capital market indices have always been the center of investigations. This great attention in recent years caused the development of models used in forecasting.

So far, in previous research studies, in order to predict the indices different models have been used such as the Autoregressive integrated moving average (ARIMA) model, autoregressive conditionally model (ARCH), and the Artificial Neural Network (ANN). The Artificial Neural Network model (data mining model) compared to regression methods such as ARIMA and ARCH has shown better performance. Therefore, the present study intends to gain a more accurate conclusion by machine learning algorithm using algorithmic models of Rough Set, Decision Tree (decision-Tree) and Logistic Regression to estimate the future stock index. This article starts with a summarization of the theoretical foundations of stocks index, forecasts, and the models of decision tree, Rough Set and logistic regression. Then background literature will be reviewed, subsequently, the main discussion of the paper, i.e. the design of the model and their comparison with each other will be presented. The final section presents the findings and recommendations of the research.

Theoretical Foundations

Forecast

In a general definition, the prediction of the future conditions and events is called forecast and how this is done, is called forecasting. (AFSAR, 2006)

Index

Index, in general, means figure, representative or indicator. In terms of applications, the word Index (INDEX) is a quantity that represents several homogeneous variables. Index is a tool for measuring and comparing the phenomena that have certain nature and properties on the basis of which the changes in certain variables can be investigated during one period. (PARSE KHEBRE Brokerage Company).

Machine learning algorithm

It means the design and development of algorithms on the basis of which computers or other machines gain learning ability. Its purpose is to achieve machines that are able to extract knowledge (learning) from the environment.

Stock Exchange

The formal and structured capital market where buying and selling of shares of companies and governmental or private institutions' Stock Exchange are done under the terms, rules and specific regulations (LUNNI, 2008).

Review of Literature

Local Studies

Monajemi et al., (2010), in their study entitled "The prediction of stock market prices in Stock Exchange using the Nerve -Fuzzy network using genetic algorithms, and its comparison with artificial neural network" showed that in terms of performance evaluation criteria, the prediction of the stock price of the next day through the hybrid model of Nerve - fuzzy network and genetic algorithm is more accurate than neural network. In other words, prediction of the stock price using Nerve- fuzzy network and genetic algorithm reduces the stock price estimation error in relation to the neural network technique.

Moshiri and Morovat (2007) forecasted the total index of stock output by linear and nonlinear models. Using the daily and weekly data of indices in the period 1998 to 2004, and different forecast methods such GARCH, ARIMA, and neural network models, they predicted the total index. The result suggested that the neural network model had fewer errors than the other two models. However, the statistical test of significance showed that the difference is not significant. In other words, the accuracy of prediction models is not statistically different.

Adel Azar et al. (2007) in a study predicted the stock index by three approaches of classical methods, artificial intelligence approach and hybrid approach. The findings of this research suggest that the Nerve-fuzzy networks are superior over ARIMA method, and have the unique features of quick convergence and high accuracy and are appropriate for the prediction of stock price.

Sinai, Mortazavi, and Teimoori Asldar (2006) predicted the stock price index at Tehran Stock Exchange by artificial neural network, and presented some evidence on the chaotic behavior of stock prices on the Stock Exchange. They selected two sets of data as the input for the neural network, and selected several interruptions of Index and macroeconomic factors as the independent variables. In this study, the linear ARIMA model was used to predict the price index in the next weeks. Results from the study show that the neural network outperformed the linear ARIMA model to predict the price index.

Abbaspoor (2003) conducted a study to predict the stock price "IRAN KHODRO" company in the Tehran stock market using artificial neural network and used the daily data between 2000 to 2001 . Based on the findings of the research, the variables affecting the stock price of " IRAN KHODRO " company include currency exchange rate, oil prices, the P / E ratio (price to earnings) and the volume of the stock exchanges. The results of the study show the superiority of the results of the price forecast by the artificial neural network compared to the Box - Jenkins.

Foreign Studies

In another study, Yakup Kara et al. (2011) attempted to predict the direction of stock price index in Istanbul through neural network models and the Support Vector Machine (SVM), and used the daily data from 1997 to 2007, along with 10 technical indices as input variables of the model. Nerve- Fuzzy Network managed to forecast 75.74%; and the Support Vector Machine (SVM) model 71.52%, and the better performance of Nerve - Fuzzy Network in comparison with the Support Vector Machine model was confirmed. Further, the best predictive performance is related to 2001.

Ming-Chi Lee (2009), predicted the NASDAQ index with a hybrid model of Support Vector Regression (SVR) and compared it with the neural network. In this research, the Support Vector Regression (SVR) was combined with the function of F-score and Supported Sequential Forward Search (FSSFS) and was used by 29 technical indices as a set of complete features to change the index . The data of research was gathered from 2001 to 2007, 80% of which was used for the teaching of the model and 20% for the testing. The results showed the superiority of the hybrid model of Support Vector Regression (SVR) compared to the neural network.

