



Estimating the England Premier League Ranking with Artificial Neural Network

Hasan Aka, Zait Burak Aktuğ & Faruk Kılıç

To cite this article: Hasan Aka, Zait Burak Aktuğ & Faruk Kılıç (2021) Estimating the England Premier League Ranking with Artificial Neural Network, Applied Artificial Intelligence, 35:5, 393-402, DOI: [10.1080/08839514.2021.1901030](https://doi.org/10.1080/08839514.2021.1901030)

To link to this article: <https://doi.org/10.1080/08839514.2021.1901030>



Published online: 17 Mar 2021.



Submit your article to this journal [↗](#)



Article views: 639



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 4 View citing articles [↗](#)

ARTICLE



Estimating the England Premier League Ranking with Artificial Neural Network

Hasan Aka ^a, Zait Burak Aktuğ ^a, and Faruk Kılıç ^b

^aSport Sciences Faculty, Niğde Ömer Halisdemir University, Niğde, Turkey; ^bGazi University, Department of Machine, Technical Sciences Vocational College, Ankara, Turkey

ABSTRACT

The aim of this study is to estimate the teams' league rankings at the end of the season by using different parameters peculiar to soccer with artificial neural networks (ANNs). In this study, the values belonging to stealing the ball, number of passes (pass on target, forward pass, and pass before goal scoring), number of possessions during the match, attack time resulting in the goal scoring and number of shots in 1140 competitions played in 2015/2016, 2016/2017, and 2017/2018 England Premier League seasons have been evaluated. Season ranking in the 2017/2018 season has been estimated by analyzing the data in the first two seasons (2015/2016, 2016/2017). All data have been separated randomly for training and test. League ranking has been modeled numerically as 0 and 1. Because the generated value is between 0 and 1, the league ranking has been obtained by multiplying this value by 100 for a trained network. Thanks to the ANN model developed by training and testing according to the findings, the training, validation, test, and all regression values of the English Premier League have been obtained as 0.99779, 0.98123, 0.96981, and 0.98769, respectively. With respect to this result, it has been seen that number of shot, stealing the ball, attack time, and number of possessions parameters are determinant in team ranking at the end of the season along with the other parameters in the England Premier League. We think that analyzing matches with the ANN model provides fast and objective results for team managers, trainers, athletes, and betting shops.

ARTICLE HISTORY

Received 11 July 2020
Revised 16 October 2020
Accepted 5 March 2021

Introduction

Giving feedbacks about their performances to the athletes is an important compound in coaching profession (Maslovat and Franks 2008). However, trainers' feedbacks have been limited by a great deal of incidents taking place in team sports competitions. It has been confirmed that trainers remember a small number of incidents (42–59%) in the competitions (Laird and Waters 2008). This situation has required that matches and performances should be analyzed with numbers and data by using scientific tools. Match

analysis has provided an opportunity to objectively record the player and team behaviors during the competition in terms of significant criteria (Carling et al. 2008). Because of this, athletes' technical, tactical, physical, and behavioral performances before, during, and after the match have been determined by an expert analyst using computer-based programs (Baca 2014; Carling, Williams, and Reilly 2005).

Nowadays, parameters affecting performance have enhanced and formed an extensive data set with increasing usage of software and hardware technologies in sports. In order to correctly interpret these data, different kinds of analyses programs and artificial intelligence technologies have been used by expert analysts. One of the artificial intelligence algorithms used in sport is the artificial neural network (ANN) model. ANN is a mathematical model having qualifications similar to the biological structure of human brain (Bartlett 2006). ANN has the ability to induce the stored data obtained with intercellular connections gathered at the end of an entirely known training period (Haykin 1999). ANN generalizes both past and present data thanks to its ability to learn and make estimations (Lutz 2015; Schöllhorn, Jäger, and Janssen 2008). Analysis, generalization, association, optimization, learning, and ranking processes can be carried out successfully with ANN (Öztemel 2003).

