PJSS
ISSN: 1309-0356

Pamukkale Journal of Sport Sciences
2011, Vol. 2, No. 3, Pg:22-36

Received: 29/05/2011
Accepted: 07/09/2011

Hematinezhad, M[1]., Ramezaniyan, M.R[1]., Gholizadeh, M.H[1]., Shafiee, S.H[1]., Ghazi Zahedi[2]

[1]Guilan University, Rasht-Iran
[2]Amirkabir University of Technology (Tehran Polytechnic), Tehran-Iran
Shafieeshahram@gmail.com

ORIGINAL ARTICLE

Predicting National Team Rank in Asian Game Using Model Tree

Abstract

Many people are interested in predicting the outcome of sporting contests. However, one of the reasons that sport attracts so much attention is that the outcome of a contest is not perfectly predictable. In this paper we tried to predict the success of nations at the Asian Games through macro-economic, political, social and cultural variables. we used the information of variables include urban population, Education Expenditures, Age Structure, GDP Real Growth Rate, GDP Per Capita, Unemployment Rate, Population, Inflation Average, current account balance, life expectancy at birth and Merchandise Trade for all of the participating countries in Asian Games from 1970 to 2006 in order to build the model and then this model was tested by the information of variables in 2010. The prediction is based on the number of golden medals acquired each country. In this research we used WEKA software that is a popular suite of machine learning software written in Java. Japans's stability is entirely consistent with it's variables in all of the courses held. The value of correlation coefficient between the predicted and original ranks is 75.5%. We tried to design the pattern that: To improve sport in each country and get the better international ranks according to it's facilities, potential sources and the comparison with other countries. Managers and planners take the appropriate policies and determine long-term, middle-term and short-term goals in sport according to political, cultural, economic and social factors.

Key Words: *Prediction, Asian Game, Macro variable and Model Tree*

Introduction

Achieving international and especially Asian sporting success has become increasingly important to a growing number of countries. Both politicians and the media count medals as a measure for international success, despite the International Asian Games Committee's protestation that the Asian Game medal table is not an order of merit. Elite sporting success has frequently been regarded as a resource valuable for its malleability and its capacity to help achieve a wide range of non-sporting objectives (Green & Houlihan, 2005). As a result, governments have become more willing to intervene directly in elite sport development by making considerable financial investments, thus leading to the increasing institutionalization of elite sport systems (e.g., Bergsgard, Houlihan, Mangset, Nodland, & Rommetvedt, 2007; Green & Houlihan, 2005). The keynote idea from this 'global sporting arms race as described by Oakley and Green (2001) is that elite sporting success can be produced by investing strategically in elite sport. From this power struggle emerged an interest in elite sport systems and the desire to explain (mainly) Asian successes. In particular, the question of why some nations succeed and others fail in international competition has been raised. This issue is obviously a key concern of policymakers who wish to improve their position in the Asian medal table.

Several papers have studied sports data predictions, given that sports events are quite distinct from completely randomized events such as lotteries. The previous literature on sports forecasting can be divided into several groups, according to the type of forecasts made. For example, several papers have aimed to predict the result of a particular match between two contestants (Boulier & Stekler, 2003; Caudill, 2003; Klaassen & Magnus, 2003). Other papers have aimed to predict the point spread between two contestants (Smith & Schwertman, 1999), and still other papers have aimed to predict the winner of sports events involving several contestants, such as tournaments (Anderson, Edman, & Ekman, 2005; Clarke & Dyte, 2000), leagues (Rue & Salvesen, 2000) or races (Bolton & Chapman, 1986). Two main methods for predicting the outcome of a sports event exist, namely statistical models and expert evaluations. Thus, some scholars have compared the accuracies of these competing methods (Anderson et al., 2005; Boulier & Stekler, 2003; Forrest, Goddard, & Simmons, 2005). With regard to statistical models, a myriad of different models have been used. For instance, if the objective is to construct

winning probabilities for contestants in a match, the most prominent approach is to use either logit (Clarke & Dyte, 2000; Klaassen & Magnus, 2003) or probit (Abrevaya, 2002; Boulier & Stekler, 1999) models. Another alternative is to use the maximum score estimator (Caudill, 2003). If a tie is possible (e.g., in soccer), either ordered probit models (Goddard & Asimakopoulos, 2004) or multinomial logit models (Forrest & Simmons, 2000) can be used. However, if the probabilities for a match are based on the number of points, goals or run probabilities, Poisson regression (Dixon & Coles, 1997) and negative binomial (Cain, Law, & Peel, 2000) models should be used to take the discrete nature of the data into account.

With regard to the variables that enter these statistical models, Goddard and Asimakopoulos (2004) and Goddard (2005) used several variables related to past results, as well as information about the number of goals scored and conceded. Forrest and Simmons (2000) used the performance in previous matches as an explanatory factor in their logit models of the English national soccer league. Dyte and Clarke (2000) used Federation Internationale de Football Association (FIFA) ratings to predict the numbers of goals scored by national teams competing in the 1998 FIFA World Cup. Similar to FIFA, other sports-governing bodies also produce rankings based on the past performances of contestants. These rankings have been used as predictors of victory in several different settings. For instance, Boulier and Stekler (1999) found that the ranking difference between contestants is a good predictor in professional tennis and collegiate basketball. Lebovic and Sigelman (2001) demonstrated the accuracy of collegiate football rankings in predicting match outcomes. Smith and Schwertman (1999) showed that the difference in rankings is a good predictor of the victory margin in collegiate basketball. Caudill and Godwin (2002) developed a heterogeneous skewness model that takes into account not only differences in rankings but also their degree. In tennis in particular, Klaassen and Magnus (2003) proposed a method of forecasting the winner of a match at the beginning of the match, as well as during it. For this purpose, they used a measure based on nonlinear differences in rankings, similar to that used by Caudill and Godwin (2002). Clarke and Dyte (2000) used tennis rankings to estimate the chance of winning as a function of the difference in rating points, and were able to estimate a player's chance of a tournament victory once the draw for the tournament became available.