Kelly Logan (2007), forecasted the amount of money in the economy of America by the Least-angle Regression (LARS) and Bayesian methods. She used the variables such as long-term interest rates, short-term interest rates, unemployment rates, the deposit amount, and costs for monetary services between 1960 and 2009 on a monthly.

M.Tsang et al. (2007) investigated the effectiveness of neural network model (NN) on the prediction of Stocks prices in Hong Kong. This system was applied on the events of two Banking Joint Stock Companies in Hong Kong and Shanghai. The system indicated an overall success rate of over 70 percent. This study suggested the superiority of the forecast based on the Least-angle Regression (LARS) model.

zhi yank zhank (2006) attempted to predict the trend of the stock price in Shanghai using Support Vector Machine (SVM). He extracted the daily index price in Shanghai stock market from 2003 to 2005. Further, the recommendations of the nearly 400 capital market analysts and their prediction were used as input variables. The results of his study indicated that the Support Vector Machine (SVM) has a high predictability; and a combination of the Support Vector Machine (SVM) with smart models has even better results compared to the Support Vector Machine (SVM).

METHOD DATA ANALYSIS

Machine Learning

Machine Learning is one of the most important branches of artificial intelligence research which is currently going through a period of growth and evolution. It means the design and development of algorithms based on which computers or other machines gain the ability of learning. Its purpose is to achieve machines that are able to extract knowledge (learning) from the environment. Learning machines have been used to accelerate and automate this process. Research in the field of machine learning is focused on the production of systems capable of extracting the concepts and their relationships in an environment based on observing some examples of them (at least one sample per meaning), and use this knowledge to identify other phenomena in the future. These machines, in general, use the induction technique for their learning. Obtaining knowledge is one of the most important applications of the learning machine in the sense that the act of learning extracts basic information from the environment and uses it for the analysis of future events. Also, another application of the learning machine is to extract large amounts of data. Learning machines are used in intelligent systems to increase knowledge and change it, increase efficiency and automatic error correction.

Decision Tree

Decision trees are one of the most powerful, well-known, and common tools used for classification and prediction, which is a subset of machine learning algorithms. Decision tree is a data structure that can be used to split a large collection of records to smaller sets of records. Decision- tree use a series of questions and very simple decision rules to do this. With each successful division, the elements that are in each set are more similar to each other. As an overview of a decision tree, it can be considered a hierarchical structure in which the intermediate nodes are used to test a feature. The branches are indicators of test output; the leaves indicate class tag or the distribution of class tags. The number of sub-trees of a node determines its grade. Leaf node are rated as zero. In the decision trees each node of the tree does the act of categorization based on the values of one the features, and the final decisions are made in the leaves.

Rough set Theory

Finding an equivalent term in Persian for the term ROUGH SETS is difficult. In the dictionary, ROUGH equivalents are coarse, rude, approximate, turbulent and uneven, among which the word ‘approximate’ is more like the concept of the founder of the theory. But none of these words have the exact meaning of the Latin word; hence, in this study the term "Rough Set" is used. Rough Set Theory is founded in early 1980 by professor PAWLAK. This theory deals with the analysis of data tables. In this theory the data tables can be obtained by measurement, or expert and specialists.

The main aim of Rough sets is to obtain approximate concepts of the acquired data. This theory is a powerful mathematical tool for reasoning in cases of ambiguity and uncertainty which can provide a method for eliminating and reducing irrelevant knowledge that is more than the needs of databases. This process is done by eliminating redundant data on the basis of education (main task of the system) without loss of essential data of the database. As a result of data reduction, a set of abridged and meaningful rules would result that makes the decision-making process much easier. In fact, we can say that the Rough Sets model, by reducing the data space and selecting important terms, perform a shift from a space of the raw data and terminologies to asemantics space (meaning). Thus, due to the explosive growth of data volumes, the Rough sets can be very effective in decision support systems. Rough set theory has many similarities with fuzzy set theory, intuition theory, Boolean logic methods and analysis; however, the rough set theory is considered an independent theory. Rough set theory is a smart mathematical tool that deals with the collections and the relationships between them. Rough theory is built on the basis of the information is of concern to any member of the international community. This method attempts to suggest a way to convert data into knowledge and it is a useful method to discover hidden patterns of data. The main advantage of Rough set theory is that it doesn't need the additional information of the data such as probability in statistics and membership grid in the fuzzy theory.

