

# Makine Öğrenmesi ile Spor Karşılaşmalarının Modellenmesi

## Modelling Sports Games Using Machine Learning

Berk Karaoğlu

Bilgisayar Mühendisliği Bölümü

Bilkent Üniversitesi

berk.karaoglu@bilkent.edu.tr

### Özet

*Büyük ve çok değişkenli sistemlerin davranışlarını tahmin etmek bir çok bilim alanında araştırma konusu olmuştur. Sporun tüm dünya tarafından takip edilen bir alan olması sebebiyle, bir spor müsabakasının sonucunu tahmin etmek üzerine yapılan çalışmaların sayısı da artmaktadır. Bir spor müsabakasının sonucu bir çok farklı subjektif değişkene bağlı olsa da, temel olarak takımların ofansif ve defansif yeteneklerine göre şekillenir. Bu çalışmada takımların sezon başından itibaren attığı ve yediği gol ortalaması temel alınan basit bir formül kullanılarak, çeşitli makine öğrenmesi algoritmalarının sadece takımların ofansif ve defansif yeteneklerini dikkate alarak maçın sonucunu ne kadar başarılı tahmin edebildiği incelenmiştir. Avrupa'dan 16 futbol liginde yapılan testler neticesinde DecisionTable çoğunlukla en yüksek başarıyı veren algoritma olmuştur.*

**Anahtar Kelimeler:** Makine öğrenmesi, spor karşılaşmalarının modellenmesi

### Abstract

*Predicting the behaviour of complex and multi-variate systems have always been a study area among different science areas. As sports have a world-wide popularity, the number of studies about predicting the result of sports games increase day by day. Although the result of a sports game depends on many subjective variables, mostly competitors' offensive and defensive skills affect the final result. This study uses a simple formula based on competitors' scored and conceded goal averages since the beginning of the season; and investigates how machine learning algorithms perform, when only teams' offensive and defensive skills are taken into account. According to the tests done using 16 European football leagues, DecisionTable mostly performs the best.*

**Keywords:** Machine learning, modelling sports games

### 1. Giriş

Büyük ve çok değişkenli sistemlerin davranışlarını tahmin etmek bir çok bilim alanında araştırma konusu olmuştur. Yakın gelecekteki borsa ve döviz kurları tahmini, önümüzdeki ayın hava durumu tahmini, güneş ve rüzgâr santrallerinin elektrik üretim miktarının tahmini gibi bir çok örnek verilebilir. Geçmiş verilere dayanarak, henüz gerçekleşmemiş olayların sonucunu tahmin etmek bir çok kişi ve kuruluş için önem arz etmektedir. Kurumsal kâr amacı güden çalışmalar dışında, araştırmacıların

kendi ilgi alanları çerçevesinde yaptıkları çalışmaların sayısı da oldukça fazladır.

Sporun tüm dünya tarafından takip edilen bir alan olması sebebiyle internet üzerinde geçmiş müsabakalara dair veri toplamak, diğer bir çok alana göre çok daha kolaydır. Bu sebeple spor müsabakalarının sonucunun modellenmesi geçmiş akademik çalışmaların bir çoğuna konu olmuştur. Geçmiş çalışmaların çoğu futbola odaklansa da basketbol, amerikan futbolu, tenis, vb. gibi diğer spor dalları için de çalışmalar mevcuttur. Bu çalışma, Avrupa'dan 16 futbol liginin 2013-2014 ve 2014-2015 sezonları süresince oynanan futbol maçlarının sonucunu, çeşitli makine öğrenmesi algoritmaları kullanarak ve sadece takımların ofansif ve defansif güçleri dikkate alındığında ne kadar başarılı tahmin edebildiklerini incelemektedir.