Technics giving accurate and fast results have been required for athletes' and trainers' success. Analyzing matches with developed ANN models can provide faster inferences considering that the teams have a lot of trainings and matches in a week especially in top-level soccer leagues like the Premier League. In literature, it has been seen that there are limited numbers of researches investigating the match analyses with the ANN model and present researches are for estimating the competition results (Arabzad, Araghi, and Soheil 2014; McCabe and Trevathan 2008). There is no research done with the ANN model in determining the team ranking in soccer. It has been thought that estimating the team ranking at the end of the season can give team managers, coaches, athletes, and betting shops advantages. Concordantly, the aim of this study is to estimate the England Premier League team ranking with the ANN model according to varieties like stealing the ball, number of passes (pass on target, forward pass, and pass before goal scoring), number of possessions during the match, attack time resulting in the goal scoring, and number of shots.

Methods

In this study, machine learning method has been used in estimating the team ranking in soccer. ANN model in MATLAB (Neural Network Toolbox) software has been used in order to develop the model. In the study, the values belonging to stealing the ball, number of passes (pass on target,

forward pass, and pass before goal scoring), number of possessions during the match, attack time resulting in the goal scoring, and number of shots in 1140 competitions played in 2015/2016, 2016/2017, and 2017/2018 England Premier League seasons have been used. England Premier League match data used in the study have been obtained from one of the international analysis companies.

Model Parameters

Competitions in soccer have been played as two 45- min periods. Teams must score at least one goal more than their opponents in order to beat them. In matches, teams get 3 points for win, 1 point for draw, and 0 (zero) point for defeat. League ranking has been constituted from the points the teams gathered in the matches during the season. Stealing the ball, number of passes, number of possessions during the match, attack time, and number of shots have been determined as input parameters in this study.

Input Parameters

Stealing the ball: It is to take the ball from the opposing team.

The number of passes on target: It is the number of passes that players can make to their teammates.

Forward pass: It is the beginning of an attack as a result of the pass that a player can make to his/her teammates.

The number of passes before the goal scoring: It is the number of passes that teams make before the goal scoring.

Possession of the ball: It the total time that teams can control the ball during the match.

Attack time: It is the total time that teams try to score in their opponents' area.

Number of shots: It is the number of shots that teams make toward their opponents' goal area.

Normalization

The highest and lowest values have been determined as 1 and 0, respectively, and all inputs and outputs have been subjected to the normalization process in the study. The formula used in normalization process has been presented in the equation

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (\text{Sözen, Arcaklioğlu, and Özkaymak 2005}).$$

where x' is the normalized value, x is the initial value, $\max(x)$ is the maximum value, and $\min(x)$ is the minimum value.

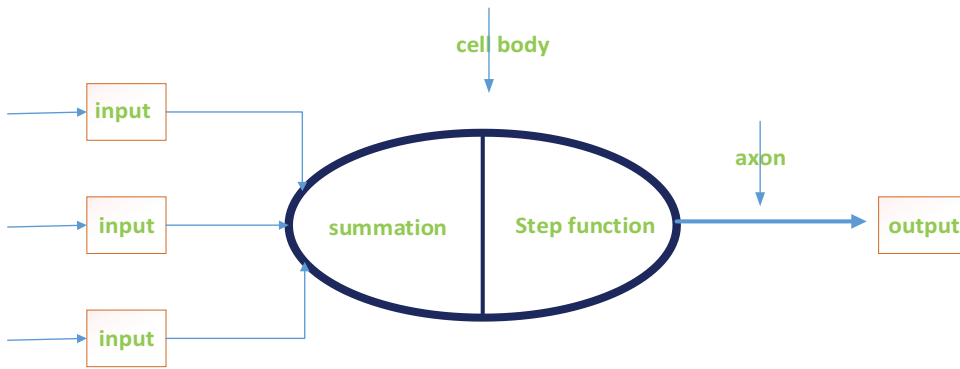


Figure 1. The simple neuron model.

Mean Square Error (MSE) Validation

MSE term is a benchmark used for determining the performance of the predictor. It is desired that the MSE value is close to zero (Salman, Kukrer, and Hocanin 2017).

$$MSE = \frac{1}{n} \sum_{i=1}^n [Y_i(m) - Y_i(p)]^2$$

Artificial Neural Network and Modeling

A neuron is a mechanism deciding to trigger or not depending on the threshold value by evaluating the input value (Figure 1). Neurons can make decisions by communicating with one another (Menet et al. 2020).

There is a weight on the transmission line while inputs are transmitted to neurons from the access level. Weight factors are effective between neurons and inputs. Each input has been transmitted to neurons by multiplying it by weight values. The neurons having memory and learning qualifications have formed a model by creating a network between the input and the output. Formed models have been built for obtaining one output value against seven input values.