Logistic Regression

Logistic Regression is a special type of regression where binary response variable is related to a set of explanatory variables, which can be discrete and/or continuous. Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).The important point here to note is that in linear regression, the expected values of the response variable are modeled based on combination of values taken by the predictors. In logistic regression Probability or Odds of the response taking a particular value is modeled based on combination of values taken by the predictors. Like regression (and unlike log-linear models that we will see later), we make an explicit distinction between a response variable and one or more predictor (explanatory) variables. We begin with two-way tables, then progress to three-way tables, where all explanatory variables are categorical. Then we introduce binary logistic regression with continuous predictors as well. In the last part we will focus on more model diagnostics and model selection.

Command: Statistics
- Regression
  - Logistic regression

Logistic regression is applicable, for example, if

we want to model the probabilities of a response variable as a function of some explanatory variables, e.g. "success" of admission as a function of gender.

we want to perform descriptive discriminate analyses such as describing the differences between individuals in separate groups as a function of explanatory variables, e.g. student admitted and rejected as a function of gender

we want to predict probabilities that individuals fall into two categories of the binary response as a function of some explanatory variables, e.g. what is the probability that a student is admitted given she is a female

we want to classify individuals into two categories based on explanatory variables, e.g. classify new students into "admitted" or "rejected" group depending on their gender.

Sample size calculation for Logistic regression is a complex problem, but based on the work of Peduzzi et al. (1996) the following guideline for a minimum number of cases to include in your study can be suggested. Use simple Logistic regression when you have one nominal variable and one measurement variable, and you want to know whether variation in the measurement variable causes variation in the nominal variable.

Simple Logistic regression is analogous to linear regression, except that the dependent variable is nominal, not a measurement. One goal is to see whether the probability of getting a particular value of the nominal variable is associated with the measurement variable; the other goal is to predict the probability of getting a particular value of the nominal variable, given the measurement variable.

You can also analyze data with one nominal and one measurement variable using a one-way ANOVA or a Student's *T_TEST*, and the distinction can be subtle. One clue is that Logistic regression allows you to predict the probability of the nominal variable. For example, imagine that you had measured the cholesterol level in the blood of a large number of 55-year-old women, then followed up ten years later to see who had had a heart attack. You could do a two-sample *t*–test, comparing the cholesterol levels of the women who did have heart attacks vs. those who didn't, and that would be a perfectly reasonable way to test the null hypothesis that cholesterol level is not associated with heart attacks; if the hypothesis test was all you were interested in, the *T_TEST* would probably be better than the less-familiar Logistic regression. However, if you wanted to *predict* the probability that a 55-year-old woman with a particular cholesterol level would have a heart attack in the next ten years, so that doctors could tell their patients "If you reduce your cholesterol by 40 points, you'll reduce your risk of heart attack by *X*%," you would have to use Logistic regression.

Another situation that calls for Logistic regression, rather than an ANOVA or *T_TEST*, is when you determine the values of the measurement variable, while the values of the nominal variable are free to vary. For example, let's say you are studying the effect of incubation temperature on sex determination in Komodo dragons. You raise 10 eggs at 30 °C, 30 eggs at 32°C, 12 eggs at 34°C, etc., then determine the sex of the hatchlings. It would be silly to compare the mean incubation temperatures between male and female hatchlings, and test the difference using an ANOVA or *T_TEST*, because the incubation temperature does not depend on the sex of the offspring; you've set the incubation temperature, and if there is a relationship, it's that the sex of the offspring depends on the temperature.

Simple Logistic regression finds the equation that best predicts the value of the *Y* variable for each value of the *X* variable. What makes Logistic regression different from linear regression is that you do not measure the *Y* variable directly; it is instead the probability of obtaining a particular value of a nominal variable. For the spider example, the values of the nominal variable are "spiders present" and "spiders absent." The *Y* variable used in Logistic regression would then be the probability of spiders being present on a beach. This probability could take values from 0 to 1. The limited range of this probability would present problems if used directly in a regression, so the odds, *Y*/(1-*Y*), is used instead. (If the probability of spiders on a beach is 0.25, the odds of having spiders are 0.25/(1-0.25)=1/3. In gambling terms, this would be expressed as "3 to 1 odds *against* having spiders on a beach.") Taking the natural log of the odds makes the variable more suitable for a regression, so the result of a Logistic regression is an equation that looks like this: $\ln[Y/(1-Y)]=a+bX$

Logistic regression, a special case of a generalized linear model, is appropriate for these data since the response variable is binomial. The logistic regression model can be written as:

$$P(\text{failure}) = \frac{e^{Xb}}{1+e^{Xb}}$$

Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a *logit transformation* of the probability of presence of the characteristic of interest:

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

Where p is the probability of presence of the characteristic of interest. The logit transformation is defined as the logged odds:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ absence\ of\ characteristic}$$

And

$$logit(p) = \ln\left(\frac{p}{1-p}\right)$$

Rather than choosing parameters that minimize the sum of squared errors (like in ordinary regression), estimation in Logistic regression chooses parameters that maximize the likelihood of observing the sample values.