However, the success of nations participating in Asian Games cannot rely on using past sporting data as the primary input to the statistical model. Too many disparate events are involved for it to be practical to derive aggregate medal predictions based on a micro model for each of the several hundred individual competitions that make up the Games. When analyzing Asian Game performances by national teams the tradition has therefore been not to use sporting data but rather to use more fundamental variables that capture the economic, demographic, social, cultural and political factors which are relevant in determining the shape of the final medals table.

Bernard and Busse (2004) made what is probably the best known contribution to the field. They estimated a tobit regression equation (which allows for the large number of countries with no medals in a Games), employing data from 1960 to 1996 (excluding, in their preferred result, those years affected by large-scale political boycotts). Their dependent variable was the medal share of country i in year t. The key explanatory variables were population and GDP per capita (the very similar values of the coefficients estimated on these two variables led Bernard and Busse to conclude that "total GDP is the best predictor of national performance"). In addition, categorical variables were also included to represent: the host country of Games; membership or former membership of the Soviet bloc; and 'other planned economies'. The coefficient estimates on all variables were positive and the levels of statistical significance suggested that the model had successfully identified the key structural determinants of Olympic performance.

Forrest et al (2010) reported in their paper the results of an exercise to forecast national team medal totals at the Beijing Olympic Games, 2008. The starting point was an established statistical model based on a regression analysis of medal totals in earlier Games, with past performance and GDP among the principal covariates. Final forecasts were successful in predicting the principal changes in medal shares relative to the 2004 Games, namely the surge in medals for China and Great Britain and the substantial fall in medals for Russia (Forrest, sanz, & Tena, 2010). Condon et al (1999) construct several models that try to predict a country's success at the Summer Olympic Games. Their data set consists of total scores for over 271 sporting events for 195 countries that were represented at the 1996 Summer Games and information they gathered on 17 independent variables. They build linear regression models and neural network models and compare the predictions of both types of models. Overall, the best neural network model outperformed the best regression model (Condon, Golden, & Wasil, 1999).

Leitner et al (2010) reported in their paper different methods for assessing the abilities of participants in a sports tournament, and their corresponding winning probabilities for the tournament are embedded in a common framework and their predictive performances compared. In this paper bookmaker consensus model correctly predicts that the final will be played by the teams from Germany and Spain (with a probability of about 20.5%), while showing that both finalists profit from being drawn in groups with relatively weak competitors (Leitner, Zeileis, & Hornik, 2010). Leitner et al (2010) in another paper forecast Spain team with a probability of 17.86% as the winner of the tournament; the second best team is Brazil with a winning probability of 15.27%. In addition to the forecast of the winning probability, information about the groups of the preliminaries and the deferent continental confederations can be obtained from the model competitors (Leitner, Zeileis, & Hornik, 2010).

Strumbelj et al examine the effectiveness of using bookmaker odds as forecasts by analyzing 10,699 matches from six major European soccer leagues and the corresponding odds from 10 different online bookmakers. They show that the odds from some bookmakers are better forecasts than those of others, and provide empirical evidence that (a) the effectiveness of using bookmaker odds as forecasts has increased over time, and (b) bookmakers offer more effective forecasts for some soccer leagues for than others (Strumbelj, & Sikonja, 2010). Song et al (2007) compare the forecasts of the outcomes of NFL Games made by 31 statistical models with those of 70 experts who predicted the winners of 496 NFL Games played in the 2000 and 2001 seasons. They also analyze the betting line predictions. There are nearly 18,000 expert and 12,000 statistical forecasts. The difference in the accuracy of the experts and statistical systems in predicting Game winners was not statistically significant. The variation in the success rates was higher among experts than statistical systems, but the betting line outperformed both. Moreover, having more information did not always improve the forecasting accuracy. Neither the experts nor the systems could profitably beat the betting line (Song, Boulier, Stekler, 2007). In this paper we tried to predict the success of nations in Asian Games through macro-economic, political, social and cultural variables.

Methodology

This part consists of 3 steps:

a) Economic, political, social and cultural variables that are important contributors for international sporting success were identified by a comprehensive literature review. Then the variables were given to relevant experts in order to rank them according to their importance, and then each variable was given a specific point according to its rank. At the end, the first eleven variables were selected as effective variables to predict the ranking of participating countries in Asian Games.

b) In second step, the information of selected variables for the participating countries were collected from 1974 to 2010. Additionally, the information of the countries include Uzbekistan, Kazakhstan, Tajikistan and Kyrgyzstan was given from 1994 to2010. The countries include Afghanistan, North Korea and Iraq were removed from this research, because of the lack of their information.

c) The ranking of participating countries in Asian Games takes place in 2 methods:

1) The ranking of participating countries is based on the number of golden medals acquired each country.

2) The ranking of participating countries is based on the number of total medals acquired each country.

In this research, the prediction is based on the number of golden medals acquired each country. The number of countries in various periods participated in Asian Games is variable. Independent variables are macro economic, political, social and cultural variables and dependent variable is the success of participating countries according to the number of golden medals acquired in Asian Games. As a logical extension of the literature review, we propose a model of the determinants of success in Asian Games.

