Turkish Journal of Sport and Exercise /
Türk Spor ve Egzersiz Dergisi
http://dergipark.gov.tr/tsed
Year: 2022 - Volume: 24 - Issue: 1 - Pages: 22-29
10.15314/tsed.932501

Turkish Journal of
Sport and Exercise
Türk Spor ve Egzersiz Dergisi
ISSN: 2147-5652
https://dergipark.org.tr/tr/pub/tsed

The Estimation of German Football League (Bundesliga) Team Ranking via Artificial Neural Network Model

Zait Burak AKTUĞ[1A], Serkan İBİŞ[1B], Hasan AKA[1C], Faruk KILIÇ[2D]

[1] *Niğde Ömer Halisdemir University, Faculty of Sport Science, Niğde, Turkey*
[2] *Gazi University, Department of Machine, Technical Sciences Vocational College, Ankara, Turkey*

Address Correspondence to Z. B. Aktuğ: e-mail: zaitburak@gmail.com
(Received): 04/5/2021 (Accepted): 15.04.2022
A:Orcid ID: 0000-0002-5102-4331 B:Orcid ID: 0000-0002-5154-3086 C:Orcid ID: 0000-0003-0603-9478 D:Orcid ID: 0000-0002-9978-1972

Abstract

This study was conducted to estimate the places of teams in league ranking by the analysis of the time intervals of the scored and conceded goals in football using Artificial Neural Network (ANN). In the study, the data of the minutes of the scored and conceded goals (0-15, 16-30, 31-45, 46-60, 61-75, 76-90) in total 918 matches played in 3 seasons (2015/2016, 2016/2017, 2017/2018) in German Soccer League (Bundesliga) were used. Total 12 input values (scored and conceded goals) and 1 output (league ranking) value were obtained. 4 different models were determined. 3 seasons league rankings were estimated by training the first 2 season data. All data were separated randomly for training and testing. League ranking was obtained by normalizing between the range of 0,1 – 0,9. Since the produced value was in the range of 0 – 1, it was multiplied by 100 for a trained network and the league ranking was obtained. It was determined that the model developed according to our findings estimated the league ranking with above 99% accuracy for many teams (test data set) according to the minutes of the scored and conceded goals. The lowest mean square error value was obtained as 0.00004. Consequently, it was determined that the minutes of scored and conceded goals in soccer affect the league ranking of the teams. Obtained ANN prediction model can be a guide for coaches to determine the offensive and defensive organizations.

Keywords: Artificial neural network, ranking, prediction, soccer league

Yapay Sinir Ağları Modeli ile Almanya Futbol Ligi (Bundesliga) Takım Sıralamasının Tahmini

Özet

Bu çalışma futbolda atılan ve yenilen gollerin zaman aralığının Yapay Sinir Ağları (YSA) modeli ile analiz edilerek takımların lig sıralamasındaki yerinin tahmin edilmesi amacıyla yapılmıştır. Çalışma Almanya Futbol Ligi'nde (Bundesliga) oynanmış 3 sezonda (2015/2016, 2016/2017, 2017/2018) toplam 918 maçta atılan ve yenilen gol dakikalarına (0-15, 16-30, 31-45, 46-60, 61-75, 76-90) ait veriler kullanılmıştır. Toplam 12 girdi (atılan-yenilen goller) değerine karşılık 1 çıktı (lig sıralaması) değeri elde edilmiştir. Birbirinden farklı 4 model belirlenmiştir. Takımların ilk 2 sezon verileri eğitilerek 3. sezon lig sıralaması tahmin edilmiştir. Verilerinin tamamı eğitim ve test için rastgele yöntemle ayrılmıştır. Lig sırlaması 0,1 - 0,9 aralığında normalize edilerek lig sıralaması elde edilmiştir. Üretilen değer 0 – 1 aralığında olduğundan eğitilen bir ağ için 100 ile çarpılarak lig sıralaması elde edilmiştir. Bulgularımıza göre geliştirilen modelin atılan ve yenilen gol dakikaları değişkenine göre birçok takım için (test veri kümesi) lig sıralamasını %99'un üzerinde doğruluk oranıyla tahmin ettiği belirlenmiştir. En düşük ortalama kare hatası değeri 0.000044 olarak elde edilmiştir. Sonuç olarak futbolda atılan ve yenilen gol dakikalarının, takımların lig sıralamasını etkilediği tespit edilmiştir. Elde edilen YSA tahmin yöntemi antrenörlerin hücum ve savunma organizasyonlarını belirlemede yol gösterici olabilir.