This Logistic regression line is shown on the graph; note that it has a gentle S-shape. All Logistic regression equations have an S-shape, although it may not be obvious if you look over a narrow range of values.

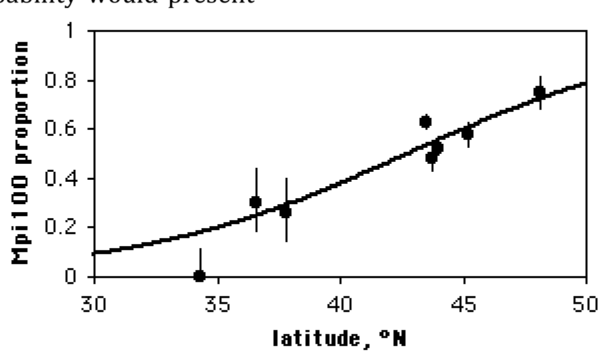

Figure1. MPI allele frequencies vs. latitude in the amphipod *Megalorchestia californiana*. Error bars are 95% confidence intervals; the thick black line is the Logistic regression line.

METHODS AND POPULATION

This present study, is applied in terms of purpose, and the design of the research study is descriptive and a correlation survey which attempts to extend the quantitative data obtained from the sample to the population. The population and its statistical samples include Tehran Stock Exchange indices in a period of 9 year from 2007 to 2015.

In general, data collection procedure can be divided into two categories of library research and field method. To gather the required information, the website of the Central Bank and the software Rah AVARD NOVIN has been utilized. For the implementation of models, the Rosetta and WEKA software, and to run the statistical tests, SPSS software was used.

Independent variables in this study are the price of gold, oil, dollar, copper, silver, inflation and construction certificate; and the dependent variable is the predicted changes of Stock Index.

Descriptive Statistics

Table 1. Results K - S

		Forecast Rough set	Forecast decision tree	Forecast logistic regression
N		3240	3240	3240
Normal parameters	Mean	0.0006	0.0071	0.8509
	Std. Deviation	0.024	0.083	0.778
Test Statistic k-s		1.289	1.197	1.263
Significant quantities		0.071	0.114	0.082

According to the table 1, greater than 0.05 significance level for all models is the accuracy. The value logistic regression model, and Rough set of normally distributed random tree addressing. So it can be for review and comparison of mean precision, independent sample T_TEST and ANOVA parametric use.

Evaluation of Models
The results of the decision tree

Table 2. The classification Table of the variable "index changes " using Random Decision Tree model

		predicted Chang=INC	Chang=No _Chg	Chang=Dec	Percent correct
	Chang= INC	1213	0	0	100
	Chang = No _Chg	10	1080	0	99/08
	Chang= Dec	0	12	924	98/71
observed Total					%99/32

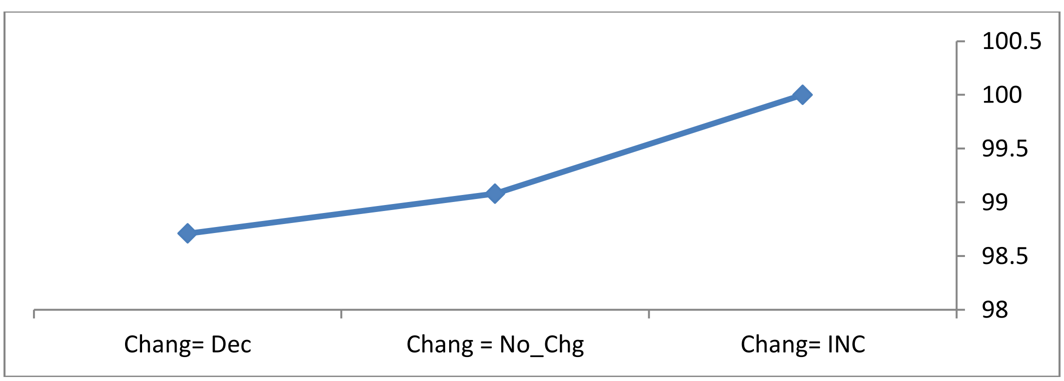

Chart 1 . The dot chart for the prediction accuracy of different levels of variable " index changes " based on Random DecisionTree model.

