Football Match Results Prediction Using Artificial Neural Networks; The Case of Iran Pro League

Article · October 2014

CITATIONS
7

READS
6,787

4 authors:

S. Mohammad Arabzad
Jay Pharmed Spadana Co.
42 PUBLICATIONS **233** CITATIONS
SEE PROFILE

Me E Tayebi Araghi
Islamic Azad University
8 PUBLICATIONS **45** CITATIONS
SEE PROFILE

Sadi-Nezhad Soheil
University of Waterloo
55 PUBLICATIONS **824** CITATIONS
SEE PROFILE

Nooshin Ghofrani
sepehr aeen group corporation
1 PUBLICATION **7** CITATIONS
SEE PROFILE

International Journal of Applied Research on Industrial Engineering

Vol. 1, No. 3 (2014) 159–179

Available online at www.grandjournals.com

International Journal of
Applied Research on Industrial Engineering

Football Match Results Prediction Using Artificial Neural Networks; The Case of Iran Pro League

S. Mohammad Arabzad[1*], M.E. Tayebi Araghi[1], S.Sadi-Nezhad[1],Nooshin Ghofrani[2]

[1]Department of Industrial Engineering, Tehran Science and Research Branch, Azad Islamic University,Iran

[2]Department of management, Research Institute of ShakhesPajouh, University of Isfahan, Iran

ARTICLE INFO

Article history :
Received:
31 August 2014
Received in revised format:
29 September 2014
Accepted:
3 October 2014
Available online:
9 October 2014

Keywords :
Prediction, Sport Science, Football match results, Artificial Neural Networks (ANNs), Iran Pro League

ABSTRACT

Predicting the results of sports matches is interesting to many, from fans to punters. It is also interesting as a research problem, in part due to its difficulty, because the result of a sports match is dependent on many factors, such as the morale of a team (or a player), skills, current score, etc. So even for sports experts, it is very hard to predict the exact results of sports matches. This research discusses using a machine learning approach, Artificial Neural Networks (ANNs), to predict the outcomes of one week, specifically applied to the Iran Pro League (IPL) 2013-2014 football matches. The data obtained from the past matches in the seven last leagues are used to make better predictions for the future matches. Results showed that neural networks have a remarkable ability to predict the results of football match results.

1. Introduction

Predicting the results of sports matches is interesting to many, from fans to punters. It is also interesting as a research problem, in part due to its difficulty, because the result of a sports match is dependent on many factors, such as the morale of a team (or a player), skills, current score, etc. So even for sports experts, it is very hard to predict the exact results of sports matches.

In the previous literature, two approaches have been used to model match outcomes in association football: first, modeling the goals scored and conceded by each team; and second, modeling win–draw–lose match results directly. Due to the sensitivity of the results, in this paper goals scored approach were used.

Artificial Neural Networks (ANNs) have been used successfully in many scientific, industrial and business domains as a method for extracting knowledge from vast amounts of data. However, the use of ANN techniques in the sporting domain has been limited. In professional sport, data is stored on many aspects of teams, games, training and players. Sporting organizations have begun to realize that there is a wealth of untapped knowledge contained in the data and there is great interest in techniques to utilize this data (McCullagh & Whitfort, 2013). ANNs have been used as the primary learning algorithm for learning the patterns in the data. An ANNs is modeled in the way the neurons in the human body pass/suppress signals. Weights are associated with each of the input and activation functions are applied to their weighted sum (including a bias term) to get new set of inputs. We can have multiple such hidden layers before reaching a final layer with the required classification. Using back-propagation the weights are adjusted in the direction of the minimum mean-squared error on the training data.

This research discusses using a machine learning approach, ANN, to predict the outcomes of one week, specifically applied to the Iran Pro League (IPL) 2013-2014 football matches. The data obtained from the past matches in the seven last leagues are used to make better predictions for the future matches.

The remainder of this paper is organized as follows: section 2 gives a summary of previous works on football prediction and their limitations. Section 3 introduces the Iran Pro League (IPL) as a case study of this paper. Section 4 presents machine learning technique, Artificial Neural Networks (ANN), as a tool to predict the match results. In section 5, we briefly describe our methodology and its implementation. How we applied our methodology to football is covered in section 5. Results and discussion are presented in Section 6. Finally, conclusions and future work are given in section 7.

2. Literature Review

Football is perhaps the World's pre-eminent sport, so it is not surprising that there has been a substantial amount of research on football prediction. Actually, among all sports, football prediction is one of the most widely and deeply researched area. We thus survey prediction-related researches for football, a representative of the sports as our target domain, and categorize them into a few groups as below. These studies mostly deal with mathematical/statistical models or methods (Min et al., 2008) but there are a few researches based on machine learning techniques.

Statistical analysis has been done on football prediction. Many researchers suggested their own models or processes to analyze the results of football matches. They usually showed that their methods represent the results of football matches well. Many models and methods, such as Poisson regression models, a logistic regression model using seed positions, and an updating process for the intra-match winning probability, were used to analyze the results of football matches (Karlis & Ntzoufras, 1999; Falter & Perignon, 2000; Forrest & Simmons, 2002; Crowder et al., 2002; Dixon & Coles, 2002; Constantinou, 2012). Most of these works also give some predictions as well, but they are more focused on statistical analysis of the results of

football matches. Crowder et al. (2002) focused on modeling the soccer teams in the English Football Association League using refinements of the independent Poisson model. Dixon and Coles (2002) developed and fitted a parametric result prediction model to English league and cup football. The technique was based on a Poisson regression model but was complicated by the data structure and the dynamic nature of teams' performances. Constantinou (2012) presented a Bayesian network model for forecasting Association Football matches. The model (pi-football) was used to generate forecasts about the outcomes of the English Premier League (EPL) matches during season.

Statistical prediction on football results also has been researched. Usual process of these works is as follows: models are developed by fitting them to real data. Actually this process is similar to many machine learning approaches. Some of the works in this area took more statistical approaches in predicting football matches; they use little knowledge/information and are heavily based on pure statistical models, such as an ordered probit model and Poisson models (Koning, 2000; Koning et al., 2003; Goddard & Asimakopoulos, 2004; Goddard, 2005; Hvattum & Arntzen, 2010). Other works used models or methods that are more dependent on the information or knowledge of football matches; a Bayesian dynamic generalized linear model to estimate and predict the time dependent skills of teams, an ordered probit regression model with match venue information, an independent Poisson models with a method of modeling a team's offensive and defensive strengths, a ranking system based on the seasonal coefficients of variation (CVs) of the end-of-season points, etc. were used to predict football matches (Rue & Salvesen, 2000; Halicioglu, 2000; Crowder et al., 2002; Goddard, 2003; Goddard & Asimakopoulos, 2004; Halicioglu, 2005). Hvattum and Arntzen (2010) examined the value of assigning ratings to football teams based on their past performance in order to predict match results. The ELO rating system was used to derive covariates that are then used in ordered logit regression models. Goddard (2005) used Bivariate Poisson and Ordered probit regression models to estimate forecasting models for goals scored in English league football matches. They considered two types of outcomes comprised of the goals by each team and win–draw–lose match results. Goddard and Asimakopoulos (2004) implemented an ordered probit regression model to estimate English league football match results. As well as past match results data, the significance of the match for end-of-season league outcomes, the involvement of the teams in cup competition and the geographical distance between the two teams' home towns all contributed to the forecasting model's performance.

Machine learning techniques and related methods have been applied to football prediction. Examples include Bayesian learning, decision tree, naive Bayesian learning, expert Bayesian network, K-nearest neighbor, fuzzy logic representation with genetic and neural optimization methods in tuning the fuzzy model, etc. (Rotshteinet al., 2005; Joseph et al., 2006; Flitman, 2006; Minet al., 2008; Grunz et al., 2012). They used previous match results, i.e. win/draw/lose or scores, as training data and forecasted the results of a league or tournament matches. Rotshtein et al. (2005) proposed a model to predict the result of football match from the previous results. Acceptable simulation results can be obtained by tuning fuzzy rules using tournament data. The tuning procedure implies choosing the parameters of fuzzy-term membership functions and rule weights by a combination of genetic and neural optimization techniques. Joseph et al. (2006) used and compared Bayesian networks (BNs) with other machine learning techniques for predicting the outcome (win, lose or draw) of matches played by Tottenham Hotspur Football Club. The results were even more impressive for BNs given that, in a number of key respects, the study assumptions place them at a disadvantage. Flitman

(2006) developed a model that will readily predict the winner of Australian Football League games together with the probability of that win. This model has developed using a genetically modified neural network to calculate the likely winner, combined with a linear program optimisation to determine the probability of that occurring in the context of the tipping competition scoring regime. Minet al. (2008) proposed a framework for sports prediction using Bayesian inference and rule-based reasoning, together with an in-game time-series approach. The two different approaches cooperated in predicting the results of football matches. Grunz et al. (2012) introduced a hierarchical architecture of artificial neural networks to find tactical patterns in soccer game. The hierarchical architecture is capable of recognizing different tactical patterns and variations in these patterns. They worked on defense player structure using self-organizing map (SOM) as one of the most applicable ANN's technique.