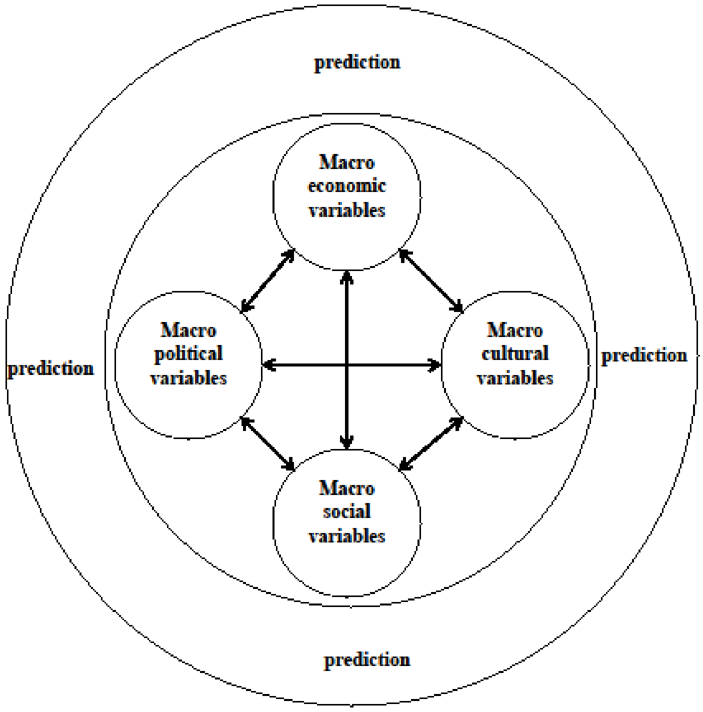


Fig 1: Model of Predicting the success of nations at the Asian Games

In this research we used WEKA (Waikato Environment for Knowledge Analysis) software that is a popular suite of machine learning software written in Java, developed at the University Of Waikato, New Zealand. WEKA is free software available under the GNU General Public License. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. WEKA, an open source collection of data mining algorithms written in java, is a solid exploratory tool for those interested in mining their collected data (Witten & Frank, 2005).

Results

Model trees (Qui92) are a technique for dealing with continuous class problems that provide a structural representation of the data and a piecewise linear fit of the class. They have a conventional decision tree structure but use linear functions at the leaves instead of discrete class labels. The first implementation of model trees, M5, was rather abstractly defined in (Qui92) and the idea was reconstructed and improved in a system called M5′ (WW97). Like conventional decision tree learners, M5′ builds a tree by splitting the data based on the values of predictive attributes. Instead of selecting attributes by an information theoretic metric, M5′ chooses attributes that minimize intra-subset variation in the class values of instances that go down each branch (Holmes, Hall & Frank, 1999).

The M5 algorithm is the most commonly used classifier of decisions trees family. Structurally, a model tree takes the form of a decision tree with linear regression functions instead of terminal class values at its leaves. The M5 model tree is a numerical prediction algorithm and the nodes of the tree are chosen over the attribute that maximizes the expected error reduction as a function of the standard deviation of output parameter (Zhang & Tsai, 2007). M5 model trees were discovered and brought by Quinlan (1992) and his theory was expanded in a method called M5/ by Wang and Witten (1997). Model trees have several advantages, making them a suitable regression method for performance analysis. The prediction accuracy of model trees is comparable to that of techniques such as ANNs and is known to be higher than the prediction of regression trees such as the CART method (Ould-Ahmed-Vall et al., 2007). Both, the derived tree structure and the regression models at the leaves, can be used to further the knowledge of nature and severity of performance problems. Model trees are also known to efficiently handle large data sets with a high number of attributes and high dimensions. At first, M5 model trees algorithm constructs a regression tree by recursively splitting the instance space. Fig. 1 illustrates a tree structure of training procedure corresponding to a given 2-D input parameter domain of x_1 and x_2. The splitting condition is used to minimize the intra-subset variability in the values down from the root through the branch to the node. The variability is measured by the standard deviation of the values that reach that node from the root through the branch, with calculating the expected reduction in error as a result of testing each attribute at that node. In this way, the attribute that maximizes the expected error reduction is chosen. The splitting process would be done if either the output values of all the instances that reach the node vary slightly or only a few instances remain. The standard deviation reduction (SDR) is calculated as follows (Quinlan, 1992):

$$\boldsymbol{SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i)}$$

Where T is the set of examples that reach the node, Ti are the sets that are resulted from splitting the node according to the chosen attribute and sd is the standard deviation (Wang & Witten, 1997). After the tree has been grown, a linear multiple regression model is built for every inner node, using the data associated with that node and all the attributes that participate in tests in the sub- tree rooted at that node. Then linear regression models are simplified by dropping attributes if it results in a lower expected error on future data. After this simplification, every sub-tree is considered for pruning. Pruning occurs if the estimated error for the linear model at the root of a sub-tree is smaller or equal to the expected error for the sub-tree. After pruning, there is a possibility that the pruned tree might have discontinuities between nearby leaves. Therefore, to compensate discontinuities among adjacent linear models in the leaves of the tree a regularization process is made. This process is started once the tree has been pruned and usually improves the prediction, especially for models based on training sets containing a small number of instances (data points) (Zhang & Tsai, 2007).

