Football Player Rating DATASET

Mümine KAYA KELEŞ, Serkan ÖZKAYA

[1]*Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi, Mühendislik Fakültesi,* Bilgisayar *Mühendisliği Bölümü, Adana*

[2] *Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi, Mühendislik Fakültesi,* Bilgisayar *Mühendisliği Bölümü, Adana*

Manuscript Title

Abstract

In many real-life sports games, spectators are interested in predicting the outcomes and watching the games to verify their predictions. Traditional approaches include subjective prediction, objective prediction, and simple statistical methods. These estimates depend on the quality, ability and stability of the players. For these predictions to be the most reliable, we need to know the quality of footballers. The most obvious feature of a footballer's quality is his salary. The more stable and skilled a football player is, the more he gets paid and plays for good teams.So in this study we will class out according to the quality level of the footballers for reliable guessing.

1.INTRODUCTION

Rankings of soccer players and data-driven evaluations of their performance are becoming more and more central in the soccer industry. On the one hand, many sports companies, websites and television broadcasters, such as Opta, WhoScored.com and Sky, as well as the plethora of online platforms for fantasy football and e-sports, widely use soccer statistics to compare the performance of professional players, with the purpose of increasing fan engagement via critical analyses, insights and scoring patterns. On the other hand, coaches and team managers are interested in analytic tools to support tactical analysis and monitor the quality of their players during individual matches or entire seasons. Not least, soccer scouts are continuously looking for data-driven tools to improve the retrieval of talented players with desired characteristics, based on evaluation criteria that take into account the complexity and the multi-dimensional nature of soccer performance. While selecting talents on the entire space of soccer players is unfeasible for humans as it is too much time consuming, data-driven performance scores could help in selecting a small subset of the best players who meet specific constraints or show some pattern in their performance, thus allowing scouts and clubs to analyze a larger set of players thus saving considerable time and economic resources, while broadening scouting operations and career opportunities of talented players.The aim of this study is to estimate the quality level of a player according to 9 different characteristics. According to the quality level found in this, the football player is included in a class and has the characteristics of that class. Our goal is to get the most reliable information when evaluating a football team, predicting the outcome of a match or predicting the status of a football player.

1. LITERATURE REVIEW

Statistics are very important in football. Many studies have been carried out to obtain this statistical information. Team rankings, footballer rankings, match prediction from statistical information, performance and development of players and teams as a result of statistically informed results.our work will be to classiation of footballers according to their quality.

2. MATERIAL AND METHOD

Material

As Materyel, I used python to pull data from the Internet. I searched the sites where I could pull data on the subject of the assignment and found 5 sites so that I could get the result I wanted in this regard. Three of these sites were not for pulling data because they blocked me during data withdrawal. Of the remaining 2 sites, it chose the site where I could pull the largest number of data in terms of the number of attributes. The name of that site is mackolik.com.After I made the site selection, I went to the coding stage. I first uploaded beatifulsoup and request libraries to the codes. After installing the libraries, I pulled the attributes I wanted from mackolik.com by encoding them according to their id numbers from the page where the characteristics of the footballers were located.I used weka to analyze the data I pulled from the Internet.

Data Set

I used 10 attributes in my data set. These attributes are as follows: Player's Name,Footballer's Home Country,Footballer's Salary,Footballer's Age,Number of Seasons Played by Footballer,Number of Matches Played by Footballer,Number of goals played by footballer,Total number of minutes played by footballer,Number of games played by footballer in the first 11. I analyzed with a total of 22433 data.