Not only the prediction-related research but also other football-related research has been done. For example, Hirotsu and Mike (2006) proposed a game theoretic approach to modeling tactical changes of formation in a football match and demonstrated that the decisions of each team's head coach affects the probability of winning the match, using data from the Japan professional football league. Wright (2009) applied Operations research (OR) methodologies to several areas of sports, such as the scheduling of fixtures and sports officials, the making of decisions with respect to tactics and strategies during sporting events, and the forecasting of results. Zerguiniet al. (2010) determined the incidence and characteristic features of soccer injuries among African competition-level players. Recalde et al. (2013) presented a mathematical programming to Ecuadorian football federation (FEF) authorities to create feasible sports schedules could easily exceed the benefits obtained by the empirical method. Çali et al. (2013) determined the prevalence and possible causes of non-specific low back pain in male professional football players in the Turkish Super League. Conlin et al. (2013) contributed to the empirical research on delegation to experts by considering agents representing football players in contract negotiations with National Football League (NFL) teams. Arabzad et al. (2013) proposed a two-phase approach to select and rank the best football players of the English Premier League. Arabzad et al. (2014) proposed a novel approach based on Data Envelopment Analysis (DEA) and TOPSIS techniques to team seeding in sports tournaments with the case of Euro 2012 football tournament. Krenn (2014) dealt to a question whether uniform color had any impact on judging tackles in football. Jelineket al. (2014) investigated the effectiveness of using computer-based machine learning regression algorithms and meta-regression methods to predict performance data for Australian football players based on parameters collected during daily physiological tests. Coutts et al. (2014) compared the metabolic power demands between positional groups, and examine temporal changes in these parameters during Australian Football match-play.

3. The Iran Pro League (IPL)

The Iran Pro League (IPL) also known as Persian Gulf Cup is a professional football league competition for clubs located at the highest level of the Iranian football league system. The Premier League is the top tier of an extensive pyramid-like structure, above the Azadegan League (or 1st Division), the 2nd division, the 3rd Division and the lower local leagues. IFFHS (2013) ranked IPL as the 36th strongest league in the world as of 2012.

The 2001-2002 season saw the introduction of a professional football league in Iran. The first winner of the Iran Pro League was Persepolis F.C. Saipa was the 6th team to win the 6th new

edition of the league which meant 6 different teams won 6 leagues in the row. But in 2008 Persepolis regained the title after 6 years by the dramatic win again the rival Sepahan on the 96th minute of the final match and became the first team that won two titles in the new edition of the Iranian League. The next season Esteghlal did the same thing and won the league for second time on the final match day. Sepahan's dominance started in 2009-10 season where they won the league before the final match day and the same thing happened in the next season. But in 2011-12 they had a more difficult job to repeat the glory and they won it on the final match day for the third time in a row which is a new record in the history of the Iranian League. Totally, Sepahan winning 4 four championships in the league is identified as honorable known.

In 2012, the league considered reducing the amount of teams to 16, for the 2012-2013 IPL Season. This idea was dismissed for the 2012-2013 season. Up to 2012, there were 18 clubs in the league although the competition started in 2001–02 with 14 teams. Each club plays the others twice, once at their home stadium and once at that of their opponents for a total of 30 games for each club, with a total of 240 games in each season. Teams receive three points for a win and one point for a draw. Teams are ranked by total points and then goal difference and then goals scored. At the end of each season, the club with the most points is crowned as champion. Since the 2001-2002 Season, 12 seasons of IPL has been completed. To now, 31 different clubs have attended in these seasons. Among them, Esteghlal, Sepahan and Persepolis respectively have collected the most points.

As of 2008 four teams from Iran qualify for the AFC Asian Champions League annually. This includes the top three teams of the IPL together with the winner of the Hazfi Cup. If the winner of the Hazfi Cup is also among the three top IPL teams then the fourth best IPL team also qualifies. Sepahan as the first team of Iran has qualified to this tournament for nine times in the 11held periods have held and is identified as the second team of Asia in terms of collected points.

4. Artificial Neural Network (ANN)

Artificial Neural networks are built from elements that behave somewhat like individual nerve cells (or neurons).Much research has been conducted looking at the application of this technology for predicting future (Salchenberger et al., 1992). This paper investigates the development of a neural network system to predict the football match results. To ensure the nonlinear aspect of the problem is catered for, the networks have a minimum of three levels. The reader is referred to Lipperman (1987) for an overview of neural network theory. This paper outlines only the very basics of neural net theory.

Whilst many different architectures and learning algorithms for neural network models exist, most successful applications utilize the 3-layer back-propagation model, as illustrated in Fig. 1. When the input neurons receive data, a calculation is performed at each neuron, with a subsequent signal sent to each connected internal neuron, which in turn passes a signal to each output neuron. The output layer then forms the prediction.

The calculations performed at each neuron are determined by an activation function, which are usually logistic, linear, linear threshold or hard limiting (on/off). For example, the activation functions below provide a logistic (i.e. sig modal) function in the ranges (0,1) and [−1, 1], respectively.

$$f(x) = \frac{1}{1 + e^{-x}},$$

$$f(x) = \tanh(x). \tag{1}$$

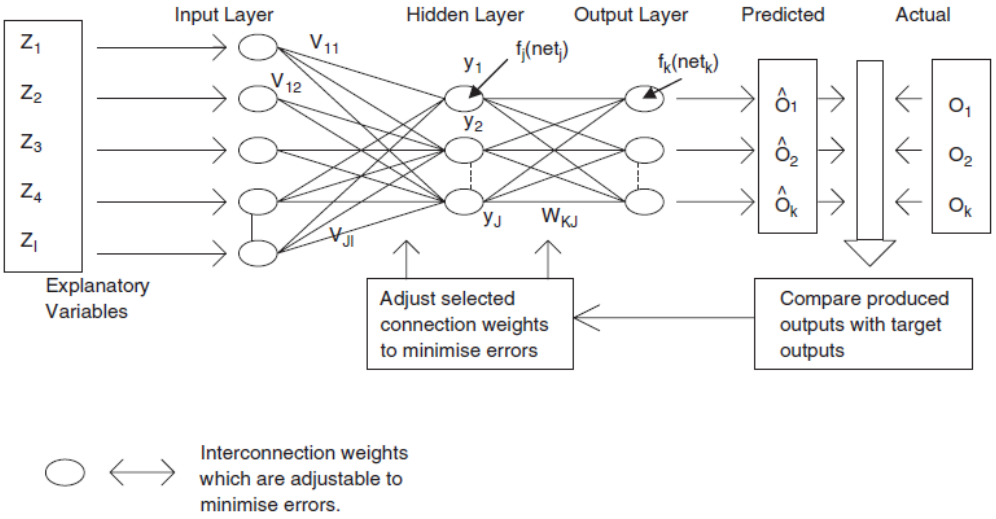


Fig 1 Basic back-propagation neural network

The size of the signal passed between any two neurons depends on both the activation function and the weight of the connection. In the network illustrated in Fig. 1 we have *I* inputs, *J* hidden neurons, and *K* outputs, with the inputs being denoted z_i, the outputs from the internal layer y_j, and the final outputs o_k. The weights connecting neurons are denoted v_{ji} and w_{kj} for connections to and from the internal layerrespectively.

The network learns by using test data. Explanatory variables are supplied and the resultant output is compared to the desired output. The network then adjusts the interconnection weights between layers. This process is repeated until the network performs well on the training set. The network can then be assessed on data not included in the test set, to evaluate its performance.