Anahtar kelimeler: Yapay sinir ağları, futbol, sıralama, tahmin

INTRODUCTION

In order to be successful in senior soccer, most of the players should reach physical, technical, and tactical skills (22). To be able to adapt to the competitions with high intensity for an elite player is only possible through a controlled training process (9). In this process, analyzing matches and the league to increase productivity is necessary for a good team ranking in the league (6, 20). The importance of competition analysis applications is gradually increasing due to their positive contributions to improve team performance using different methods and techniques. Today, with the analysis programs used by an analyst in the technical team in senior teams, match analyses of both home and opposing teams are performed and used in technical and tactical evaluations (8, 19). Coaches can reach a great number of accurate data by using these programs (5). Another program in analyzing competitions is the ANN model. The artificial neural network model is used in many fields including sports (13, 17, 23). Artificial neural networks are computer systems that can automatically apply the talents such as learning a skill, which is an important characteristic of the human brain, producing and discovering new information without getting any help (18). Artificial neural networks, which are implemented simulating electronic circuits or computer software, have talents such as collecting information after a learning process and storing it by means of intercellular connections (21). This model provides to reveal the mutual relations that are unknown and difficult to be understood between the data. To be able to train ANN and to reach the targeted results, intensive input and output sequences are needed. Analysis, generalization, association, optimization, learning, and classification in different issues can be performed by ANN (18). The artificial neural network model has found a wide range of application areas due to its superiorities such as presenting easy solutions to complex problems by means of its talents of being able to learn simple structures special to the problem, to do parallel processing, and to tolerate mistakes (21).

A great number of data can be reached by classical analysis programs used in soccer. However, time and experience are required to interpret this data accurately. A competition and league analysis to be performed by ANN model can decrease the effect of human factor especially in the interpretation of the data, and thus decrease the rate of error. The number of studies on the prediction of league rankings in football using the ANN model is limited (2, 14). There is no study to predict the end of season team rankings of the German football league (Bundesliga). It is thought that estimating the league ranking accurately by using the ANN model in soccer will provide important advantages for coaches, clubs, and betting shops. In this regard, the objective of this study is to estimate the league ranking according to the minutes of the scored and conceded goals in the German soccer league (Bundesliga) by using the ANN model.

MATERIAL AND METHOD

In this study, a machine learning method was used to estimate the team ranking. To develop the model, the ANN model in MATLAB (Neural Network Toolbox) software was used. In the study, the data of the minutes of the scored and conceded goals in 3 seasons (2015/2016, 2016/2017, 2017/2018) in the German soccer league (Bundesliga) were used. The match data of the German soccer league used in the study were obtained from an international analysis company.

Model Parameters

Competitions in football are played as 2 halves of 45 minutes. Total 12 input variables were determined as both scored and conceded goals in the first and second halves of the matches.

Input Parameters

Scored goal variable; The time of scoring goal in the matches played by the teams were evaluated in 15-minute sections. Accordingly, scored goal input variable was separated into 6 sections as 0-15, 16-30, 31-45, 46-60, 61-75, 76-90.

Conceded goal variable; The time of conceding a goal in the matches played by the teams were evaluated in 15-minute sections. Accordingly, conceded goal input variable was separated into 6 sections as 0-15, 16-30, 31-45, 46-60, 61-75, 76-90.

Normalization

In the study, all inputs and outputs were subjected to normalization treatment, and the highest value was 0,9 and the lowest value was 0,1 (17). The formula used in normalization is stated in the equation.

x' Normalized value.

x Initial value.

max (x) Maximum value.

min (x) Minimum value.

$$x' = 0.8\ (\frac{x-\min(x)}{\max(x)-\min(x)}) + 0.1 \quad (23).$$

Artificial Neural Network and Modeling

A neuron is the mechanism deciding whether transmits the information or not depending on the threshold value by evaluating the input value arriving to it (Figure. 1). Neurons can decide by communicating with each other (16).