WEKA

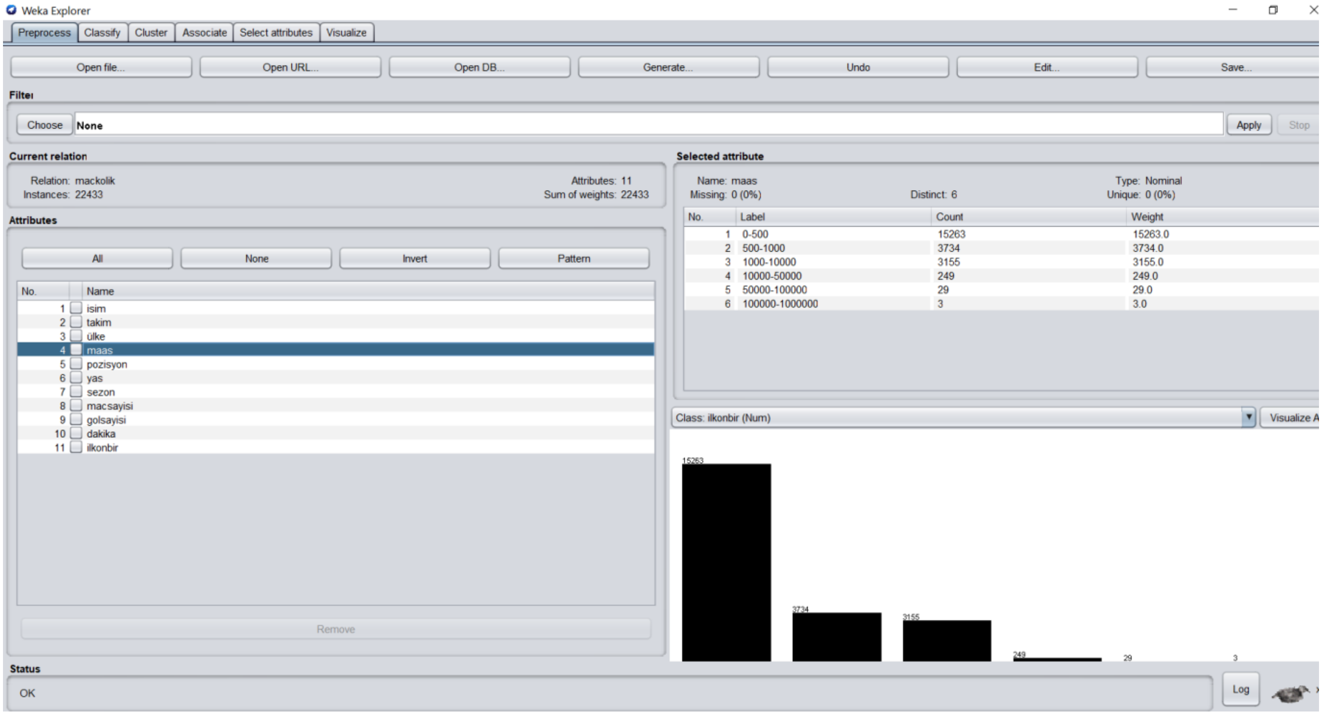

Figure 1. WEKA Tool

No.	1: isim Nominal	2: takim Nominal	3: ülke Nominal	4: maas Nominal	5: pozisyon Nominal	6: yas Nominal	7: sezon Numeric	8: macsayisi Numeric	9: golsayisi Nominal	10: dakika Numeric	11: **ilkonbir** Numeric
1	Ali Cansun	Leventspor	Türkiye	0-500	Forvet	30-40	13.0	184.0	0-100	9135.0	91.0
2	Eser Yaðmur	Kartalspor	Türkiye	0-500	Forvet	30-40	11.0	207.0	0-100	10971.0	113.0
3	Onur Okan Topaloðlu	Büyükçekmece Bld.	Türkiye	0-500	Kaleci	30-40	13.0	190.0	0-100	16958.0	189.0
4	Serdar Özkan	Bursaspor	Türkiye	0-500	Ortasaha	30-40	18.0	403.0	0-100	24466.0	269.0
5	Baykal Yýlmaztaþ	Büyükçekmece Bld.	Türkiye	0-500	Defans	30-40	10.0	126.0	0-100	9547.0	107.0
6	Cafercan Aksu	Adýyaman 1954 SK	Türkiye	0-500	Ortasaha	30-40	18.0	422.0	100-200	30963.0	347.0
7	Faruk Atalay	Bucaspor	Türkiye	0-500	Ortasaha	30-40	9.0	152.0	0-100	8579.0	90.0
8	Ýlhan Çelikay	Bayrampaþa	Türkiye	0-500	Defans	30-40	9.0	157.0	0-100	12181.0	137.0
9	Ýlker Erbay	Büyükçekmece Bld.	Türkiye	0-500	Defans	30-40	15.0	259.0	0-100	20173.0	227.0
10	Ýrfan Baþaran	Kocaelispor	Türkiye	0-500	Ortasaha	30-40	13.0	296.0	0-100	18558.0	209.0
11	Kemal Özcan	Pozantýspor	Türkiye	0-500	Ortasaha	30-40	3.0	29.0	0-100	627.0	6.0
12	Uður Tülümen	Sivas Belediyespor	Türkiye	0-500	Forvet	30-40	13.0	267.0	0-100	16878.0	178.0
13	Sinan Turan	Kozan Spor FK	Türkiye	0-500	Ortasaha	30-40	15.0	293.0	0-100	19797.0	229.0
14	Bekir Küçükertaþ	Kepez Belediyespor	Türkiye	0-500	Kaleci	30-40	12.0	160.0	0-100	13934.0	155.0
15	Can Arat	Sarýyer	Türkiye	0-500	Defans	30-40	15.0	197.0	0-100	15554.0	174.0
16	Dündar Denizhan	Yataðanspor	Türkiye	0-500	Ortasaha	30-40	15.0	360.0	0-100	23775.0	276.0
17	Erman Yýldýrým	Torbalýspor	Türkiye	0-500	Defans	30-40	12.0	196.0	0-100	14848.0	178.0
18	Fahri Tatan	Þanlýurfaspor	Türkiye	0-500	Ortasaha	30-40	16.0	394.0	0-100	26877.0	305.0