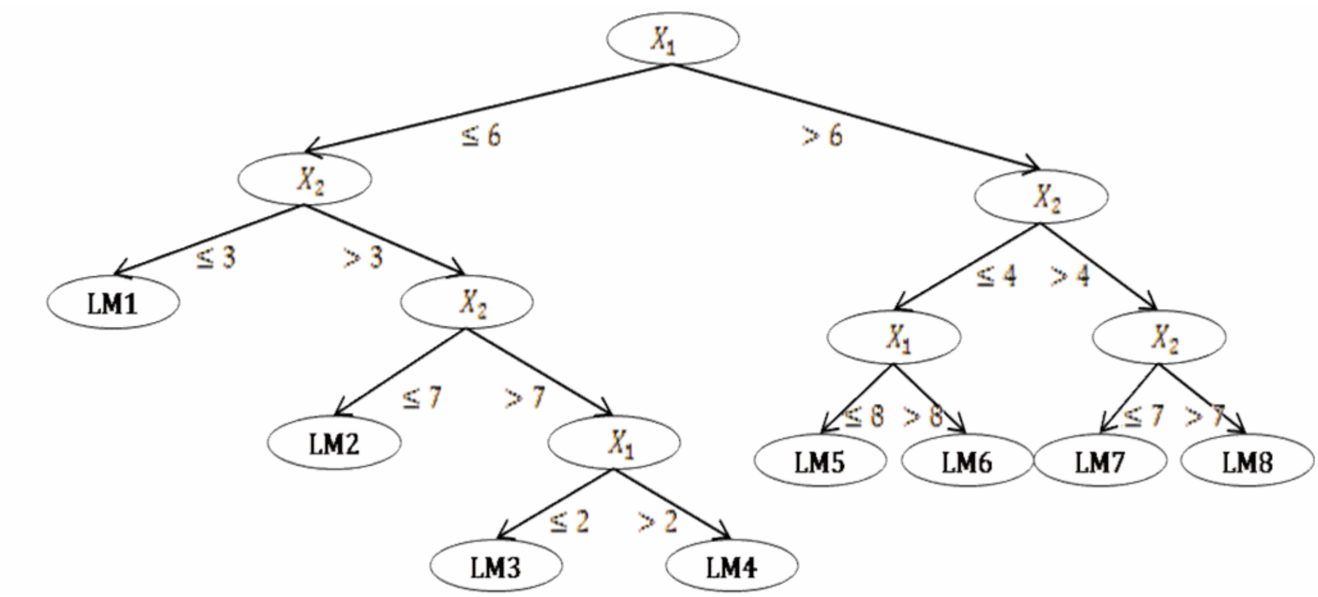


Fig.1. Example of M5 model tree (Models 1–8 are linear regression models).

The method for generating rules from model trees, which we call M5′Rules, is straightforward and works as follows: a tree learner (in this case model trees) is applied to the full training dataset and a pruned tree is learned. Next, the best leaf (according to some heuristic) is made into a rule and the tree is discarded. All instances covered by the rule are removed from the dataset. The process is applied recursively to the remaining instances and terminates when all instances are covered by one or more rules. This is the basic separate and conquers strategy for learning rules; however, instead of building a single rule, as it is done usually, we build a full model tree at each stage, and make its “best” leaf into a rule. This avoids potential for over-pruning called hasty generalization (FW98). In contrast to PART, which employs the same strategy for categorical prediction, M5′Rules builds full trees instead of partially explored trees? Building partial trees leads to greater computational efficiency, and does not affect the size and accuracy of the resulting rules.

Fig 2 shows model tree which represents the prediction of success of participating countries in Asian Games. In table1 given 10 rules and the prediction model was taken from these rules. By collecting the information for variables include Urban Population, Education Expenditures, Age Structure, GDP Real Growth Rate, GDP Per Capita, Unemployment Rate, Population, Inflation Average, current account balance, life expectancy at birth and Merchandise Trade for all of the participating countries from 1970 to 2006 and with the necessary and special training, the prediction model was designed. Then this model was tested by the information of participating countries in 2010.