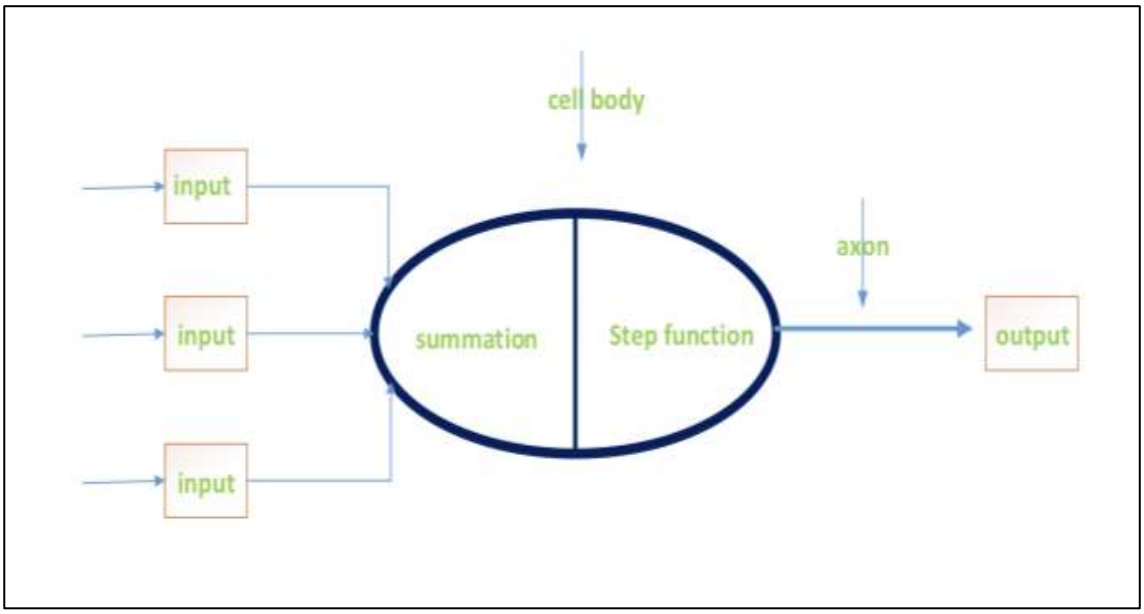


Figure 1. The simple neuron model

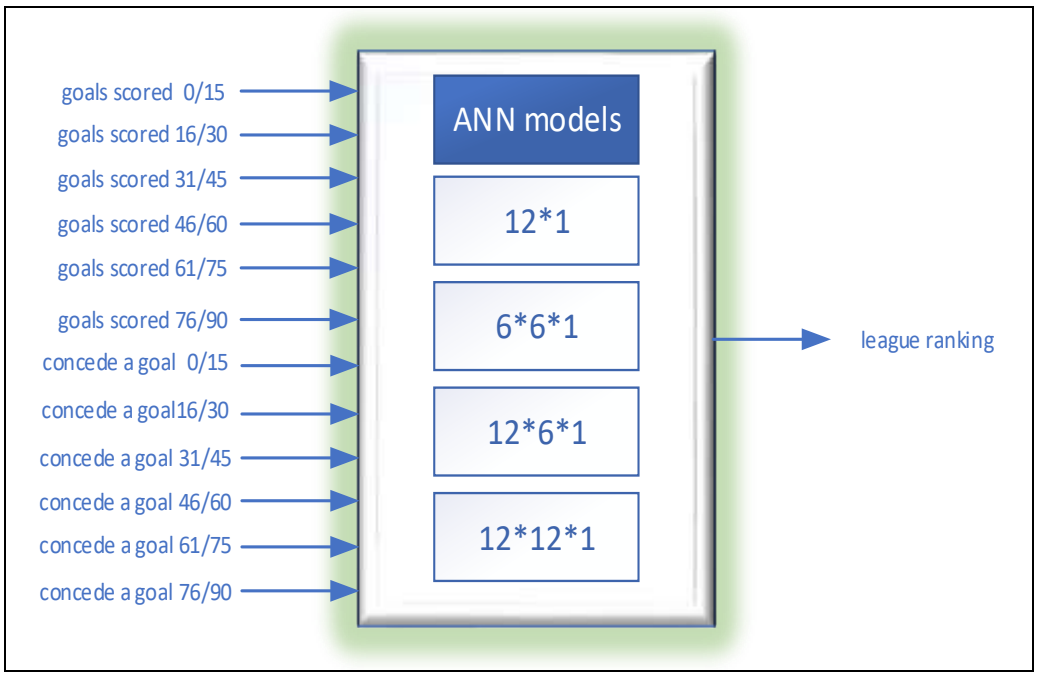


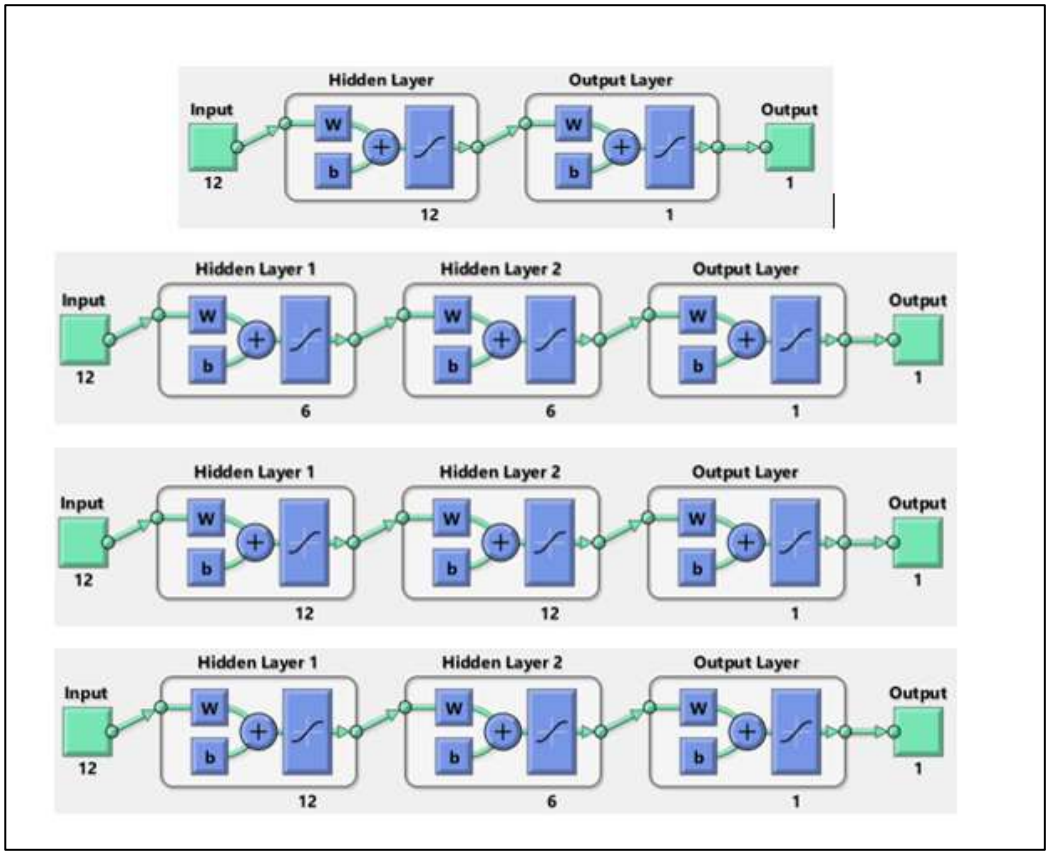
Figure 2. MATLAB artificial neural network models and input, output variables

While the inputs from the entrance level are being transmitted to neurons, there is a weight on the transmission line. Weight factors also work between neurons and inputs. Each input is transmitted to

neurons by being multiplied with weight values. Weights between inputs and neurons are related to the learning process. Neurons having the features of memory and learning create a model between input and output by establishing the network. The created models are designed so that one output will be obtained for twelve inputs.

FINDINGS

The special neural network view of the operated models is given in figure 3. In the models, sequencing was made as input, hidden layer, and output in order from left to right. Models with different layers have 12 input and 1 output values.



The network characteristics of the four established models were chosen as the same. Feed-forward backdrop was operated as the network type. In the networking screen, TRAINGDX was chosen as the training function and the LEARNGDM as the Adoption learning function.

Table 1. Training of best validation performances of models

Model type	Number of layers	MSE
12*1	2	0.00017632
6*6*1	3	0.00053017
12*6*1	3	0.00023022
12*12*1	3	0.000044188

MSE: Mean Square Error

The fewer the mean square error values of the four established models are the less error occurs. In table 1, the smallest error value was obtained as 4.4188e-05 in 12*12*1 model.