19	Hüsnü Zeybekoðlu	Tavas Birlikspor	Türkiye	0-500	Ortasaha	30-40	15.0	326.0	0-100	19820.0	233.0
20	Mithat Yaþar	Turgutluspor	Türkiye	0-500	Ortasaha	30-40	16.0	414.0	0-100	27737.0	314.0
21	Serhat Akýn	Berliner AK 07	Türkiye	0-500	Forvet	30-40	14.0	274.0	0-100	17217.0	191.0
22	Recep Biler	Giresunspor	Türkiye	0-500	Kaleci	30-40	13.0	183.0	0-100	15825.0	176.0
23	Semih Þentürk	Eskiþehirspor	Türkiye	0-500	Forvet	30-40	18.0	438.0	100-200	21159.0	218.0
24	Serkan Sütlü	Beylikdüzüspor	Türkiye	0-500	Forvet	30-40	11.0	172.0	0-100	9753.0	104.0
25	Serkan Silan	Vefa	Türkiye	0-500	Forvet	30-40	17.0	411.0	0-100	29993.0	331.0
26	Süleyman Aydýn	Antalya Kemerspor	Türkiye	0-500	Defans	30-40	7.0	102.0	0-100	8357.0	95.0
27	Tolga Altun	Batman Petrolspor	Türkiye	0-500	Ortasaha	30-40	8.0	155.0	0-100	11657.0	134.0
28	Ulaç Çaðlayan	Dersimspor	Türkiye	0-500	Defans	30-40	13.0	228.0	0-100	18235.0	204.0
29	Ali Þen Kandil	Faroz Yalýspor	Türkiye	0-500	Ortasaha	30-40	11.0	157.0	0-100	9377.0	109.0
30	Cemre Atmaca	Çankaya FK	Türkiye	0-500	Forvet	30-40	18.0	397.0	100-200	26219.0	291.0
31	Ergin Keleþ	Altay	Türkiye	0-500	Forvet	30-40	16.0	363.0	0-100	22717.0	242.0
32	Feridun Sungur	Haçka Spor	Türkiye	0-500	Defans	30-40	16.0	245.0	0-100	18953.0	209.0
33	Mustafa Yalçýnkaya	Turgutluspor	Türkiye	0-500	Defans	30-40	11.0	195.0	0-100	14089.0	149.0
34	Osman Kurtuldu	52 Orduspor FK	Türkiye	0-500	Kaleci	30-40	16.0	222.0	0-100	19820.0	218.0

Figure 2. WEKA Tool

Method

This topic aims to classige the quality of footballers and estimate which class they should belong to according to the characteristics of the attributes of a new future footballer. Therefore, this problem is a classification algorithm problem. Attributes in accordance with the classification algorithm were created and some numeric values were categorized into nominal values.7 attributes are Nominal and 3 are numeric values.