More specifically, the back propagation algorithm has four steps (Lipperman, 1987), designed to basically ensure the root mean squared error of the network on the training data is minimized. Using the notation introduced in Fig. 1 (and denoting the desired output at node *k* as d_k), these four steps can be summarizedas:

1. Sum for a single pass of the training set, the following error term, calculated for each training pattern:

$$\varepsilon = \frac{1}{2}\sum_{k=1}^{K}(d_k - O_k)^2 \tag{2}$$

2. Calculate the change in the set of weights connecting to the output level as follows:

$$\Delta w_{kj} = -\eta \frac{\partial_{\mathcal{E}}}{\partial_{w_{kj}}},$$

$$\Delta w_{kj} = \eta(d_k - o_k) f_k'(net_k) y_j, \tag{3}$$

Where η is the learning rate.

3. Now calculate the change of weights connecting to the inner level using:

$$\Delta v_{ji} = \eta f_j'(net_j) z_i \sum_{k=1}^{K} \delta_{ok} w_{kj}, \quad (4)$$

Where $\delta_{ok} = (d_k - o_k) f_k'(net_k)$.

4. Repeat this on each complete pass of the training data until no further improvement in the error term calculated in step 1 is being achieved (after a number of iterations).

The essential difference between neural networks and other forecasting techniques is that the networks use the training data to develop their representation for the modeled entity. This eliminates the situation associated with most models which must pre-determine assumptions about the modeled environment. This suggests that, in those cases where we are forced to make the most assumptions in order to model those using traditional models, neural networks may provide better results.

However, in designing a neural network, there are a number of parameters that need to be selected. These include learning rate, momentum, initial interconnection weights (all to do with learning) and the number of neurons in the hidden layer(s). The choice of these parameters can greatly influence the network's performance (Flitman and Ong, 1997). The traditional method of selecting these parameters is to adopt a trial and error approach.

5. Research Methodology

In this paper, a NN-based approach is proposed to predict the football matches results of the last week of 13thIran Pro League. To this end, some important criteria which potentially affect the match results are needed. Therefore, a comprehensive database of match results comprised of seven leagues of Iran Pro League was conducted. Mining the gathered data will enable the NN model to learn and consequently predict the results. Figure 2 shows the research methodology in four steps.

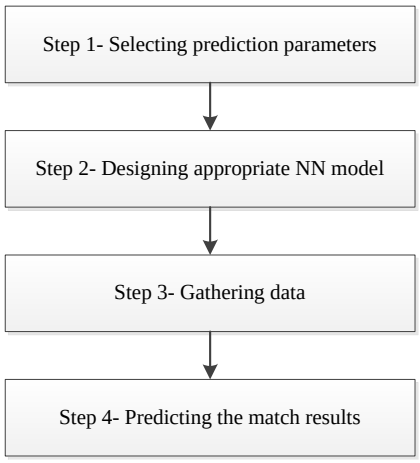


Fig 2 Research Methodology

Step 1- Selecting prediction parameters

In order to predict the match results, seven match-based criteria were selected as prediction parameters. These criteria is comprised of *The teams, Condition of teams in recent weeks,*

Condition of teams in the league, *Quality of opponents in the last matches*, *League of match*, *Week of match* and *Match results*. Prediction parameters are obtained from the seven criteria (Table 1). The parameters are considered as inputs of NN model and the obtained results of last leagues and weeks are the outputs of NN model (Figure 3).

Step 2- Designing appropriate NN model

The most applicable NN model, MLP, was chosen as the most appropriate model to predict the results of football matches. The number of 10 inputs and one output are made the structure of MLP model. Also, two hidden layers with the number of 20 neurons on each layer were considered to deal with the complexity of the problem. Other designing parameters are presented in Table 2.

Table 1 Prediction Criteria and Parameters

Match-based prediction criteria	**Prediction parameters**	***Symbols***
Teams	*The code of Home team*	I_1
	The code of Away team	I_2
Condition of teams in recent weeks	*Average of obtained points in recent 4 matches for Home team*	I_3
	Average of obtained points in recent 4 matches for Away team	I_4
Condition of teams in the league	*Average of obtained points in the league for Home team*	I_5
	Average of obtained points in the league for Away team	I_6
Quality of opponents in the last matches	*Average of obtained points by previous opponents in recent 4 matches for Home team*	I_7
	Average of obtained points by previous opponents in recent 4 matches for Away team	I_8
League of match	*The number of the league*	I_9
Week of match	*The number of the week*	I_{10}
Match results	*The number of goals by Home team*	O_1
	The number of goals by Away team	O_2

Table 2 MLP designing Parameters

Designing parameters	***Symbols***
Number of neurons in input layer	*10*
Number of hidden layers	*2*
Number of neurons in each hidden layers	*20*
Number of neurons in output layer	*2*
Learning rate	*0.05*
Learning function	*Scaled Conjugate Gradient*
Activation function for first layer	*Logsig*
Activation function for second layer	*Poslin, Purelin, Satlin, Satlins, Tansig, Logsig*

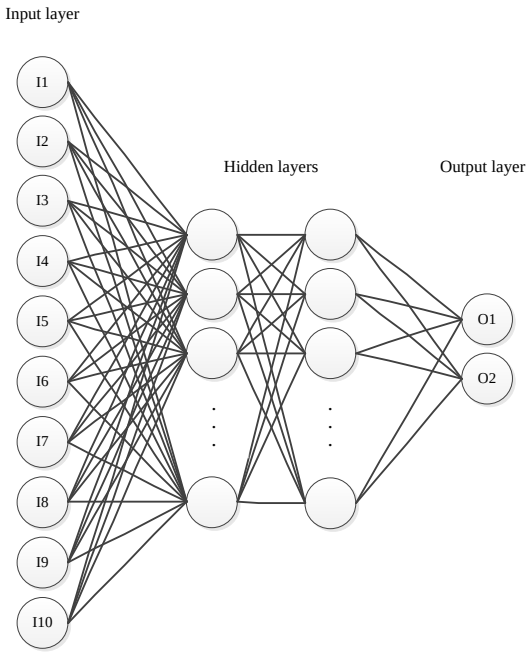


Fig 3 The Structure of NN Model

Step 3- Gathering data

In order to predict the match results of last week of 13th ILP league, data related to the six previous ILP leagues was gathered. So, the number of 1836 records of match results related to the six last IPL leagues plus 29 weeks of present league (232 records) were considered as NN learning data for predicting the last week of present league. The base of all criteria is referred to the match results. Among the seven criteria, only two criteria comprised of *Teams* and *Match results* have directly extracted from Iran Football Federation. Other criteria have indirectly

obtained from these two criteria. As it mentioned before, the seven prediction criteria are turned into prediction parameters and are used as inputs and outputs of NN model. Due to space limitation, illustrating the whole database in the article is impossible. But, Table 3 shows a part of match prediction database. It should be noticed that whole data in database are normalize before using in NN prediction model.

Table 3 APart of Match Prediction Database

Row	*Inputs*										*Outputs*	
	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	O_1	O_2
54	Perspolis	Fajr	2.5	1.0	2.6	1.0	0.8	2.1	6	7	3	3
71	Sepahan	Esteghlal T.	1.3	1.3	2.0	1.0	1.1	1.8	8	7	2	1
570	Foolad	Saba	1.8	1.5	1.8	0.8	1.5	1.6	30	8	2	2
987	Paykan	Naft T.	1.8	0.8	1.0	1.1	1.3	1.8	8	10	1	3
1073	Sanat-mes	Pas	2.5	0.5	1.8	1.0	1.2	1.4	18	10	0	0
1778	Zob-ahan	Damash	0.3	1.0	0.9	1.2	1.6	1.4	28	12	1	0
2068	Tractor S.	Esteghlal K.	0.3	0.8	1.4	1.0	1.7	1.5	29	13	1	0

Step 4- Predicting the match results

After setting prediction parameters, collecting required data, and designing NN prediction model the match results of last week of ILP league would be predictable. As it mentioned before, the five top teams comprised of Foolad FC, Naft Tehran FC, Esteghlal Tehran FC, Perspolis FC, and SepahanFC have a chance to achieve championship cup (first top team) and entitlement to participate in the AFC Championship League for next year (second and third top teams). As the standing before last week shows every five teams have a chance to be located as first to fifth position in the end of league. On the other hand, there is high intensive competition in the bottom of standing where five teams comprised of Sanat-mes FC, Damash FC, Fajr-sepasi FC, Zob-ahan FC, and Esteghlal Khuzestan FC have a chance to fall to Azadegan League. So, because of high density in the top and bottom of the standing, the prediction model must be able not only to predict the chance of victory, but also the number of goals by each team. A part of standing for these clubs is shown in Table 4. The fixture of last week is comprised of Damash vs. Sanat-mes(Match 1); Fajr-sepasivs.Malavan(Match 2); Rah-ahanvs.Zob-ahan(Match 3);Sepahanvs.Naft(Match 4);Gostareshvs. Foolad (Match 5); Esteghlal Khuzestan vs.Perspolis(Match 6);Esteghlal Tehran vs. Tractor (Match 7); Saba vs. Saipa (Match 8).