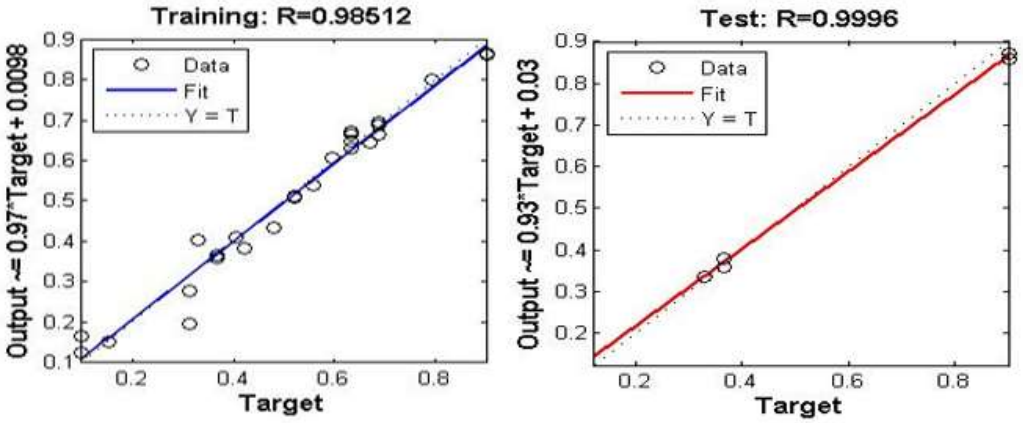


Figure 4. Regression analysis for 12*1 model.

According to Figure 4, the training chart value of the regression analysis performed with 12*1 model type with 2 layers was determined as 0,9851, and the test chart value as 0,9996.

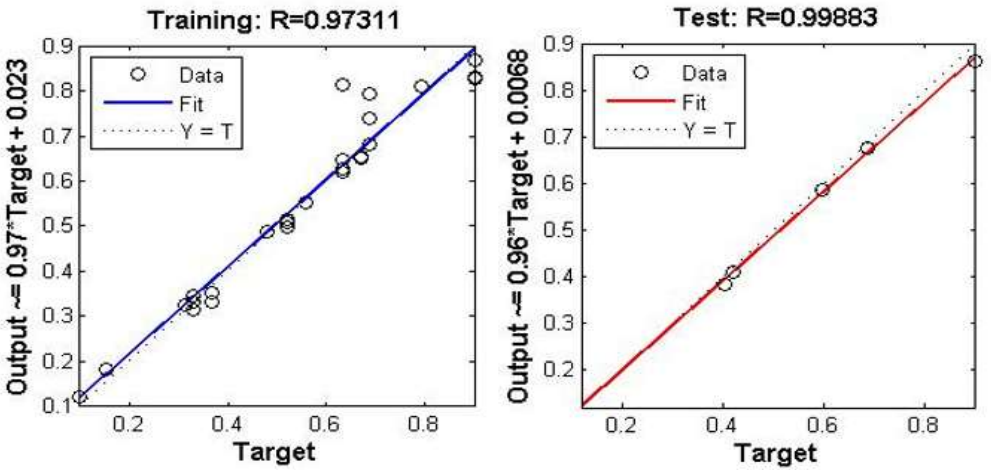


Figure 5. Regression analysis for 6*6*1 model.

According to Figure 5, the training chart of the regression analysis performed with 6*6*1 model type was determined as 0,9731, and the test chart as 0,9988.

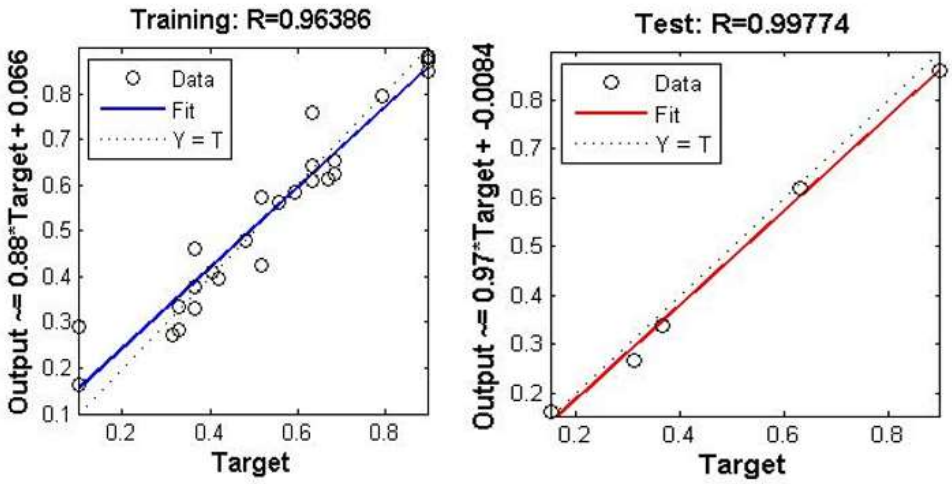


Figure 6. Regression analysis for 12*6*1 model.

According to Figure 6, the training chart of the regression analysis performed with the 12*6*1 model type with 3 layers was determined as 0,9638, and the test chart as 0,9977.