Data Mining

The highest percentage of accuracy was observed in SimpleCart, BFTree, DTNB, J48, IBk algorithms. Except for the accuracy percentage of the algorithms, precision, recall, roc area, rmse values were analyzed and the algorithm that should be used for this subject was analyzed. After observing all the values, the most suitable algorithm for this problem was selected as IBk.

Data Preprocessing Methods

Unnecessary attributes were removed to classify. Noisy data was detected and organized in the data set.

Classification Algorithms/Clustering Algorithms/Regression Algorithms/Association Rule Mining Algorithms

The classification algorithm was used because the problem was classification and categorize footballers by quality.

```
=== Confusion Matrix ===

     a     b     c     d     e     f   <-- classified as
 13674  1097   491     1     0     0 |     a = 0-500
  2483   722   529     0     0     0 |     b = 500-1000
  1513   572  1034    34     2     0 |     c = 1000-10000
    59    26   127    34     3     0 |     d = 10000-50000
     3     1    11    13     1     0 |     e = 50000-100000
     1     0     0     0     2     0 |     f = 100000-1000000
```

Figure 3-Weka IBK Confusion Matrix

3. RESULTS AND DISCUSSION

Algorithm	Accuracy	Precision	Recall	F-Measure	ROC Area	RMSE
Bayes Net	66.0991 %	?	0,661	?	0,831	98.5021 %
Bayesian Logistic Regression	-	-	-	-	-	-
Naïve Bayes	68,4081 %	0,702	0,684	0,691	0,826	95,627 %
Naive Bayes Multinominal Updateable	68,4081%	0,702	0,684	0,691	0,826	95,627%
Multilayer Perceptron	70.5791 %	?	0,706	?	0,764	91.9344 %
SMO	69.0144%	?	0,690	?	0,555	113.9533 %
Linear Regression	-	-	-	-	-	-
Logistic	70.2314 %	0,602	0,702	0,623	0,767	92.0215%
Simple Logistic	70.1467 %	0,591	0,701	0,621	0,767	92.0445 %
IBk	70,8156 %	0,666	0,687	0,676	0,786	96,3677%
IB1	65,3234 %	0,652	0,653	0,653	0,667	117,0337%
KStar	69,9238%	?	0,699	?	0,794	92,2564%
Bagging	69,429%	?	0,694	?	0,768	94,2564%
Random Committee	68,7469%	0,666	0,687	0,676	0,786	96,3677%
Classification Via Regression	70,8153%	?	0,708	?	0,771	91,3644%
VFI	60.3174 %	0,695	0,603	0,641	0,760	112.3119 %
Hyper Pipes	67,9178%	?	0,716	?	0,706	95,2769%
JRip	70,5657%	?	0,706	?	0,578	96,6911%
Ridor	65,9475	?	0,659	?	0,622	117,9577%
Decision Table	69,7999%	0,602	0,698	0,612	0,757	92,7329%
OneR	67,8821%	?	0,679	?	0,505	114,5579%
ZeroR	68,0382%	?	0,670	?	0,500	100 %
DTNB	72,1081%	0,660	0,721	0,665	0,840	86,8934%
Random Forest	69,7276%	?	0,697	?	0,814	91,7392%
SimpleCart	72,6742%	?	0,727	?	0,790	89,0145%
Random Tree	64,6904%	0,645	0,647	0,6626	0,662	119,7689%
Random Forest	69,7276%	?	0,697	?	0,814	91,7392%
J48	71,591%	?	0,708	?	0,706	95,2756%
Id3	-	-	-	-	-	-
BFTree	72.2864 %	?	0,723	?	0,806	88.8409 %

CONCLUSION

Statistical knowledge and classification of footballers is important for many markets. We have done this study to get the most reliable results about footballers and a lot of statistical information can be obtained based on this study and estimates can be made based on this information. As a result of observations and analyses, the data set and algorithm we selected were observed as the most useful choice for this study.