Findings

The model has been lined up as input, hidden layer, and output value from left to right. One output value has been acquired against seven input values by using different layers in developed models (Figure 2).

The same network qualifications have been determined in four different developed models. Feed-forward backprop has been operated as network type. Training function and adoption learning function have been chosen as TRAINLM and LEARNGDM on the networking screen, respectively.

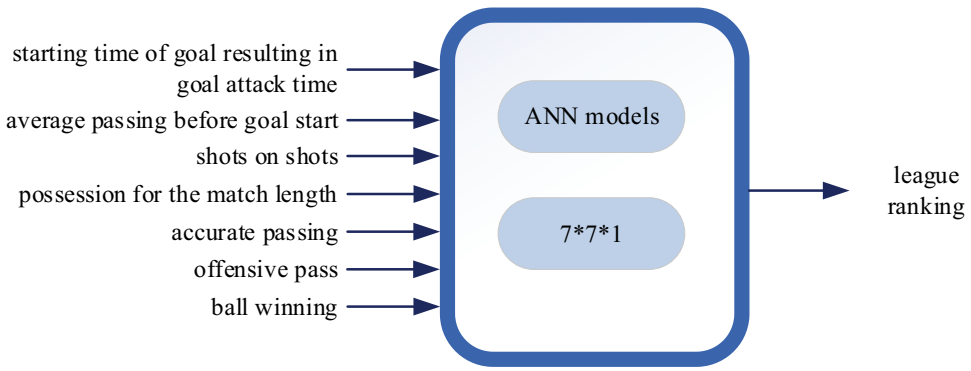


Figure 2. MATLAB ANN models and input, output parameters.

The number of epochs is in horizontal axis, and the MSE value is in vertical axis according to Figure 3. The best validation performance (BVP) value has been obtained in fourth epoch. BVP has been determined as 0.002134.

When the figure has been examined, training, validation, test, and all values of the model can be seen. ANN training with the input values of 2015–2016, 2016–2017 seasons have been executed with 0.99779 regression value. The 7*7*1 network validation has been realized as 0.98123. The test regression

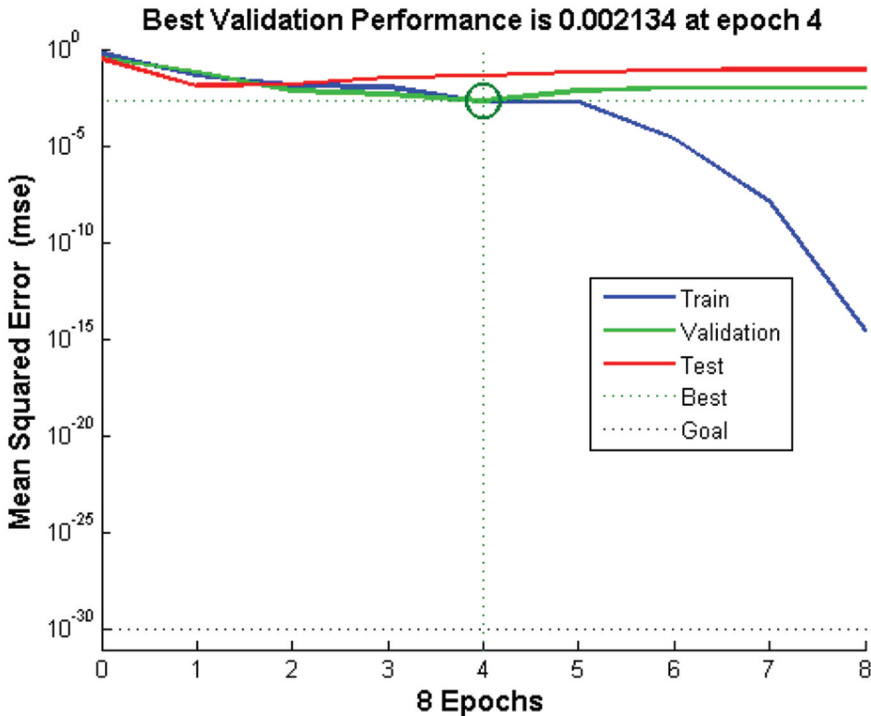


Figure 3. The best mean squared error (MSE) value of the developed model.