Table 4 A Part of League Standing Before Last Week

Rank	Team	Games Played	Games Won	Games Drawn	Games Lost	Goal Difference	Points
1	Naft Tehran	29	15	9	5	17	54
2	Foolad Khuzestan	29	15	9	5	11	54
3	Esteghlal Tehran	29	15	9	5	10	53
4	Perspolis Tehran	29	15	8	6	18	52
5	Sepahan Esfahan	29	13	12	4	15	51
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
12	Esteghlal Khuzestan	29	6	11	12	-10	29
13	Zob-Ahan Esfahan	29	5	11	13	-13	26
14	Fajr-Sepasi Shiraz	29	5	11	13	-15	26
15	DamashGilan	29	5	11	13	-10	25
16	Sanat-Mes Kerman	29	1	18	10	-15	21

6. Results and Discussion

The case of football match results prediction of ILP 13 has been dealt to predict the results of last week where the stand of 10 teams is so important. As the points obtained by the teams were so close, a NN prediction model considered the scores. The numbers of 2068 history records were conducted the learning data to predict 8 matches. The prediction model was run with the pointed structure and parameters. To have a robust result, matches were predicted for 30 times. The results have shown in Table 5.

After that, a statistical analysis is needed to ensure if a process is stable and predictable. To this end, an individuals and moving range (X-MR) chart as a pair of control charts for processes with a subgroup size of one was implemented. The individual (X) chart displays individual measurements. The moving range (MR) chart shows variability between one data point and the next. Individuals and moving range charts are also used to monitor the effects of process improvement theories. X-MR chat was used 16 times for each team of the eight matches. Results have shown in Table 6.

Table 5 Match Prediction Results for Last Week

Row	**The last weeks matches**															
	Match 1		**Match 2**		**Match 3**		**Match 4**		**Match 5**		**Match 6**		**Match 7**		**Match 8**	
1	1.01	0.78	1.12	0.87	1.04	0.81	1.07	0.82	1.09	0.82	0.83	1.02	1.76	0.56	1.20	0.78
2	1.46	0.85	1.08	0.94	1.51	0.87	1.55	0.88	0.88	1.05	0.84	0.99	1.38	0.92	1.13	0.85
3	1.33	0.90	1.11	0.85	1.37	0.92	1.41	0.93	1.16	1.16	1.12	1.15	1.46	1.21	1.29	0.93
4	1.17	0.96	0.98	0.88	1.21	0.98	1.24	0.99	0.69	1.08	0.52	1.04	1.33	1.24	1.04	0.95

Table 5 Match Prediction Results for Last Week

Row	The last weeks matches															
	Match 1		Match 2		Match 3		Match 4		Match 5		Match 6		Match 7		Match 8	
5	0.93	0.84	0.89	0.94	0.96	0.86	0.99	0.87	0.63	1.25	0.59	1.24	0.92	0.67	0.91	0.75
6	0.96	1.18	1.14	0.96	0.99	1.22	1.02	1.23	1.10	1.39	1.04	1.41	1.13	1.44	1.00	1.25
7	1.06	1.47	1.18	1.09	1.09	1.52	1.12	1.53	0.91	1.37	0.93	1.41	1.13	1.21	1.14	1.28
8	1.21	0.94	1.22	0.90	1.24	0.96	1.28	0.97	0.91	0.92	0.68	1.01	1.18	0.73	1.10	0.92
9	0.90	1.09	1.81	1.84	0.93	1.12	0.96	1.13	1.38	1.40	1.19	1.63	1.04	1.04	1.29	1.38
10	1.73	1.34	1.84	1.13	1.78	1.38	1.83	1.40	1.54	1.58	1.73	1.67	0.84	1.29	1.66	1.31
11	1.39	0.83	1.38	0.94	1.43	0.86	1.47	0.87	0.89	1.13	0.91	1.22	1.51	1.17	1.58	0.98
12	1.23	1.04	1.13	1.14	1.27	1.07	1.30	1.08	0.97	1.14	0.93	1.23	1.19	0.71	1.08	0.93
13	1.33	1.02	1.28	1.00	1.37	1.05	1.41	1.06	0.87	1.24	0.76	1.26	1.76	0.78	1.45	1.00
14	1.04	0.98	1.20	1.44	1.08	1.01	1.11	1.02	0.90	1.53	0.79	1.58	1.19	0.97	1.16	1.07
15	1.11	0.44	1.34	0.74	1.14	0.46	1.18	0.46	0.97	0.97	0.97	1.03	1.11	1.05	1.19	0.68
16	1.08	0.88	1.18	0.95	1.11	0.90	1.14	0.91	0.94	1.09	1.08	1.03	1.54	0.62	1.03	0.63
17	1.12	0.79	1.27	0.95	1.15	0.82	1.18	0.82	0.57	1.26	0.57	1.37	1.16	0.85	1.16	0.94
18	0.71	1.33	1.50	1.02	0.73	1.37	0.75	1.38	1.11	1.32	1.11	1.44	0.98	0.49	0.95	0.80
19	1.36	0.91	1.54	1.11	1.40	0.94	1.44	0.94	1.21	1.38	1.14	1.31	1.45	0.93	1.30	0.93
20	1.22	1.06	1.03	1.01	1.26	1.09	1.30	1.10	0.84	1.28	0.76	1.45	1.10	1.00	1.21	0.97
21	1.11	0.79	1.17	0.88	1.14	0.81	1.18	0.82	1.00	1.13	0.90	1.11	1.17	1.07	1.13	0.81
22	1.20	1.05	1.11	1.14	1.24	1.08	1.27	1.10	0.89	1.19	0.84	1.33	1.31	1.00	1.17	1.09
23	1.04	1.21	1.10	0.88	1.07	1.24	1.10	1.26	0.93	1.34	0.89	1.30	1.29	1.33	1.09	0.91
24	1.12	0.97	1.35	0.83	1.15	1.00	1.19	1.01	0.90	1.15	0.77	1.27	1.07	1.03	1.16	0.91
25	1.15	1.11	1.24	1.10	1.19	1.14	1.22	1.15	1.15	1.08	1.18	1.09	1.17	1.08	1.17	1.12
26	1.04	0.80	1.11	0.86	1.07	0.82	1.10	0.83	0.87	0.87	0.71	0.93	1.16	1.02	0.95	0.94
27	1.17	1.02	1.23	0.97	1.21	1.05	1.24	1.06	0.90	1.07	0.89	1.12	1.54	0.85	1.25	0.96
28	1.12	0.98	1.29	1.16	1.15	1.01	1.19	1.02	0.98	1.21	0.88	1.32	1.19	1.00	1.15	0.99
29	1.44	0.88	1.37	1.11	1.48	0.91	1.52	0.92	0.93	1.30	0.91	1.32	1.58	0.77	1.50	0.93
30	1.28	1.27	1.00	0.92	1.32	1.31	1.36	1.33	0.52	1.65	0.43	1.63	0.72	1.13	0.90	1.05

Table 6 Predicted Goal Scored Chats by Each Team of the Eight Matches

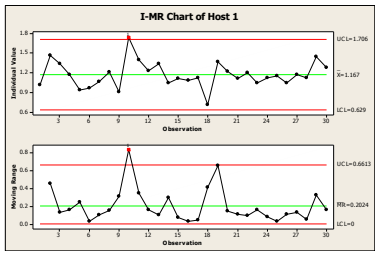
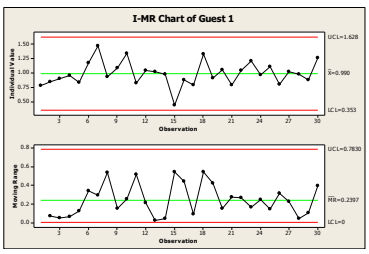
Games No.	*Host*	*Guest*
Match 1		