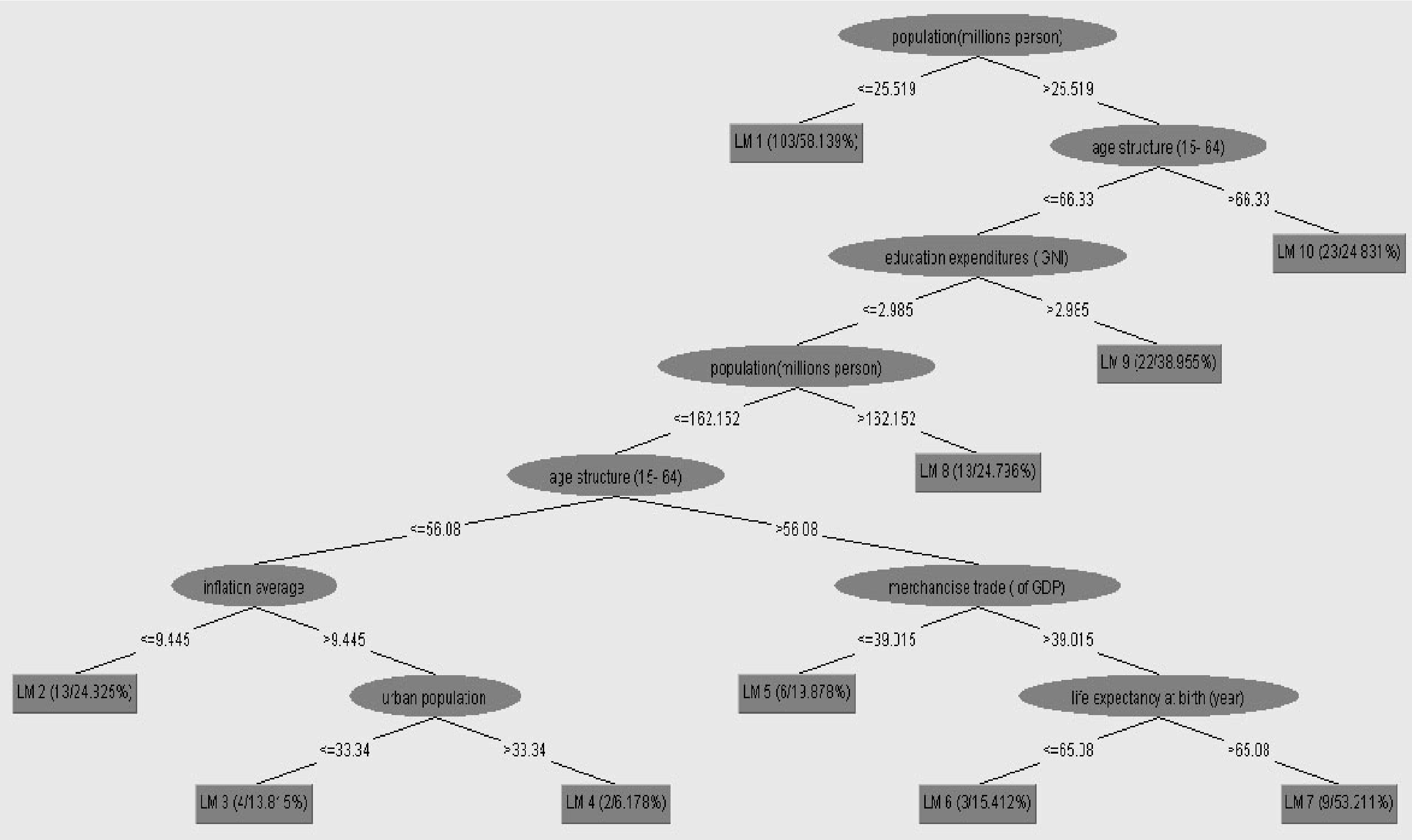


Fig.2. Performance analysis tree

For example, Rule1: GOLD= Urban Population* (-0.1092) + Education Expenditures* (-0.541) + Age Structure* (-0.297) + GDP Real Growth Rate* (- 0.016) + GDP Per Capita* (0) + Unemployment Rate* (+0.1796) + Population* (-0. 3529) + Inflation Average* (- 0.0056) + life expectancy at birth * (+0.4474) + Merchandise Trade* (+0.0033) + 15.081.

Table1. Rules for the Asian Games dataset

	LM1	LM2	LM3	LM4	LM5	LM6	LM7	LM8	LM9	LM10
Urban Population	-0.1092	-0.0893	-0.1127	-0.1141	-0.054	-0.054	-0.054	0.0388	-0.0756	-0.0883
Education Expenditures	-0.541	-0.4814	-0.4814	-0.4814	-2.2625	-1.6908	-1.6908	-0.4814	-0.9896	+0.0617
Age Structure	-0.297	+0.5517	+0.5517	+0.5517	+0.5663	+0.5633	+0.5633	+0.344	-0.0306	-0.0306
GDP Real Growth Rate	- 0.016	-0.5256	- 0.5322	- 0.5322	- 0.2673	- 0.2673	- 0.2673	-0.0171	-0.0171	-0.0171
GDP Per Capita	0	-0.0036	- 0.0036	- 0.0036	- 0.0055	- 0.0051	- 0.0051	-0.0031	-0.0003	-0.0001
Unemployment Rate	+0.1796	+0.0346	+0.0346	+ 0.0346	+ 0.0346	+ 0.0346	+ 0.0346	+0.0346	+0.0346	+0.0346
Population	-0.3529	-0.0091	- 0.0091	- 0.0091	- 0.0091	- 0.0091	- 0.0091	- 0.0148	- 0.0057	-0.0045
Inflation Average	- 0.0056	+0.0756	+0.0425	+ 0.0425	+ 0.0235	+ 0.0235	+ 0.0235	- 0.0008	- 0.0008	- 0.0008
current account balance	---	+0.0544	+0.0544	+ 0.0544	+ 0.0544	+ 0.0544	+ 0.0544	+ 0.0544	+ 0.0955	---
life expectancy at birth	+0.4474	+0.1763	+0.1763	+ 0.1763	+ 0.4981	+ 0.4921	+ 0.4757	+ 0.2369	+ 0.308	+0.0928
Merchandise Trade	+0.0033	-0.0413	- 0.0413	- 0.0413	- 0.0691	- 0.063	- 0.063	- 0.0435	+ 0.0036	+0.0036
	+15.081	-18.309	- 16.676	-16.701	-33.306	-35.829	-34.435	-15.984	- 0.5382	+6.8591

The information of model tree was reported before and after the model tree was pruned in Table 2. Correlation coefficient between original and predicted ranks for 28 participating countries in 2010 was reported 75.5%.

Table2. The information of M5 model tree for 2010 year

	number of rules	correlation coefficient	mean absolute error	root mean squared error	relative absolute error	root relative squared error	total number of instances
M5	82	0.7345	4.9017	6.8348	64.7209%	76.2674%	28
M5- Pruned	10	0.755	3.53	5.31	65.0322%	74.2026%	28

Mean absolute error (MAE)

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. The equation is given in the library references. Expressed in words, the MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally in the average.