The Results of Rough Sets

Table 3. The classification of the variable "index changes" using Rough Sets theory

Data collection	accuracy	Data collection	accuracy
First quarter 85	97.70	Third quarter 89	100
Second quarter 85	100	Fourth quarter 89	100
Third quarter 85	98.88	First quarter 90	100
Fourth quarter 85	96.62	Second quarter 90	98.92
First quarter 86	98.86	Third quarter 90	100
Second quarter 86	100	Fourth quarter 90	100
Third quarter 86	100	First quarter 91	98.86
Fourth quarter 86	100	Second quarter 91	98.92
First quarter 87	100	Third quarter 91	100
Second quarter 87	98.92	Fourth quarter 91	98.87
Third quarter 87	100	First quarter 92	98.86
Fourth quarter 87	96.62	Second quarter 92	100
First quarter 88	100	Third quarter 92	100
Second quarter 88	98.92	Fourth quarter 92	98.87
Third quarter 88	98.88	First quarter 93	100
Fourth quarter 88	100	Second quarter 93	100
First quarter 89	100	Third quarter 93	98.88
Second quarter 89	100	Fourth quarter 93	98.88

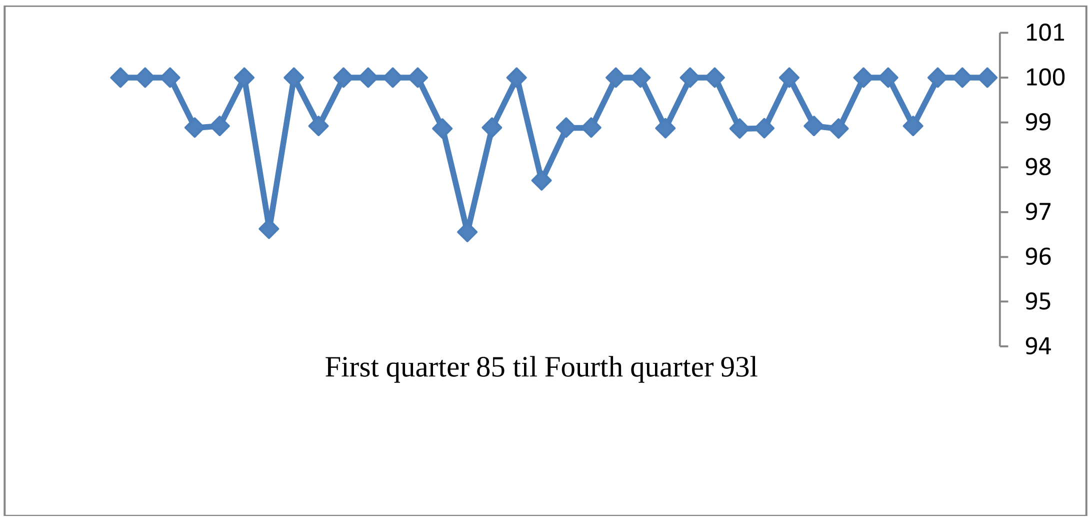

Chart 2 .The dot chart for the prediction accuracy of different levels of variable "index changes" based on the Rough Sets theory

The results of the decision tree

Table 4. The classification Table of the variable "index changes " using Random Logistic Regression model

		predicted Chang=INC	Chang=No _Chg	Chang=Dec	Percent correct
	Chang= INC	975	124	114	80.37
	Chang = No _Chg	830	143	117	13.11
observed	Chang= Dec	661	135	140	14.95
Total					38.83%

We see in this table, Logistic Regression model predicted all the class in focus 38.83% . However in first class (INCREASE) is 80.37% and in second class(NO_CHANGE) is 13.11% and in third (DECREASE) is 13.11%.

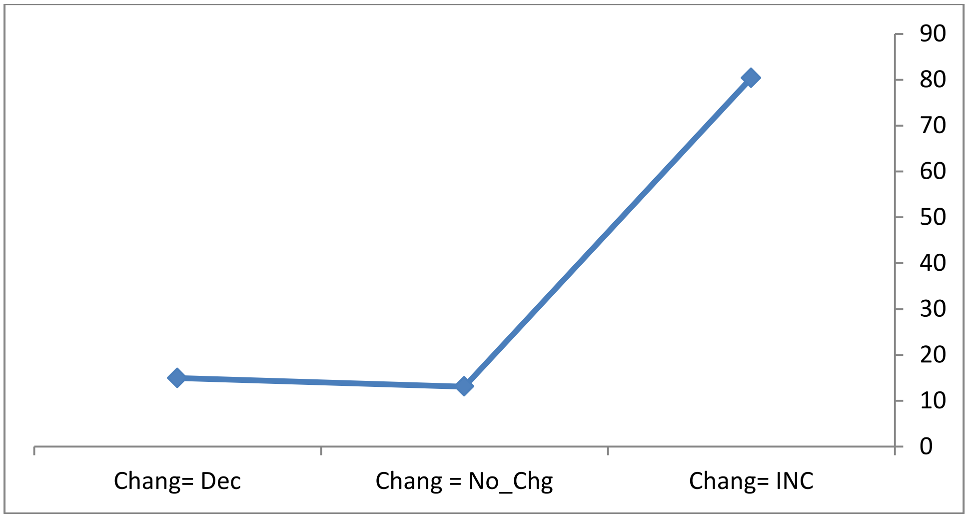

Chart 3. The dot chart for the prediction accuracy of different levels of variable " index changes " based on Logistic Regression model.