Table 6 Predicted Goal Scored Chats by Each Team of the Eight Matches

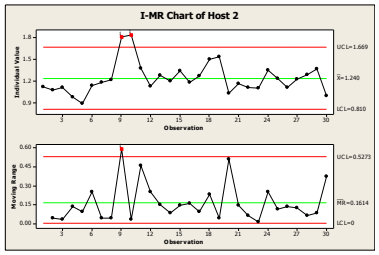
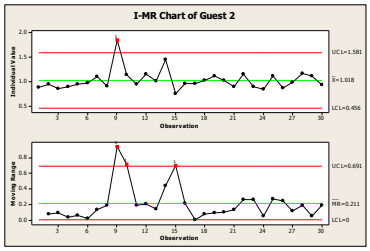
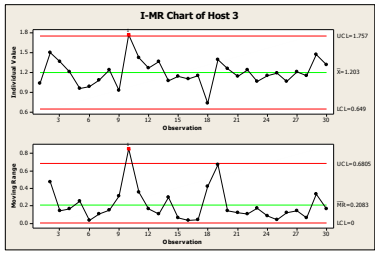
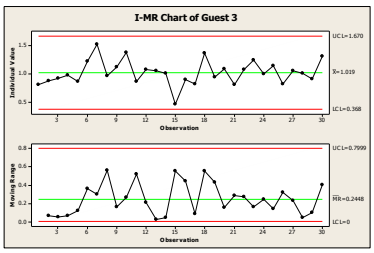
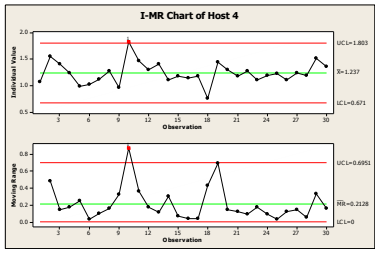
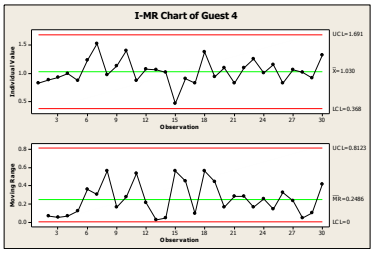
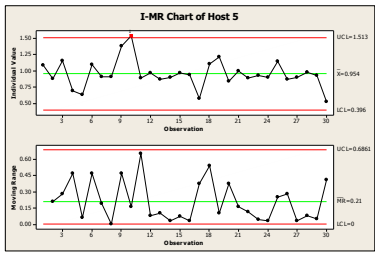
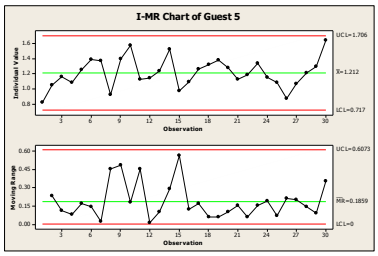
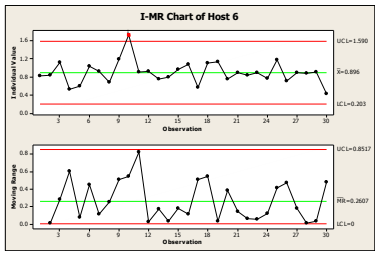
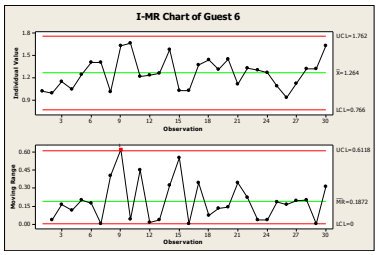
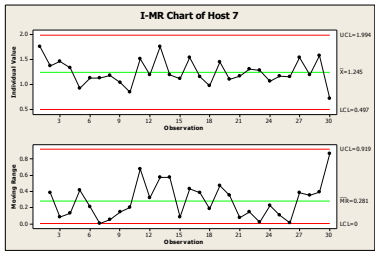
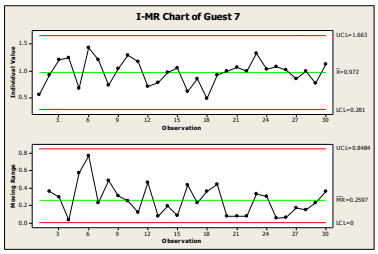
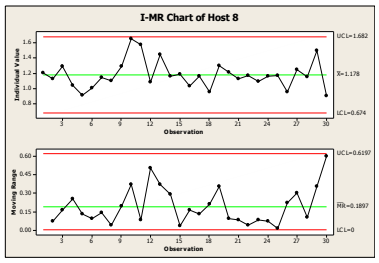
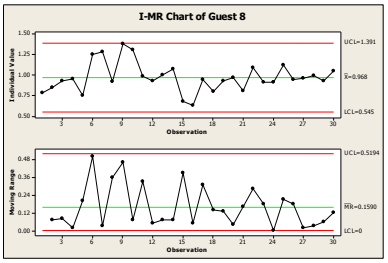
Games No.	*Host*	*Guest*
Match 2		
Match 3		
Match 4		
Match 5		
Match 6		
Match 7		

Table 6 Predicted Goal Scored Chats by Each Team of the Eight Matches

Games No.	*Host*	*Guest*
Match 8		

As it can be easily seen, there are some points which have put out of limited control area. These points should be withdrawn of data to can definitely acclaim that the points have distributed in a normal area. Tables 7 and 8 illustrate the modified data.

Table 7 The Modified Match Prediction Results for Last Week

Row	The last weeks matches															
	Match 1		Match 2		Match 3		Match 4		Match 5		Match 6		Match 7		Match 8	
1	1.01	0.78	1.12	0.87	1.04	0.81	1.07	0.82	1.09	0.82	0.83	1.02	1.76	0.56	1.20	0.78
2	1.46	0.85	1.08	0.94	1.51	0.87	1.55	0.88	0.88	1.05	0.84	0.99	1.38	0.92	1.13	0.85
3	1.33	0.90	1.11	0.85	1.37	0.92	1.41	0.93	1.16	1.16	1.12	1.15	1.46	1.21	1.29	0.93
4	1.17	0.96	0.98	0.88	1.21	0.98	1.24	0.99	0.69	1.08	0.52	1.04	1.33	1.24	1.04	0.95
5	0.93	0.84	0.89	0.94	0.96	0.86	0.99	0.87	0.63	1.25	0.59	1.24	0.92	0.67	0.91	0.75
6	0.96	1.18	1.14	0.96	0.99	1.22	1.02	1.23	1.10	1.39	1.04	1.41	1.13	1.44	1.00	1.25
7	1.06	1.47	1.18	1.09	1.09	1.52	1.12	1.53	0.91	1.37	0.93	1.41	1.13	1.21	1.14	1.28
8	1.21	0.94	1.22	0.90	1.24	0.96	1.28	0.97	0.91	0.92	0.68	1.01	1.18	0.73	1.10	0.92
9	0.90	1.09	*	*	0.93	1.12	0.96	1.13	1.38	1.40	1.19	1.63	1.04	1.04	1.29	1.38
10	*	1.34	*	1.13	*	1.38	*	1.40	*	1.58	*	1.67	0.84	1.29	1.66	1.31
11	1.39	0.83	1.38	0.94	1.43	0.86	1.47	0.87	0.89	1.13	0.91	1.22	1.51	1.17	1.58	0.98
12	1.23	1.04	1.13	1.14	1.27	1.07	1.30	1.08	0.97	1.14	0.93	1.23	1.19	0.71	1.08	0.93
13	1.33	1.02	1.28	1.00	1.37	1.05	1.41	1.06	0.87	1.24	0.76	1.26	1.76	0.78	1.45	1.00
14	1.04	0.98	1.20	*	1.08	1.01	1.11	1.02	0.90	1.53	0.79	1.58	1.19	0.97	1.16	1.07
15	1.11	0.44	1.34	0.74	1.14	0.46	1.18	0.46	0.97	0.97	0.97	1.03	1.11	1.05	1.19	0.68
16	1.08	0.88	1.18	0.95	1.11	0.90	1.14	0.91	0.94	1.09	1.08	1.03	1.54	0.62	1.03	0.63
17	1.12	0.79	1.27	0.95	1.15	0.82	1.18	0.82	0.57	1.26	0.57	1.37	1.16	0.85	1.16	0.94
18	*	1.33	1.50	1.02	*	1.37	*	1.38	1.11	1.32	1.11	1.44	0.98	0.49	0.95	0.80
19	*	0.91	*	1.11	*	0.94	*	0.94	1.21	1.38	1.14	1.31	1.45	0.93	1.30	0.93
20	1.22	1.06	*	1.01	1.26	1.09	1.30	1.10	0.84	1.28	0.76	1.45	1.10	1.00	1.21	0.97
21	1.11	0.79	1.17	0.88	1.14	0.81	1.18	0.82	1.00	1.13	0.90	1.11	1.17	1.07	1.13	0.81
22	1.20	1.05	1.11	1.14	1.24	1.08	1.27	1.10	0.89	1.19	0.84	1.33	1.31	1.00	1.17	1.09
23	1.04	1.21	1.10	0.88	1.07	1.24	1.10	1.26	0.93	1.34	0.89	1.30	1.29	1.33	1.09	0.91