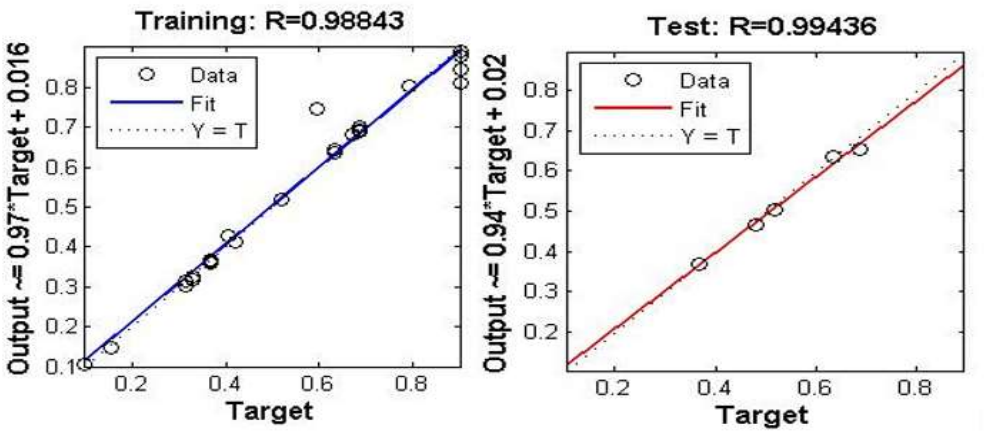


Figure 7. Regression analysis for 12*12*1 model

According to Figure 7, the training chart of the regression analysis performed with the 12*12*1 model type with 3 layers was determined as 0,9884, and the test chart as 0,9943.

Table 2. All regression values

Model type	Training	Validation	Test	All
12*1	0.9851	0.9987	0.999	0.989
6*6*1	0.9731	0.9994	0.9988	0.9808
12*6*1	0.9638	0.9966	0.9977	0.9714
12*12*1	0.9884	0.9993	0.9943	0.9897

In table 2, the four skills of the four models are given separately. The fact that the validity rates of all models are above 99% is remarkable. When compared to other models, the test value of the 12*6*1, of which the training value is the lowest, is 0.99774. It is determined that this value is also close to the other values and it is above 99%.

Table 3. Predicted league ranking results of the German League football teams

Teams	League Ranking	Model 12*1	Model 6*6*1	Model 12*6*1	Model 12*12*1
1	0,1	0,1227	0,1192	0,1644	0,1069
2	0,1470	0,5274	0,5077	0,5238	0,5024
3	0,1941	0,3369	0,3446	0,3227	0,3191
4	0,2411	0,4095	0,3845	0,4117	0,4269
5	0,2882	0,3387	0,3308	0,2819	0,3261
6	0,3352	0,5112	0,4995	0,4245	0,5203
7	0,3823	0,6725	0,8145	0,7598	0,6361
8	0,4294	0,3659	0,3653	0,3799	0,3686
9	0,4764	0,4329	0,4867	0,4817	0,4672
10	0,5235	0,5406	0,5528	0,5619	0,5505
11	0,5705	0,4019	0,3154	0,3345	0,3245
12	0,6176	0,6085	0,5861	0,5841	0,7454
13	0,6647	0,3596	0,3315	0,3298	0,3626
14	0,7117	0,6883	0,6531	0,6596	0,6777
15	0,7588	0,5085	0,5165	0,5762	0,5172
16	0,8058	0,6326	0,6469	0,6425	0,6368
17	0,8529	0,8794	0,8613	0,8793	0,8779
18	0,9	0,6431	0,6534	0,6137	0,6815

The developed model was created by MATLAB, and performance indicator charts (Figure 4, Figure 5, Figure 6, and Figure 7) are presented. The values in the team ranking column are taken in the range of 0,1 – 0,9. It is seen that the teams in the league ranking were estimated with high accuracy (above 99%) according to the results obtained from the input and output values in the developed models.