Inferential Statistics
Paired sample t-test for the decision tree model

Hypothesis 1: The decision tree model can forecast the Tehran Stock Exchange index.
H0: Decision tree model is able to forecast Tehran Stock Exchange index
H1: Decision tree model is unable to forecast Tehran Stock Exchange index

The results of the paired sample t-test for comparing the actual and predicted values of the decision tree model are shown in the table 4.

Table 4. Paired sample t-test results for the decision tree model

		LEVEN s test		T-test for comparison of means					%٩٥for SD	
		F statistic	Significant	T-statistics	Degrees of freedom	Significant bilateral	The average deviation	The standard deviation	lower limit	upper limit
Index Change	Assuming equality of variance	0/004	0/095	0/352	6477	0/725	0/007	0/020	-0/032	0/047
	Assuming Un equality of variance			0/352	6477	0/725	0/007	0/020	-0/032	0/047

According to the results of the above table 4, a significant level for LEVEN test (Sig. = 0.951) was found which is greater than 0.05, therefore, the equality hypothesis of variance of two populations (predicted values and the actual values) is confirmed. So in order to examine the comparison between the mean score of the two populations, we should refer to the relevant results and the equality of variances. The results of the paired sample t-test, in case the equality of variance is assumed, shows that the level of significance for t-test (two-tailed sig. = 0.725) is greater than 0.05, therefore, the mean of the two populations will be accepted, so it can be said that the decision tree model has had a high accuracy in the prediction of index changes variable.

Paired sample t-test for the Rough Sets

Hypothesis 2: Rough Sets model is able to predict Tehran Stock Exchange's index.

H0: Rough Sets model is able to predict Tehran Stock Exchange index.
H1: Rough Sets model is unable to predict Tehran Stock Exchange index.

The results of the paired sample t-test for the comparison of actual and predicted values of the Rough Sets model are shown in the table 5 .

Table 5 . Paired sample t-test for the Rough Sets

		LEVEN test		T-test for comparison of means					%95 for SD	
		F statistic	Significant	T-statistics	Degrees of freedom	Significant bilateral	The average deviation	The standard deviation	lower limit	upper limit
Index Change	Assuming equality of variance	0.001	0.99	٠/019	6476	0.992	-0.00069	0/020	-0.039	0.039
	Assuming Un equality of variance			٠/019	6476	0.992	-0.00069	0/020	-0.039	0.039

According to the results of the table 5, the significant level for LEVEN test (sig. = 0.99) is greater than 0.05, therefore, the hypothesis of the equality of variance of the two populations (actual and predicted values) is supported. Hence, in order to examine the comparison of the two populations' means, attention should be paid to the results of the variance equality. The results of the paired sample t-test, in case the equality of variance is assumed, shows that the level of significance for t-test (two-tailed sig. = 0.992) is greater than the alpha level 0.05; therefore, the mean equality of the two populations will be accepted, so it can be said that the Rough Sets model has had a high accuracy in the prediction of index changes variable, and that our hypothesis is supported.

Paired sample t-test for the Logistic Regression

Hypothesis 3: Logistic Regression model is able to predict Tehran Stock Exchange's index.
H0: Logistic Regression model is able to predict Tehran Stock Exchange index.
H1: Logistic Regression model is unable to predict Tehran Stock Exchange index.

The results of the paired sample t-test for the comparison of actual and predicted values of the Logistic Regression model are shown in the table 6 .

Table 6. Paired sample t-test for the Logistic Regression

		LEVEN test		T-test for comparison of means					%95 for SD	
		F statistic	Significant	T-statistics	Degrees of freedom	Significant bilateral	The average deviation	The standard deviation	lower limit	upper limit
Index Change	Assuming equality of variance	197.613	0.000	30.275	6477	0.000	0.561	٠/٠19	0.525	0.598
	Assuming Un equality of variance			30.274	6276	0.000	0.561	٠/٠19	0.525	0.598

According to the results of the table 6, the significant level for LEVEN test (sig. = 0.000) is lower than 0.05, therefore, the hypothesis of the equality of variance of the two populations (actual and predicted values) is supported. Hence, in order to examine the comparison of the two populations' means, attention should be paid to the results of the variance equality. The results of the paired sample t-test, in case the equality of variance is assumed, shows that the level of significance for t-test (two-tailed sig. = 0.000) is lower than the alpha level 0.05; therefore, the mean equality of the two populations will be rejected, so it can be said that the Logistic Regression model has had a low accuracy in the prediction of index changes variable, and that our hypothesis is rejected.