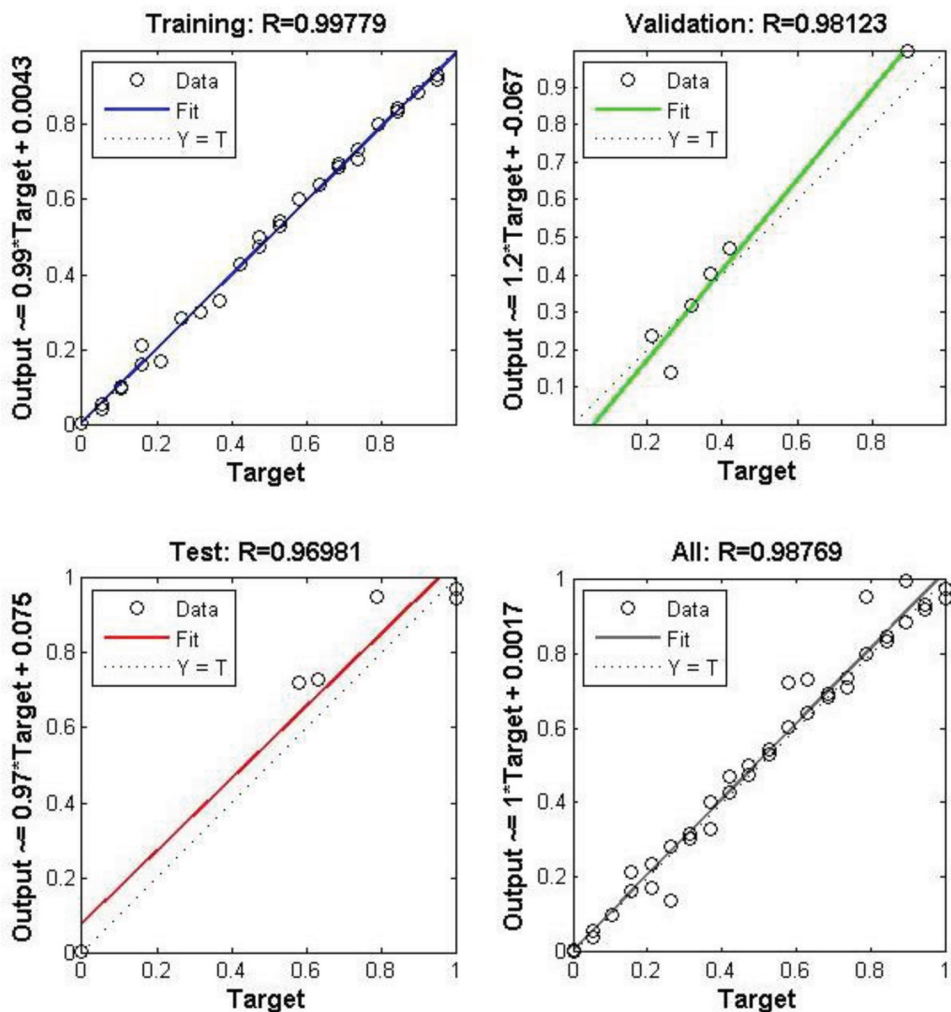


Figure 4. Training, validation, test, and all values of the model.

value of the determined target values has been obtained as a high validity value of 0.96981. All input regression value has been obtained as 0.98769.

The developed model has been operated with MATLAB, and its performance indicator charts have been presented in Figure 4. According to Table 1, the existing league ranking and the estimated league ranking are similar to each other to a great extent. The values in the team ranking columns have been taken between 0 and 1 range. In compliance with the results obtained from input and output parameters in developed models, it has been seen that the team rankings in the league have been estimated with a high accuracy rate.

Table 1. Estimated league ranking results of the England Premier League soccer teams.