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y}_j|$$

Root mean squared error (RMSE)

The RMSE is a quadratic scoring rule which measures the average magnitude of the error. The equation for the RMSE is given in both of the references. Expressing the formula in words, the difference between forecast and corresponding observed values are each squared and then averaged over the sample. Finally, the square root of the average is taken. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

The MAE and the RMSE can be used together to diagnose the variation in the errors in a set of forecasts. The RMSE will always be larger or equal to the MAE; the greater difference between them, the greater the *variance* in the individual errors in the sample. If the RMSE=MAE, then all the errors are of the same magnitude.

Both the MAE and RMSE can range from 0 to ∞. They are negatively-oriented scores: Lower values are better.

Relative absolute error

The relative absolute error takes the total absolute error and normalizes it by dividing by the total absolute error of the simple predictor. Mathematically, the **relative absolute error** E_i of an individual program i is evaluated by the equation:

$$E_i = \frac{\sum_{j=1}^{n}\left|P_{(ij)} - T_j\right|}{\sum_{j=1}^{n}\left|T_j - \bar{T}\right|}$$

Where $P_{(ij)}$ is the value predicted by the individual program i for sample case j (out of n sample cases); T_j is the target value for sample case j; and $\bar{T}$ is given by the formula:

$$\bar{T} = \frac{1}{n}\sum_{j=1}^{n}T_j$$

For a perfect fit, the numerator is equal to 0 and $E_i = 0$. So, the E_i index ranges from 0 to infinity, with 0 corresponding to the ideal.

Root relative squared error

The **root relative squared error** is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values. Thus, the relative squared error takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor. By taking the square root of the relative squared error one reduces the error to the same dimensions as the quantity being predicted.

Mathematically, the **root relative squared error** E_i of an individual program i is evaluated by the equation:

$$E_i = \sqrt{\frac{\sum_{j=1}^{n}\left(P_{(ij)} - T_j\right)^2}{\sum_{j=1}^{n}\left(T_j - \overline{T}\right)^2}}$$

where $P_{(ij)}$ is the value predicted by the individual program i for sample case j (out of n sample cases); T_j is the target value for sample case j; and $\overline{T}$ is given by the formula:

$$\overline{T} = \frac{1}{n}\sum_{j=1}^{n} T_j$$

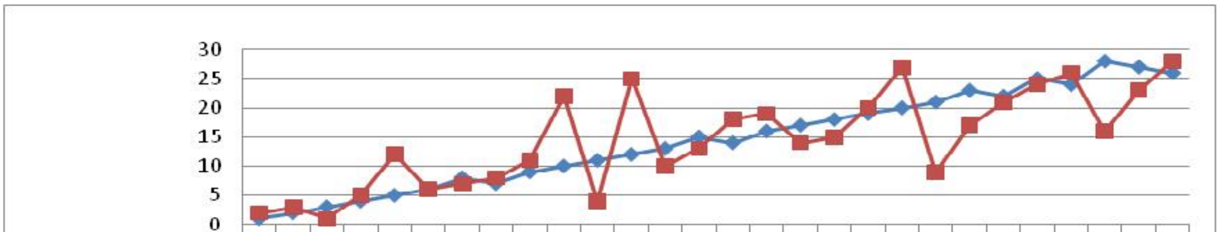
For a perfect fit, the numerator is equal to 0 and $E_i = 0$. So, the E_i index ranges from 0 to infinity, with 0 corresponding to the ideal.

The information of 28 participating countries in Asian Games in 2010 with their macro economic, political, social and cultural variables was reported in table 3.The two final columns show original and predicted records.