Table 7 The Modified Match Prediction Results for Last Week

Row	The last weeks matches															
	Match 1		Match 2		Match 3		Match 4		Match 5		Match 6		Match 7		Match 8	
24	1.12	0.97	1.35	0.83	1.15	1.00	1.19	1.01	0.90	1.15	0.77	1.27	1.07	1.03	1.16	0.91
25	1.15	1.11	1.24	1.10	1.19	1.14	1.22	1.15	1.15	1.08	1.18	1.09	1.17	1.08	1.17	1.12
26	1.04	0.80	1.11	0.86	1.07	0.82	1.10	0.83	0.87	0.87	0.71	0.93	1.16	1.02	0.95	0.94
27	1.17	1.02	1.23	0.97	1.21	1.05	1.24	1.06	0.90	1.07	0.89	1.12	1.54	0.85	1.25	0.96
28	1.12	0.98	1.29	1.16	1.15	1.01	1.19	1.02	0.98	1.21	0.88	1.32	1.19	1.00	1.15	0.99
29	1.44	0.88	1.37	1.11	1.48	0.91	1.52	0.92	0.93	1.30	0.91	1.32	1.58	0.77	1.50	0.93
30	1.28	1.27	1.00	0.92	1.32	1.31	1.36	1.33	0.52	1.65	0.43	1.63	0.72	1.13	0.90	1.05

Table 8 The Modified Predicted Goal Scored Chats by Each Team of the Eight Matches

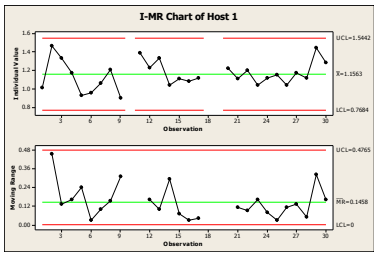
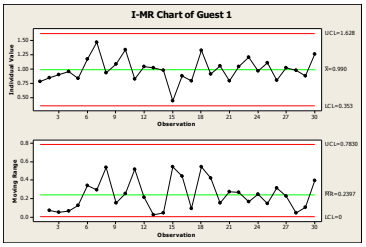
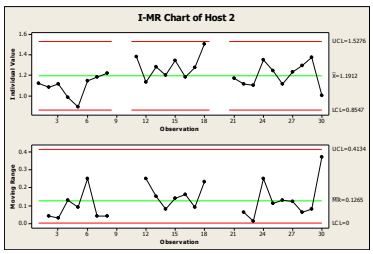
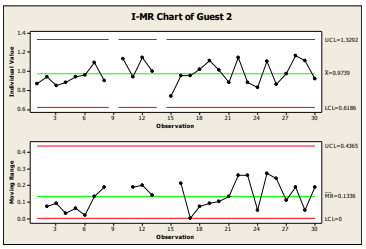
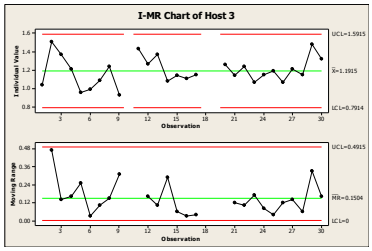
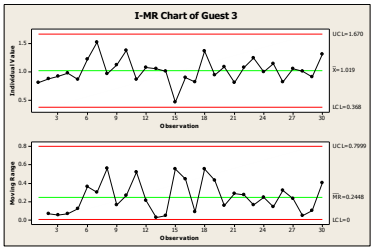
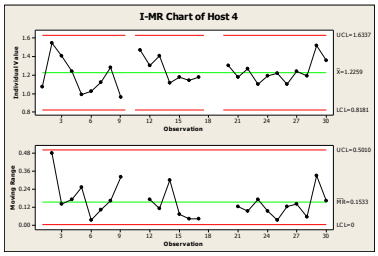
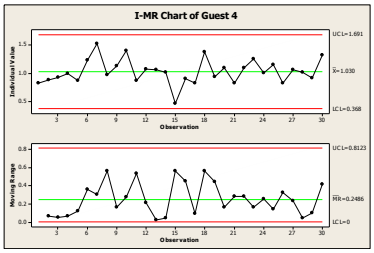
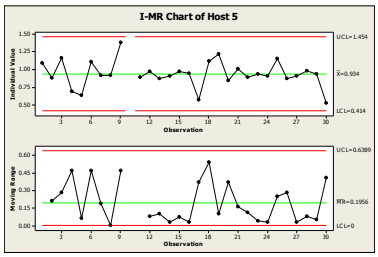
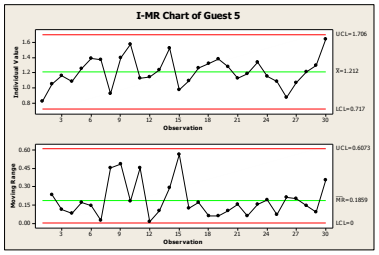
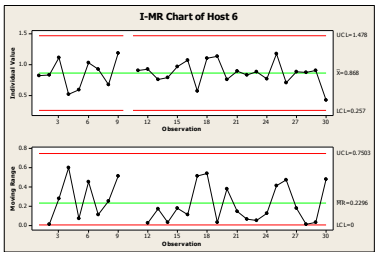
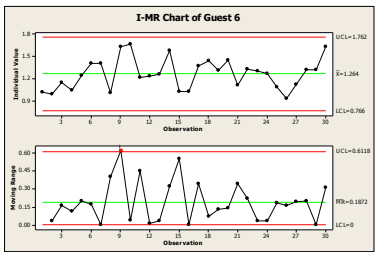
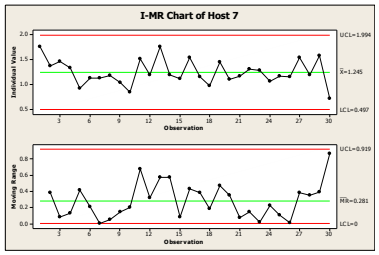
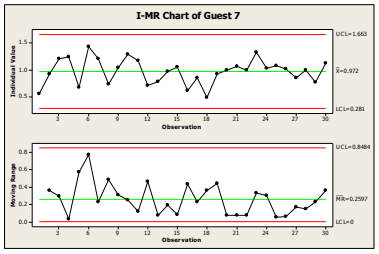
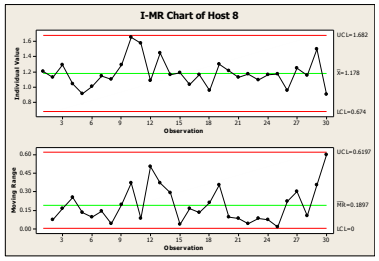
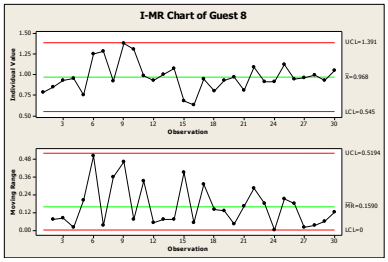
Games No.	*Host*	*Guest*
Match 1		
Match 2		
Match 3		
Match 4		

Table 8 The Modified Predicted Goal Scored Chats by Each Team of the Eight Matches

Games No.	*Host*	*Guest*
Match 5		
Match 6		
Match 7		
Match 8		

Finally, to answer the question of how the results of each football match will be, ANOVA analysis was used. In this analysis, the test is performed whether a significant difference for the calculated superiority by ANN approach between teams already there or not. For this purpose, the data of Table 5 were analyzed. The results of this analysis can be seen in Tables 9 to 16. Significant difference means sense of victory and no significant difference means draw between the teams.