	Real league ranking	Sequential league normalized values	Predicted normalized values	Predicted normalized values (ranked)
1	Manchester City	0	0.007	0.007
2	Manchester United	0.053	0.225	0.045
3	Tottenham Hotspur	0.105	0.368	0.225
4	Liverpool	0.158	0.045	0.314
5	Chelsea FC	0.211	0.314	0.347
6	Arsenal FC	0.263	0.347	0.368
7	Burnley	0.316	0.990	0.376
8	Everton FC	0.368	0.952	0.728
9	Leicester City	0.421	0.998	0.929
10	Newcastle United	0.474	0.376	0.952
11	Crystal Palace	0.526	0.998	0.952
12	AFC Bournemouth	0.579	0.952	0.963
13	West Ham United	0.632	0.728	0.985
14	Watford	0.684	0.929	0.989
15	Brighton & Hove Albion	0.737	0.995	0.990
16	Huddersfield Town	0.789	0.989	0.995
17	Southampton	0.842	0.985	0.995
18	Swansea City	0.895	0.963	0.995
19	Stoke City	0.947	0.995	0.998
20	West Bromwich Albion	1	0.995	0.998

Discussion

This study has been carried out with the aim of estimating the league ranking by analyzing the match statistics belonging to three seasons (2015/2016, 2016/2017, 2017/2018) played in the England Premier League with ANN model. While this model has been created, 1140 matches (three seasons) have been evaluated in terms of seven input parameters in order to determine the most accurate model. In evaluation, the number of shots, stealing the ball, the number of passes, the number of passes before goal scoring, the number of forward passes, the total time of attack, and the number of possessions in the first and the second seasons (2015/2016, 2016/2017) have been determined as input parameters and the league ranking in the third season (2017/2018) as output parameter. According to analysis results carried out with ANN, the England Premier League team ranking has been estimated with more than 99% accuracy in terms of the input parameters mentioned earlier. In soccer, league ranking of the teams has been determined with the total points they gather in a season. A number of parameters like players' abilities, their physical and physiological conditions, injuries, and economic conditions of the club in a season can affect the league ranking. It has been thought that estimating the league ranking of the teams with the developed ANN model can be a good source of information for the team in leading the trainings and transfer policies. Turkey volleyball league ranking has been estimated with the ANN

model by using four input parameters (Tümer and Koçer 2017). This study differs from our study because it is based on volleyball and uses only four parameters and values belonging to 66 matches. It has been stated that more accurate estimations can be done with increasing input parameters in the ANN model (Öztemel 2003). Therefore, it has been thought that making more accurate estimations is directly associated with the wideness of the data set. In literature, there is no study about estimating the league ranking in soccer with ANN, so this has limited the discussion part of our study.

It has been seen that a limited number of research using the ANN model related to analyses of the sports branches is about estimating the competition results. In a study carried out for estimating the competition results with the ANN model, match results have been estimated with more than 90% accuracy by using input parameters like current winning percentages, winning percentage of last three matches, winning percentages of home games, winning percentages of the season 2014, and handicaps in 596 competitions played in the season 2015–2016 American Professional Basketball League (Ayyıldız, 2018). In another study, the soccer match results have been estimated with 85% accuracy by Igiri and Nwachukwu (2014) using the ANN model. It has been seen that there are other result estimations having different accuracy rates in the studies carried out with ANN. Arabzad, Araghi, and Soheil (2014), McCabe and Trevathan (2008), Kahn (2003), Ivankovic et al. (2010) have estimated the seven soccer matches (83%), winning teams in terms of 11 parameters (55–68%), winning teams in NFL in terms of 5 parameters (75%), and two factors affecting the winning in basketball (80%) by using the ANN model, respectively.

It has drawn attention that these studies are related to the match results and are based on fewer competitions compared to our study. In our study, a wide data set consisting of 1140 matches has been analyzed in terms of seven parameters and the league ranking has been estimated with high accuracy. It has been thought that building our study with a wider data set will increase the result reliability of the model.

A lot of accurate data have been reached with classical analysis programs widely used in modern soccer (Baacke 2005). It has been thought that the analyst must have enough information in both analysis and soccer in order to correctly infer from the data. Wrong inferences can cause trainers to make mistakes in their technical and tactical preferences. Fast inferences done by neural networks instead of the personal inferences can be evaluated as a factor increasing the objectivity of the results in a competition analysis carried out with a developed ANN model. Besides, it has been thought that fast inferences done with the ANN model provide opportunities to analyze more matches and even trainings in the leagues having a lot of weekly matches and trainings like the England Premier League.

In conclusion, the England Premier League team ranking has been estimated with over 99% accuracy rate with the developed ANN model. It has been seen that the number of shots, stealing the ball, the number of passes, the number of passes before goal scoring, the number of forward passes, the total time of attack, and the number of possessions parameters are determinant in team ranking at the end of the season. This result can help trainers to make important inferences in determining the game setup and match tactics. We thought that making match analyses in soccer by using the ANN model provides team managers, trainers, athletes, and betting shops fast and objective results.