Table3. The information of participating countries in Asian Games in 2010

	Urban Population	Education Expenditures (% Of GNI)	Age Structure (15- 64)	GDP Real Growth Rate	GDP Per Capita (Per Person) $	Unemployment (Of Total Labor Force)	Population (Millions Person)	Inflation Average	Current Account Balance (% Of GDP)	Life Expectancy At Birth (Year)	Merchandise Trade (% Of GDP)	Original Rank	Predicted Rank
China	44	1.9	72.1	8.504	2202	4.1	1341.41	3.12	6.239	74.68	59.2	1	2
South Korea	81.7	4.2	72.3	-0.987	20329	3.3	48.91	2.9	1.607	79.05	92.27	2	3
Japan	66.6	3.7	64.3	-5.369	37555	5.145	127.471	-1.407	2.84	82.25	31.45	3	1
Iran	69	4.8	72.9	1.484	3411	12.57	75.35	8.5	2.3	70.06	46.73	4	5
Kazakhstan	58.2	4.41	70.2	1.465	6346	7.8	15.584	7.303	0.715	68.51	81.74	5	12
India	29.8	3.2	64.3	5.355	871	10.7	1215.94	13.162	-2.172	66.8	40.6	6	6
Uzbekistan	36.9	9.4	67	6.978	764	0.2	28.246	9.151	5.055	72.51	55.92	7	8
Thailand	33.7	4.9	70.5	-3.436	3577	1.39	67.653	3.245	2.496	73.6	130.86	8	7
Malaysia	71.3	4.5	63.6	-3.631	6347	3.5	28.233	2	15.379	74.12	160.71	9	11
Hong Kong	100	3.3	74.6	-3.623	29273	4.387	7.122	2	12.052	82.34	354.39	10	22
Saudi Arabia	82.3	5.7	59.5	-0.8	15886	10.476	26.106	5.2	9.1	73.12	94.03	11	4
Bahrain	83.9	2.9	70.1	3.04	14908	15	1.06	2.393	5.486	75.91	143.34	12	25
Indonesia	52.6	3.5	66	3.99	1753	7.5	134.557	4.724	1.414	70.79	51.98	13	10
Singapore	100	3.2	76.7	-3.328	33174	2.078	4.832	2.097	21.986	80.74	361.62	14	18
Qatar	95.7	3.3	76.8	11.467	34449	0.5	1.352	1.033	25.111	75.94	90.12	15	13
Kuwait	98.4	3.8	70.7	-1.51	18756	1.639	3.606	4.463	31.62	77.97	79.92	16	19
Pakistan	36.6	2.9	59.1	1.966	866	6.195	169.38	11.5	-3.824	66.53	38.11	17	14
Philippines	65.7	2.6	60.6	0.994	1499	7.2	94.013	4.953	3.492	71.83	64.82	18	15
Mongolia	57.3	5.1	67.9	0.5	1322	3	2.734	7.341	-6.641	66.57	117.07	19	20
Jordan	78.5	2	59.4	3	2707	13	6.126	5.278	-8.911	72.71	116.2	20	27
Bangladesh	27.6	2.4	61.4	5.419	436	5.1	167.671	7.385	2.088	66.15	49.31	23	17
Kyrgyzstan	36.4	6.6	64.5	1.465	589	5.57	5.431	8.428	-12.47	67.37	112.66	25	24
Vietnam	28.3	5.3	68.3	4.606	778	5	88.257	12	-6.907	74.37	158.11	21	9
Syria	54.6	4.9	59.9	3.019	1893	8.5	21.762	5.037	1.13	74.23	59.09	22	21
Tajikistan	26.5	3.5	62.1	2	465	2.2	6.536	7.039	-7.27	66.75	91.08	24	26
Laos	32	2.3	56.2	4.584	656	2.5	6.497	6.865	-10.13	64.97	44.56	26	28
Lebanon	87.1	2	67.1	7	6163	9.2	3.908	5.003	-12.79	72.05	72.47	27	23
Nepal	17.7	3.8	59.2	3.995	355	46	28.285	11.762	-2.77	66.39	37.02	28	16

In Fig 3, the diagram of original and predicted records was reported.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
original rank	1	2	3	4	5	6	8	7	9	10	11	12	13	15	14	16	17	18	19	20	21	23	22	25	24	28	27	26
predicted rank	2	3	1	5	12	6	7	8	11	22	4	25	10	13	18	19	14	15	20	27	9	17	21	24	26	16	23	28

Fig3. The diagram of original and predicted records

Discussion and Conclusion

The methods based on model tree are more useful than the others like neural networks. Because they have the possibility of production and recovery in rules. Additionally, in spite of their proper performance, they can have a good vision of model data in the estimates. In this application, M5 method is an appropriate option to estimate the numerical data and produce the linear regression which is proper for the variables represent countries. By using of pruned tree, we prevent the production of extra rules and so the results show the percentage of error has decreased. The countries with the high percentage of error in their estimates have the structure which is not followed by the public forms of model. We can plan for the change and improvement of macro variables of each country on the base of the model built by variables. Additionally, we can study the properties of the successful countries in order to plan the necessary changes in the variables of countries which have the similar situations to obtain good results in the future. Because the countries with the similar macro structures can provide the better patterns for changes. By constructing the rules, the countries with the similar structure are classified together and by this way, there is the possibility of comparison and finally we can have the programs to get the better ranks in Asian Games.

The reason that the predicted rank of each country differs from the original ones, originates from 2 factors:

1) In this research we only study the macro variables. To get the better results, it would rather be considered the role of micro and meso variables, and we can have an optimized pattern by the consensus of all of the variables.
2) In this research, we use the economic, political, cultural and social variables. We should consider that if the prediction of economic, political, cultural and social phenomena, because of their complexities and various parameters, is not impossible, it should be very difficult.

On the base of this matter, we can investigate the minor changes in the model. For example, the predicted rank of Japan was better than the original ones in 2010. It indicates Japan had the stability in all of the courses held in Asian Games. Although, it seems China got the best rank in the next courses held in Asian Games, this country never was more stable than Japan. Japans's stability is entirely consistent with it's variables in all of the courses held. The value of correlation coefficient between the predicted and original ranks is 75.5%. In the situation that we encounter with political, cultural, economic and social variables and as we know one aspect of these variables is the human being factor and it is obvious that the human being is unpredictable, so it seems this percentage for correlation coefficient factor was proper. This category was based on rules which is presented by the model.

We tried to design the pattern that:

1) To improve sport in each country and get the better international ranks according to it's facilities, potential sources and the comparison with other countries. Managers and planners take the appropriate policies and determine long-term, medium-term and short-term goals in sport according to political, cultural, economic and social factors.
2) To give the opportunity to athletes to compare themselves with athletes from the other countries in order to identify their position and plan the necessary training programs and finally, obtain the better records according to the standard templates.
3) To provide the atmosphere in which the expectations and demands of sports spectators, critics and experts are intellectual and realistic and so, they do not have emotional expectations on the results acquired by athletes in Games. This causes the destructive pressure and stress on athletes to decrease and they are able to do their best performance.