DISCUSSION

This research was conducted to estimate the league ranking by analyzing the statistics of the matches played in the German (Bundesliga) soccer league for 3 seasons (2015/2016, 2016/2017, 2017/2018) by using the ANN model. While this model was being created, 918 matches (3 seasons) were evaluated according to 12 variables (scored and conceded goals) to determine the most accurate model. In the evaluation, the minutes of the scored and conceded goals (0-15, 16-30, 31-45, 46-60, 61-75, 76-90) in the 1st and 2nd seasons (2015/2016, 2016/2017) were predicted as input variables, and the league ranking in the 3rd season (2017/2018) as the output variable. According to the analysis results by the ANN model, the Bundesliga league ranking was estimated with above 99% accuracy. In a study conducted by Tümer & Koçer (24), Turkish volleyball league ranking was predicted with 98% accuracy by using the ANN model. The input variables used in their study were determined as the teams' winning – losing the game at home and winning – losing the game away, and the output variable as the league ranking. Although this study had a different design from our study, only 66 games were analyzed, input variables were different and it was for the estimation of volleyball league ranking, it supports our study since it predicted the league ranking with high accuracy by ANN model.

In the literature, the number of studies on predicting the league rankings in football with the ANN model is limited. Aka et al., (2) English football league (over 99%); Aka et al., (1) Turkish Super League (over 99%); Kılıç et al., (14) the 2020 Super League (over 94%) predicted the end-of-season team rankings with ANN models developed according to different input variables. It is seen that other studies for the analysis of sport branches by using the ANN model are quite limited in number, and they are for the prediction of the match results. In a study by Ayyıldız (4), game results were tried to be predicted by ANN model by using 596 competition data in the National Basketball Association (NBA) in the 2015-2016 season. In this study, current winning percentages, last 3 games winning percentage, the percentage of winning at home, winning percentages of the year 2014, and handicaps were used for both the host and the guest teams as input variables. The game result was determined as the output variable of the study. According to the result of the study conducted by Ayyıldız (4), NBA games were predicted with above 90% accuracy by the ANN model. In a similar study conducted by Igiri & Nwachukwu (10), it is stated that match results in football were predicted with 85% accuracy by the ANN model. In another study conducted for football, the data of the past 7 weeks were analyzed for the estimation of the 1-week match results in the Iran Pro League. As a result of the study, the match results of 5 out of 6 teams were estimated accurately (83%) (3). In different studies conducted to estimate match results by ANN model, estimations with lower rates were determined (11,12,15).

In above-mentioned studies, the results were estimated by analyzing a few teams and matches. However, in our study, the league ranking was estimated by analyzing a great number of matches such as 918. The fact that the accuracy rate in our study was higher than the above-mentioned studies is thought to be related to the high number of samples used in match analysis.

Since there are many variables affecting the league ranking of the teams in football, predicting the league ranking is quite difficult. It is thought that the economic potential of the teams, physiological and psychological states of the players, the competence of the technical team, material used and the state of the pitch, climate etc. can affect the rankings of the teams at the end of the season. In our study, the goal variable was taken into consideration while estimating the league ranking, and, as a result, it was predicted with a high rate of accuracy. Especially the prediction of the German football league, which is one of the strongest football leagues in Europe, by using ANN model with a high accuracy rate can bring a different point of view to competition analysis in football. It is possible to reach quite a lot of data in the competition analyses performed by the classical analysis systems that are widely used today. However, the high number of data may cause analysis and evaluation times to extend. In addition, personal opinions and suggestions of the expert performing the analysis are used in the inferences obtained from classical analysis programs. This situation can limit the objectivity of the inferences despite the reliable data in the analyses. Also,

mistakes in the inferences by analysts may be the reason for the losses of time, labor, and even performance. In classical performance analysis systems, evaluations can be done according to an annual plan such as before the season, in the middle of the season, and after the season. Thus, the situation can be evaluated by generalizing the whole season (7). In the analyses performed by the ANN model, the fact that inferences are made by the neural network can be considered as a factor increasing the objectivity of the evaluations. In addition, with increasing the evaluation frequencies, it can be ensured that coaches can reach faster and more reliable inferences about their own teams and the opponents.

Consequently, the minutes of the scored and conceded goals in the matches played in 3 seasons in the German (Bundesliga) Soccer League, which is one of the strongest leagues of European soccer was analyzed by the ANN model, and the team ranking was estimated with 99% accuracy. It confirms that the minutes of scored and conceded goals are determinant in ranking together with the other factors affecting the ranking in teams' reaching the targeted ranking in football. This result can be a guide for determining offensive and defensive organizations for coaches.