Funding

This work was supported by the Ömer Halisdemir Üniversitesi [1].

ORCID

Hasan Aka  <http://orcid.org/0000-0003-0603-9478>

Zait Burak Aktuğ  <http://orcid.org/0000-0002-5102-4331>

Faruk Kılıç  <http://orcid.org/0000-0002-9978-1972>

References

- Arabzad, A., M. Araghi, and S. Soheil. 2014. Football match results prediction using artificial neural networks: The case of Iran pro league. *International Journal of Applied Research on Industrial Engineering* 1 (3):159–79.
- Ayyıldız, E. 2018. Estimation of American Basketball League (NBA) match results by artificial neural networks. *Gaziantep University Journal of Sports Science* 3 (1):40–53.
- Baacke, H. 2005. *Voleybol antrenmanı üst düzey takımlar için el kitabı 2*. İstanbul: Çağrı Baskı.
- Baca, A. 2014. *Computer science in sport: Research and practice*. London: Routledge.
- Bartlett, R. 2006. Artificial intelligence in sports biomechanics: New dawn or false hope. *Journal of Sports Science and Medicine* 5 (4):474–79.
- Carling, C., J. Bloomfield, L. Nelsen, and T. Reilly. 2008. The role of motion analysis in elite soccer: Contemporary performance measurement techniques and work rate data. *Sports Medicine* 338 (10):839–62. doi:10.2165/00007256-200838100-00004.
- Carling, C., A. Williams, and T. Reilly. 2005. *The handbook of soccer match analysis*. London: Routledge.
- Haykin, S. 1999. *Neural networks and learning machines*. India: Pearson Prentice Hall.
- Igiri, C. P., and E. O. Nwachukwu. 2014. An improved prediction system for football a match result. *IOSR Journal of Engineering* 4 (12):12–20. doi:10.9790/3021-04124012020.
- Ivankovic, Z., M. Rackovic, B. Markoski, D. Radosav, and M. Ivankovic. 2010. Analysis of basketball games using neural networks. Computational Intelligence and Informatics (CINTI) 11th International Symposium, 251–56. Obuda University Budapest, Hungary.
- Kahn, J. 2003. *Neural network prediction of NFL football games*. World Wide Web Electronic Publication, 9–15.

- Laird, P., and L. Waters. 2008. Eyewitness recollection of sport coaches. *International Journal of Performance Analysis in Sport* 8 (1):76–84. doi:10.1080/24748668.2008.11868424.
- Lutz, R. 2015. *Fantasy football prediction*. Cornell University. arXiv:1505.061140.
- Maslovat, D., and I. M. Franks. 2008. The need for feedback. In *The essentials of performance analysis: An introduction*, ed. M. Hughes and I. M. Franks, 1–8. London: Routledge.
- McCabe, A., and J. Trevathan. 2008. Artificial intelligence in sports prediction. information technology: New generations, 2008. ITNG 2008 Fifth International Conference, 1194–97. Las Vegas.
- Menet, F., P. Berthier, M. Gagnon, and J. M. Fernandez. 2020. Spartan networks: Self-feature-squeezing neural networks for increased robustness in adversarial settings. *Computers & Security* 88:1–17. doi:10.1016/j.cose.2019.05.014.
- Öztemel, E. 2003. *Yapay sinir ağları*. Türkiye: Papatya Yayınevi.
- Salman, M. S., O. Kukrer, and A. Hocanin. 2017. Recursive inverse algorithm: Mean-square-error analysis. *Digital Signal Processing* 66:10–17. doi:10.1016/j.dsp.2017.04.001.
- Schöllhorn, W., J. Jäger, and D. Janssen. 2008. Artificial neural network models of sports motions. In *Routledge handbook of biomechanics and human movement science*, ed. Y. Hong and R. Bartlett, 50–64. London: Routledge.
- Sözen, A., E. Arcaklioğlu, and M. Özkaymak. 2005. Turkey's net energy consumption. *Applied Energy* 81 (2):209–21. doi:10.1016/j.apenergy.2004.07.001.
- Tümer, A. E., and S. Koçer. 2017. Prediction of team league's rankings in volleyball by artificial neural network method. *International Journal of Performance Analysis in Sport* 17 (3):202–11. doi:10.1080/24748668.2017.1331570.