Reference

1. Abrevaya, J. (2002). Ladder tournaments and underdogs: lessons from professional bowling. *Journal of Economic Behavior and Organization*, 47, pp: 87–101.
2. Anderson, p., Edman, j., & Ekman, M. (2005). Predicting the world cup 2002 in soccer: performance and confidence of experts and non-experts. *International Journal of Forecasting*, 21, pp: 565–576.
3. Bergsgard, N. A., Houlihan, B., Mangset, P., Nodland, S. I., & Rommetveldt, H. (2007). Sport policy. A comparative analysis of stability and change. London: elsevier.
4. Bernard, A. B., & busse, M. R. (2004). Who wins the Olympic Games: economic resources and medal totals. *Review of Economics and Statistics*, 86, pp: 414–417.
5. Bolton, R., & Chapman, R. (1986). Searching for positive returns at the track: a multinomial logit model for handicapping horse races. *Management Science*, 32, pp: 1040–1060.
6. Boulier, B., & Stekler, H. (1999). Are sports seedings good predictors? An evaluation. *International Journal of Forecasting*, 15, pp: 83–91.
7. Boulier, B., & Stekler, H. (2003). Predicting the outcomes of national football league games. *International Journal of Forecasting*, 19, pp: 257–270.
8. Cain, M., Law, D., & Peel, D. (2000). The favourite-longshot bias and market efficiency in uk football betting. *Scottish Journal of Political Economy*, 47, pp: 25–36.
9. Caudill, S. (2003). Predicting discrete outcomes with the maximum score estimator: the case of the ncaa men's basketball tournament. *International Journal of Forecasting*, 19, pp: 313–317.

10. Caudill, S., & Godwin, N. (2002). Heterogeneous skewness in binary choice models: predicting outcomes in the men's ncaa basketball tournament. *Journal of Applied Statistics*, 29, pp: 991–1001.

11. Clarke, S., & Dyte, D. (2000). Using official ratings to simulate major tennis tournaments. International Transactions in Operational Research, 7, pp: 585–594.

12. Condon, E. M., Golden, B. L. & Wasil, E. A. (1999). Predicting the success of nations at the summer Olympics using neural networks. *Computers & Operations Research*, 26, pp: 1243- 1265.

13. Dixon, M., & Coles, S. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, 46, pp: 265–280.

14. Dyte, D., & Clarke, S. (2000). A ratings based poisson model for world cup soccer simulation. *The Journal of the Operational Research Society*, 51(8), pp: 993–998.

15. Forrest, D., & Simmons, R. (2000). Forecasting sport: the behavior and performance of football tipsters. *International Journal of Forecasting*, 16, pp: 317–331.

16. Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: the case of english football. *International Journal of Forecasting*, 21, pp: 551–564.

17. Forrest, D., sanz, I.,& Tena, J.D. (2010). Forecasting national team medal totals at the summer olympic games, *International Journal of Forecasting,* 26, pp: 576–588.

18. Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21, pp: 331–340.

19. Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *International Journal of Forecasting*, 23, pp: 51–66.

20. Green, M., & Houlihan, B. (2005). Elite sport development. Policy learning and political priorities. London and new york: routledge.

21. Holmes, G., hall, M., & Frank. F. (1999). Generating rule sets from model trees, proceeding ai '99 of the 12th australian joint conference on artificial intelligence: advanced topics in artificial intelligence springer-verlag London.

22. Klaassen, F., & Magnus, J. (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research*, 148, pp: 257–267.

23. Lebovic, J., & Sigelman, L. (2001). The forecasting accuracy and determinants of football rankings. *International Journal of Forecasting*, 17, pp: 105–120.

24. Leitner, CH., Zeileis, A.,& Hornik, K. (2010). forecasting sports tournaments by ratings of (prob)abilities: a comparison for the euro 2008, *International Journal of Forecasting,* 26, pp: 471–481

25. Leitner, CH., Zeileis, A.,& Hornik, K. (2010). Forecasting the winner of the FIFA world cup 2010, research report series / department of statistics and mathematics, 100. Institute for statistics and mathematics, wu vienna university of economics and business, vienna.

26. Oakley, B., & Green, M. (2001). The Production of Olympic Champions: International Perspectives on Elite Sport Development System. European Journal for Sport Management, 8, pp: 83–105.

27. Ould-Ahmed-Vall, E., Woodlee, J., Yount, C., Doshi, K. A., & Abraham, S.. (2007). Using model trees for computer architecture performance analysis of software applications. In *Proceedings of IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS'07).*

28. Quinlan, J. R. (1992). Learning with continuous classes. *In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, pp: 343-348.
29. Rue, H., & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *The statistician*, 49, Part 3, pp: 399–418.
30. Smith, T., & Schwertman, N. (1999). Can the ncaa basketball tournament seeding be used to predict margin of victory? *The american statistician*, 53(2), pp: 94–98.
31. Song. CH., Boulier, B. L., & Stekler, H. O. (2007). The comparative accuracy of judgmental and model forecasts of american football games, *International Journal of Forecasting,* 23, pp: 405–413.
32. Strumbelj, E., & Sikonja, M. R. (2010). Online bookmakers' odds as forecasts: the case of european soccer leagues, *International Journal of Forecasting,* 26, pp: 482–488.
33. Wang, Y., Witten, I.H. (1997). Induction of model trees for predicting continuous lasses. In: Proceedings of the Poster Papers of the European Conference on Machine Learning, University of Economics, Faculty of Informatics and Statistics, Prague.

34. Witten, I.H., & frank, E. (2005). Data mining: Partical mashine learning tools and techniques. 2nd edition. Morgan kaufmann, sanfrancisco, CA.
35. Zhang, D., Tsai, J.J.P., (2007). Advances in machine learning applications in software engineering, The United States of American by Idea Group Publishing (an Important of Idea Group Inc).