Table 9 ANOVA Results for Match 1

Source	DF	SS	MS	F	P-value
Factor	1	0.469	0.469	11.52	0.001
Error	58	2.366	0.041		
Total	59	2.836			

Table 10 ANOVA Results for Match 2

Source	DF	SS	MS	F	P-value
Factor	1	0.735	0.735	16.45	0
Error	58	2.592	0.045		
Total	59	3.326			

Table 11 ANOVA Results for Match 3

Source	DF	SS	MS	F	P-value
Factor	1	0.530	0.530	12.33	0.001
Error	60	2.576	0.043		
Total	61	3.106			

Table 12 ANOVA Results for Match 4

Source	DF	SS	MS	F	P-value
Factor	1	0.647	0.647	14.41	0
Error	58	2.604	0.045		
Total	59	3.251			

Table 13 ANOVA Results for Match 5

Source	DF	SS	MS	F	P-value
Factor	1	0.993	0.993	23.67	0
Error	58	2.434	0.042		
Total	59	3.427			

Table 14 ANOVA Results for Match 6

Source	DF	SS	MS	F	P-value
Factor	1	2.024	2.024	38.83	0
Error	58	3.023	0.052		
Total	59	5.048			

Table 15 ANOVA Results for Match 7

Source	DF	SS	MS	F	P-value
Factor	1	1.121	1.121	19.15	0
Error	58	3.394	0.058		
Total	59	4.514			

Table 16 ANOVA Results for Match 8

Source	DF	SS	MS	F	P-value
Factor	1	0.662	0.662	20.63	0
Error	58	1.860	0.032		
Total	59	2.521			

Figure 4 shows graphical analysis based on ANOVA. For each match, 3 plots comprised of *normal probability of the residual*, *residuals versus the fitted values* and *histogram of the residuals* are shown. The value of P-value for each match is less than 0.005. It means that there is no draw predicted in the eight matches.

Based on the amount of significant difference for each team in each match, the final predicted results are obtained (Table 17). Also, these match results are compared with actual match results. Table 18 also compares the standing of league after last week with predicted match results.

Table 17 Final MatchesPredicted and Actual Results for Last Week

Results	The last weeks matches							
	Match 1	Match 2	Match 3	Match 4	Match 5	Match 6	Match 7	Match 8
Prediction Results	2 0	2 0	0 1	2 1	0 2	0 2	2 0	2 0
Actual Results	1 1	2 1	0 1	1 0	0 1	0 1	1 2	0 0

Table 18 Comparison of Predicted With Actual Standings

Team	**Prediction Results**				**Real Results**			
	Goal difference	**Point**	**Rank**		**Goal difference**	**Point**	**Rank**	
Foolad Khuzestan	13	57	1	↑	12	57	1	↑
Perspolis Tehran	20	55	3	↑	19	55	2	↑
Naft Tehran	16	54	4		16	54	3	↑
Sepahan Esfahan	16	54	5		16	54	4	

Table 18 Comparison of Predicted With Actual Standings

Team	Prediction Results				Real Results			
	Goal difference	Point	Rank		Goal difference	Point	Rank	
Esteghlal Tehran	12	56	2	↑	9	53	5	
.	.	.			.	.		
.	.	.			.	.		
Esteghlal Khuzestan	-12	29	12		-11	29	12	
Zob-ahan Esfahan	-12	29	13		-12	29	13	
Fajr-sepasi Shiraz	-13	29	14	↓	-14	29	14	↓
DamashGilan	-8	28	15	↓	-10	26	15	↓
Sanat-mes Kerman	-17	21	16	↓	-15	22	16	↓

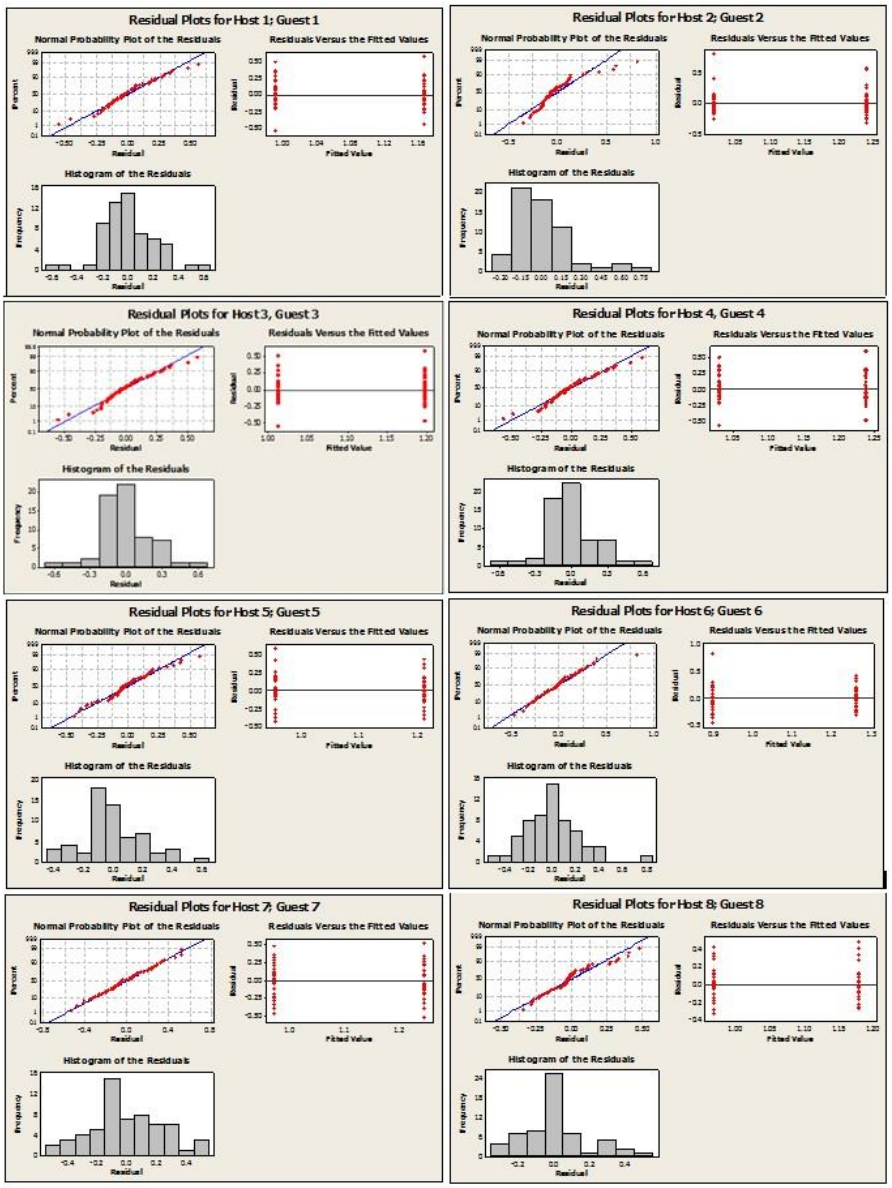


Fig 4 ANOVA Analysis for Testing the Superiority for Each Match

Although there were possible way for each of top five teams to be selected for next year AFC Champions League or even be championed before last week competitions, two out of top three teams were predicted truly. Also, in both actual and predicted results Foolad Khuzestan FC were selected as champion. On the other hand, there was nearly any possible sequence for five last teams of standing. Actual results exactly the same as actual results shows that Sanat-mes Kerman, DamashGilan and Fajr-sepasi Shiraz must be fall into Azadegan League (or 1st Division).

7. Conclusion

Predicting the results of sports matches is interesting to many, from fans to punters. It is also interesting as a research problem, in part due to its difficulty, because the result of a sports match is dependent on many factors, such as the morale of a team (or a player), skills, current score, etc. So even for sports experts, it is very hard to predict the exact results of sports matches. In this research a machine learning approach, Artificial Neural Networks (ANNs), was used to predict the outcomes of football matches. Results showed that five out of six teams were truly predicted for participating in AFC Champions League or Azadegan League.