REFERENCES

1. Aka H, Aktuğ, ZB, Kılıç F. Türkiye Süper Lig sezon sonu takım sıralamasının geliştirilen yapay sinir ağları modeli ile tahmin edilmesi. Spor ve Performans Araştırmaları Dergisi, 2020; 11(3): 258-268.
2. Aka H, Aktuğ ZB, Kılıç F. Estimating the England Premier League ranking with artificial neural network. Applied Artificial Intelligence, 2021; 35(5): 393-402.
3. Arabzad A, Araghi M, Soheil S. Football match results prediction using artificial neural networks: the case of Iran Pro League. International Journal of Applied Research on Industrial Engineering, 2014; 1(3): 159-179.
4. Ayyıldız E. Estimation of American Basketball League (NBA) match results by artificial neural networks. Gaziantep University Journal of Sports Science, 2018; 3(1): 40-53.
5. Baacke H. Voleybol Antrenmanı Üst Düzey Takımlar İçin El Kitabı 2. İstanbul: Çağrı Baskı, 2005.
6. Brito de Souza D, López-Del Campo R, Blanco-Pita H, Resta R, Del Coso J. An extensive comparative analysis of successful and unsuccessful football teams in La Liga. Frontiers in Psychology, 2019; 10(25661): 1-8.
7. Carling C, Williams A, Reilly T. Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance. New York, USA: Routledge, 2007: 164.
8. Fernandez-Echeverria C, Mesquita I, González-Silva J, Claver F, Moreno MP. Match analysis within the coaching process: a critical tool to improve coach efficacy. International Journal of Performance Analysis in Sport, 2017; 17(1-2): 149-163.
9. Harper DJ, Carling C, Kiely J. High-intensity acceleration and deceleration demands in elite team sports competitive match play: a systematic review and meta-analysis of observational studies. Sports Medicine, 2019; 49(12): 1923-1947.
10. Igiri CP, Nwachukwu EO. An improved prediction system for football a match result. IOSR Journal of Engineering, 2014; 4: 12-20.
11. Ivankovic Z, Rackovic M, Markoski B, Radosav D, Ivankovic M. Analysis of basketball games using neural networks. In Computational Intelligence and Informatics (CINTI) 11th International Symposium, Obuda University Budapest, Hungary, 2010: 251-256.
12. Kahn J. Neural Network Prediction of NFL Football Games. World Wide Web Electronic Publication. 2003.
13. Karaatli M, Helvacioğlu ÖC, Ömürbek N, Tokgöz G. Yapay sinir ağları yöntemi ile otomobil satış tahmini. Uluslararası Yönetim İktisat ve İşletme Dergisi, 2012; 8(17): 87-100.
14. Kılıç F, Aka H, Aktuğ ZB. Futbolda yapay sinir ağları modeli ile lig sıralaması tahmini. International Journal of Contemporary Educational Studies, 2020; 6(2): 379-391.
15. McCabe A, Trevathan J. Artificial intelligence in sports prediction. In information technology: New generations, 2008. ITNG 2008 Fifth International Conference, Las Vegas, 2008: 1194-1197.
16. Menet F, Berthier P, Gagnon M, Fernandez JM. Spartan networks: Self-feature-squeezing neural networks for increased robustness in adversarial settings. Computers & Security, 2020; 88: 1-17.
17. Özden S, Kılıç F. Performance evaluation of GSA, SOS, ABC and ANN algorithms on linear and quadratic modelling of eggplant drying kinetic. Food Science and Technology, 2020; 40(3): 635-643.
18. Öztemel E. Yapay Sinir Ağları. Türkiye: Papatya Yayınevi, 2003.
19. Palao J, Hernández-Hernández E. Game statistical system and criteria used by Spanish volleyball coaches. International Journal of Performance Analysis in Sport, 2014; 4(2): 564-573.
20. Rampinini E, Impellizzeri FM, Castagna C, Coutts AJ, Wisløff U. Technical performance during soccer matches of the Italian Serie A League: effect of fatigue and competitive level. Journal of Science and Medicine in Sport, 2009; 12(1): 227-233.
21. Sağıroğlu Ş, Beşdok E, Erler M. Mühendislikte Yapay Zeka Uygulamaları - 1: Yapay Sinir Ağları. Kayseri: Ufuk Kitap Kırtasiye–Yayıncılık Tic. Ltd. Şti, 2003: 299-426.
22. Sarmento H, Anguera MT, Pereira A, Araújo D. Talent identification and development in male football: a systematic review. Sports Medicine, 2018; 48(4): 907-931.
23. Sözen A, Arcaklioğlu E, Özkaymak M. Turkey's net energy consumption. Applied Energy, 2005; 81(2): 209-221.
24. Tümer AE, Koçer S. Prediction of team league's rankings in volleyball by artificial neural network method. International Journal of Performance Analysis in Sport, 2017; 17(3): 202-211.