Comparison mean of these three models

We used ANOVA test for comparison the models. Result of ANOVA test shown in table7 in under:

Table 7.ANOVA test for changes

Total	square errors	Df.	Mean square error	F	sig
Change between group	1549/92	2	774/960	3792	0/000
Change enter group	1985/833	9717	0/204		
Overall changes	3535/735	9719			

According to the results of the table 7, the significant for ANOVA test (sig. = 0.000) is lower than 0.05, therefore, so the hypothesis of different significant between three models accepted.

We used t-test for found different in three models. For this gold, there are two class of statistic tests Actual-Equal variance (for example LSD) and Equal variance not Assumed. So we used LEVENE test for found Actual-Equal variance. Result of LEVENE test in shown table 8 in under:

Table 8. LEVENE test for competition changes and models

LEVENE TEST	Df.1	Df.2	Sig.
0/876	2	9717	0/583

According to the results of the table 8, the significant level for LEVEN test is greater than 0.05. so, the hypothesis Actual-Equality of variance of the Tree populations in 0.05 level are accepted. So we can use LSD test for found the different between three models. This mean is LSD test is useful for found the different. Result of LSD test shown in under table:

Table 9. Result of LSD test for competition the models

(I)	(J)	Mean difference (I-J)	Std. error	sig	%95 Confidence Interval of The Difference lower	upper
Logistic Regression	Decision Tree	0/844	0/011	0/000	0/82	0/87
	Rough set	0/850	0/011	0/000	0/83	0/87
Decision Tree	Logistic Regression	-0/844	0/011	0/000	-0/87	-0/82
	Rough set	0/006	0/011	0/564	-0/02	0/03
Rough set	Logistic Regression	-0/850	0/011	0/000	-0/87	-0/83
	Decision Tree	-0/006	0/011	0/564	-0/03	0/02

According to the results of the table 10, focus prediction of two models (Decision Tree and Rough set) have not the significant different(sig. = 0.564) , and focus prediction of Logistic Regression models have the significant different with Decision Tree and Rough set models (sig. = 0.000). We use under figure for comparison Error Mean of prediction three models:

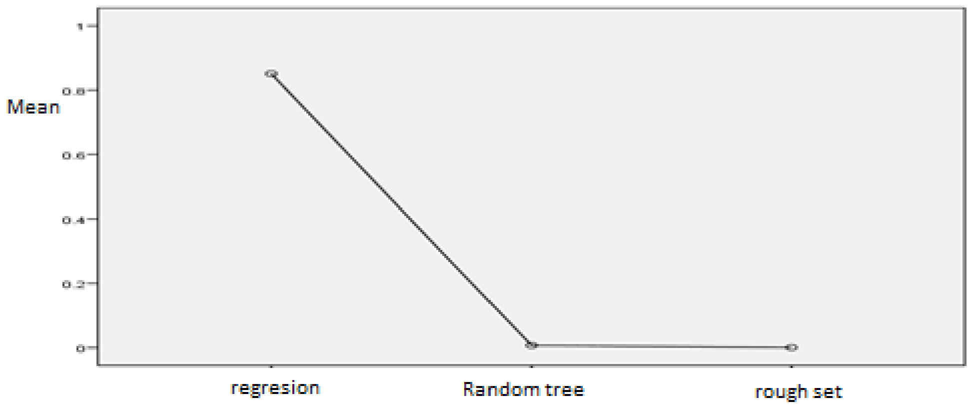

Chart 4. result of LSD in each model

According to the this figure, Error of prediction Logistic Regression greater than another models, and the Rough set model had Error of prediction lower than another models.

Assortment of the Results of Statistical Analysis and Correctness or Incorrectness of the Hypotheses

In this study, three main hypotheses were proposed, and their results after the analysis and statistical tests are as follows:

Table 10. Summary of the Hypotheses Results

Hypothesis		Title of the hypothesis	Result
Hypothesis 1	H0	Decision tree model is able to forecast Tehran Stock Exchange index	H0 is confirmed
	H1	Decision tree model is unable to forecast Tehran Stock Exchange index	H1 is rejected.
Hypothesis 2	H0	Rough Sets model is able to predict Tehran Stock Exchange index	H0 is confirmed.
	H1	Rough Sets model is unable to predict Tehran Stock Exchange index	H1 is rejected.
Hypothesis 3	H0	Logistic Regression model is able to predict Tehran Stock Exchange index.	H0 is rejected.
	H1	Logistic Regression model is unable to predict Tehran Stock Exchange index.	H0 is confirmed.

Conclusions and Suggestions for Further Research

In this paper, investigated the effectiveness of artificial intelligence on the forecasting of Tehran stock exchange index was investigated. According to research findings, some suggestions can be presented as follows:

Suggestions arising from the research

The forecast of Index using the Nearest Neighbor method and its comparison with models of decision tree, Rough Sets model, and logistic regression.