Although some unavoidable factors affecting prediction of football matches have considered, there are some others neglected. They are introduced as major constraints of this research as bellow:

- Distance between matches, match referees and deprived players
- Impact of other competitions such as HAZFI-Cup and AFC Champions League on teams during the league
- Amount of clubs' investment, team coaches and the average age of teams
- Weather conditions

Suggestion for future research is considering the above affecting factors and comparing the results with the considered factors in this paper. Also, using Genetic Algorithm (GA) is advised to optimally set the ANNs parameters.

Acknowledgments

The authors would like to thank the Foolad Sepahan Isfahan Fan Club for their collaboration on data collection.

8. References

Arabzad, S.M., Ghorbani, M. and Shahin, A. (2013)"Ranking players by DEA the case of English Premier League",*International Journal of Industrial and Systems Engineering*, Vol. 15, No. 4, pp. 443–461.

Arabzad, S.M., Ghorbani, M. and Shirouyehzad, H. (2014)"A new hybrid method for seed determination in sport competitions: the case of European Football Championship 2012",*International Journal of Industrial and Systems Engineering*, Vol.17, No. 3, pp. 259-274.

Çali, A., Gelecek, N. and Subasi, S.S. (2013)"Non-specific low back pain in male professional football players in the Turkish super league",*Science & Sports*, Vol. 28, No. 4, pp. 93-98.

Condon, E.M., Golden, B.L. and Wasil, E.A. (1999)"Predicting the success of nations at the Summer Olympics using neural networks",*Computers & Operations Research*, Vol. 26, No. 13, pp. 1243-1265.

Conlin, M., Orsini, J. and Tang, M.C. (2013)"The effect of an agent's expertise on National Football League contract structure",*Economics Letters*, Vol. 121, No. 2, pp. 275-281.

Coutts, A.J., Kempton, T., Sullivan, C., Bilsborough, J., Cordy, J. and Rampinini, E. (2014)"Metabolic power and energetic costs of professional Australian Football match-play",*Journal of Science and Medicine in Sport*, In Press.

Crowder, M., Dixon, M., Ledford, A. and Robinson, M. (2002)"Dynamic modelling and prediction of English Football League matches for betting",*Journal of the Royal Statistical Society: Series D (The Statistician)*, Vol. 51, No. 2, pp. 157-168.

Dixon, M.J. and Coles, S.G. (1997)"Modelling association football scores and inefficiencies in the football betting market",*Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 46, No. 2, pp. 265-280.

Falter, J.M. and Pérignon, C. (2000)"Demand for football and intra match winning probability: an essay on the glorious uncertainty of sports",*Applied Economics*, Vol. 32, No. 13, pp. 1757-1765.

Flitman, A.M. (2006)"Towards probabilistic footy tipping: a hybrid approach utilising genetically defined neural networks and linear programming",*Computers & operations research*, Vol. 33, No. 7, pp. 2003-2022.

Flitman, A.M. and Ong, E.S. (1997)*Using neural networks to predict AFL game outcomes* International Conference on Computational Intelligence and Multimedia Applications. In: VermaB,Yau X., editors. Proceedings of IEEE conference on computational intelligence and multimedia applications. Queensland: Griffith University; pp. 291–295.

Forrest, D. and Simmons, R. (2002)"Outcome uncertainty and attendance demand in sport: the case of English soccer", *The Statistician*,Vol. 51, No. 2, pp. 291–241.

Goddard, J. and Asimakopoulos, I. (2003) *Modelling football match results and the efficiency of fixed-odds betting*, Working Paper, Department of Economics, Swansea University.

Goddard, J. (2005)"Regression models for forecasting goals and match results in association football",*International Journal of forecasting*, Vol. 21, No.2, pp. 331-340.

Goddard, J. and Asimakopoulos, I. (2004)"Forecasting football results and the efficiency of fixed-odds betting",*Journal of Forecasting*, Vol. 23, No.1, pp51-66.

Grunz, A., Memmert, D. and Perl, J. (2012)"Tactical pattern recognition in soccer games by means of special self-organizing maps",*Human movement science*, Vol.31, No. 2, pp. 334-343.

Halicioglu, F. (2005)"Can we predict the outcome of the international football tournaments: The Case of EURO 2000", *Dogus University Journal*,Vol. 6, No. 1, pp. 59–69.

Halicioglu, F. (2005) *Forecasting the professional team sporting events: evidence from Euro 2000 and 2004 football tournaments*, (No. 0508001), EconWPA.

Hvattum, L.M. and Arntzen, H. (2010)"Using ELO ratings for match result prediction in association football",*International Journal of forecasting*, Vol. 26, No. 3, pp. 460-470.

Iyer, S.R. and Sharda, R. (2009)"Prediction of athletes performance using neural networks: An application in cricket team selection",*Expert Systems with Applications*, Vol. 36, No. 3, pp. 5510-5522.

Jäger, J.M. and Schöllhorn, W.I. (2012)"Identifying individuality and variability in team tactics by means of statistical shape analysis and multilayer perceptrons",*Human movement science*, Vol. 31, No. 2, pp. 303-317.

Jelinek, H.F., Kelarev, A., Robinson, D.J., Stranieri, A. and Cornforth, D.J. (2014)"Using meta-regression data mining to improve predictions of performance based on heart rate dynamics for Australian football", *Applied Soft Computing*, Vol. *14*, No. 1, pp. 81-87.

Joseph, A., Fenton, N.E. and Neil, M. (2006)"Predicting football results using Bayesian nets and other machine learning techniques", *Knowledge-Based Systems*, Vol. 19, No. 7, pp. 544-553.

Karlis, D. and Ntzoufras, I. (1999) *On Modelling Association Football Data*, Technical Report, Department of Statistics, AUEB.

Koning, R.H. (2000)"Balance in competition in Dutch soccer",*Journal of the Royal Statistical Society: Series D (The Statistician),*Vol. 49, No. 3, pp. 419-431.

Koning, R.H., Koolhaas, M., Renes, G. and Ridder, G. (2003)"A Simulation Model for Football Championships",*European Journal of Operational Research*, Vol. 148, No. 2, pp. 268–276.

Krenn, B. (2014)"The impact of uniform color on judging tackles in association football",*Psychology of Sport and Exercise*, Vol. 15, No. 2, pp. 222-225.

Lippmann, R. P. (1987) "An introduction to computing with neural nets", *ASSP Magazine, IEEE,*Vol. 4, No. 2, pp. 4-22.

McCullagh, J. and Whitfort, T. (2013)"An investigation into application of artificial neural networks to the prediction of injuries in sport",*International Journal of Medical, Pharmaceutical and Engineering*, Vol. 7, No. 7, pp. 47-51.

Min, B., Kim, J., Choe, C.,Eom, H. and McKay, R.I. (2008)"A compound framework for sports results prediction: A football case study",*Knowledge-Based Systems*, Vol. 21, No. 7, pp. 551-562.

Pfeiffer, M. and Hohmann, A. (2012)"Applications of neural networks in training science",*Human Movement Science,* Vol. 31, No. 2, pp. 344-359.

Recalde, D., Torres, R. and Vaca, P. (2013) "Scheduling the professional Ecuadorian football league by integer programming",*Computers & Operations Research*, Vol. 40, No. 10, pp. 2478-2484.

Rotshtein, A.P., Posner, M. and Rakityanskaya, A.B. (2005)"Football predictions based on a fuzzy model with genetic and neural tuning",*Cybernetics and Systems Analysis*, Vol. 41, No. 4, pp. 619-630.

Rue, H. and Salvesen, O. (2000)"Prediction and retrospective analysis of soccer matches in a league",*Journal of the Royal Statistical Society: Series D (The Statistician)*, Vol. 49, No. 3, pp. 399-418.

Zerguini, Y., BunguKakala, J. P., Junge, A. and Dvorak, J. (2010)"Étude des blessureslorsd'uneCompétitionnationaleafricaine de football.À propos du championnat EPFKIN 2006–2007 à Kinshasa (RD Congo)",*Journal de Traumatologie du Sport*, Vol. 27, No. 4, pp. 171-176.