Taking longer periods of time into consideration and involving the independent variables such as subsidies and currency of various countries.

Practical suggestions for future research

The application of machine learning algorithm to predict the commercial issues in optimizing the results

Investigation and comparison of machine learning algorithm with other mathematical methods for industry profitability

The investigation of the results of effective factors in forecasting index under the influence of unavoidable circumstances such as international sanctions

REFERENCES

Adel Azar; and Afsar, Amir (2007), "The comparison of classical and artificial intelligence methods in predicting the stock price index and designing a hybrid model", Journal of Humanities teacher, Issue (4).

Afsr, A. (2006). Forecast modeling of Stock price using the neural - fuzzy network and hybrid methods. MA. thesis. Industrial Management. TARBIAT MODARES University.

Allahyari, Ebrahim. (2009). Examining the weak form of capital markets effectiveness on Tehran Stock Exchange. Stock Exchange Quarterly. No. 4

Fallahpour, Saeed; Golarzi, Gholam Hossain; Fatooreh Chian, Nasser. (2014). Forecast of the trend of stock prices using the Support Vector Machine based on the genetic algorithm in Tehran Stock Exchange. *Financial Research*. Volume 15, Issue 2. pp. 269-288.

Ghahari, Milad; Rahmanifard, Vahid. (2011). Instructions on using the software Weka. Islamic Azad University, South Tehran.

Homayon, Asadullah; Mohammadi, Hamid, Keshtkar, Rasool. (2010). Evaluating Regression models of Iran stock indices. Research and Economic Policy Quarterly. The 18th year. 56. pp.95-112.

Kelly, Logan.(2007). Measuring the Economic Stock of Money, MPRA Paper No. 5528, posted 07. Bryant University

Kooreh Pzan, Amin. (2006). Principles of fuzzy set theory and its applications. Jahad Daneshgahi Publications, AMIRKABIR Industrial Unit..

Luny, N. (2006). Stock price prediction using Artificial Neural Networks and its comparison with the VAR. Business Administration financial field. Azad University of ARAK.

Mehrara, Mohsen et al. (2010). Modeling and forecasting Tehran Stock Exchange index and determining the variables affecting it. Research and Economic Policy Quarterly. The 17th year, 50. pp. 31-51.

Ming-Chi Lee,(2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction, Department of Computer Science and Information Engineering, National Pingtung Institute of Commerce, No. 51 Minsheng E. Rd., Pingtung 900, Taiwan, ROC

Monajemi, Seyed Amir hosein, Vabazry, Mehdi; and Rayati Shavazi, AliReza (2010), "the prediction of stock prices in the Stock Exchange using a neural- fuzzy network and genetic algorithms, and its comparison with the artificial neural network", Journal of quantitative economics (former economic examination), Volume 6, Issue (3).

Moshiri, Saeed, Morovat, Habib. (2007). Forecast of the total output index in Tehran Stock Exchange using linear and nonlinear models. Journal of Commercial Research. 41.

Raee, Reza., Pooyanfar, Ahmed. (1012). Advanced Investment Management. Tehran: SAMT Publications.

Rai, Reza., Fallahpour, Saeed. (2009), "The application of the Support Vector Machine in predicting the financial distress of companies by using financial ratios", Reviews of Accounting and auditing, Volume 15, Issue 53.

Safaeefard, Behnaam. The effect of the smart phone quality on the everyday tasks speed using Rough Sets, Neural Networks, decision tree and Bayesian. B.S. student of software technology engineering, Islamic Azad University of HAMEDAN.

Sinai, Hassan Ali, Mortazavi, Saeed, and Teimoori, Yasser, (2006), "The forecasting of Tehran Stock Exchange index using artificial neural networks", Reviews of Accounting and auditing, The twelfth year, Issue 41, pp. 59-83.

Tsang, P. M., Kwok, P., Choy, S.O., Kwan, R., Ng, S.C., Mak, J., Tsang, J., Koong, K., Wong, T,L., 2007. Design and implementation of NN5 for Hong Kong stock price forecasting, Engineering Applications of Artificial Intelligence 20:453–461.

Varmazyar, Mobina. Factors influencing blood pressure in FARHANGIAN Clinic. M.A. student of Islamic Azad University, HAMEDAN Branch.

Yakup Kara, Melek Acar Boyacioglu ,mer Kaan Baykan.(2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications 38.5311–5319

Yim, J. (2002). A comparison of neural networks with time series models for forecasting returns on a stock market index .paper .School of Economics and Finance.

Zare, Asef; Nassajian, Gholamreza. The application of rough set theory in the decision-making theory. Islamic Azad University, GONABAD Branch.

Zhang, Z.Y., et al. (2005).Stock time series forecasting using support vector machines employing analyst recommendations, Springer-Ver lag Berlin Heidelberg.