Futbol maç sonuçlarının modellendiği geçmiş çalışmalar temel olarak istatistiksel metodlar ve makine öğrenmesi metodları olarak ikiye ayrılabilir. İstatistiksel metodlar, takımların geçmiş bir zaman dilimi içerisinde (sezon başından itibaren veya son  $n$  hafta içinde) attığı ve yediği gol ortalamaları üzerinden doğrusal regresyon yaparak takımların gelecek maçta kaç gol atması beklendiğini hesaplar.

Bir diğer yöntem ise makine öğrenmesi kullanılmasıdır. Benzer şekilde takımların geçmiş dönemde attığı ve yediği gol ortalamaları ile elde edilen istatistiksel verilere, maç sonucu verisi de eklenerek (Ev Sahibi Galibiyeti-Berberlik-Konuk Takım Galibiyeti) makine öğrenmesi algoritmalarının, hangi durumlarda nasıl bir sonuç elde edildiğini öğrenmesi ile uygulanır. Bu yöntemde, makine öğrenmesi algoritmaları ile takımların ofansif ve defansif güç değerlerinin maç sonucuna nasıl etki ettiği anlaşılır ve bu doğrultuda henüz sonucu bilinmeyen bir maç için ev sahibi galibiyeti, beraberlik ve konuk takım galibiyeti sınıflarından birinin seçimi yapılır.

Makalenin 2. Bölümü'nde spor karşılaşmalarının sonucunu tahmin etme üzerine geliştirilen düşünceler ve geçmiş çalışmalar anlatılmıştır. 3. Bölüm'de model oluşturmada kullanılan veri kümesi tanıtılıp, makine öğrenmesi için kullanılacak özellikler anlatılmıştır. 4. Bölüm'de takımların ofansif ve defansif güç değerlerinin hesaplanması ve bu güç değerleri ile gelecek maçta atmasını beklediğimiz ortalama gol sayısının nasıl hesaplandığı anlatılmaktadır. 5. Bölüm'de her futbol ligi için, çeşitli makine öğrenmesi algoritmaları kullanılarak, sonuçlar karşılaştırılmaktadır. 6. Bölüm'de ise bu çalışma neticesinde elde edilen sonuçların değerlendirilmesi yapılmıştır.

## 2. Geçmiş Çalışmalar

Henüz oynanmamış bir spor müsabakasının sonucunun tahmini hakkında bir çok istatistiksel ve makine öğrenmesi çalışmaları yapılmıştır. Bu çalışmalar, takımların önceki maçlarda attığı ve yediği gol sayıları üzerinden doğrusal regresyon yaparak atacağı ortalama gol sayısını tahmin etmek ve Poisson olasılık dağılımı ile gol olasılıklarını hesaplamaktadır. Bir diğer yöntem ise maç sonucu tahminini bir sınıflandırma problemi olarak değerlendirip, makine öğrenmesi algoritmalarıyla modellemeyi amaçlamaktadır.

Bir futbol maçında mücadele eden takımların attığı gol sayıları Poisson dağılımı gösterir ve bu takımların Poisson değişkenleri bir takımın ofansif, diğer takımın da defansif gücüne bağlıdır [1]. Ancak, takımların ofansif ve defansif güçleri sezon boyunca sabit değildir. Oyuncuların form durumları, moralleri, sakat veya cezalı olmaları, takımın ofansif ve defansif gücüne etki ettiğinden ötürü güncel maçlara daha fazla ağırlık verilerek değerlendirilmesi gerekir [2,3,4].

Spor karşılaşmalarının sonucunu etkileyen bir çok subjektif değişken vardır. Hava durumu, maçın önemi, takımın morali, vb. gibi örnekler verilebilir. Dolayısıyla maç sonucu sadece takımların ofansif ve defansif kabiliyetlerine bağlı değildir. Bu sebeple, Poisson değişkenlerinin hesaplanmasında kullanılan fonksiyon daha karmaşık olmalı ve bu değişkenleri de barındırmalıdır [5, 6, 7].

Takımların geçmiş maçlarda attığı ve yediği gol sayıları üzerinden doğrusal regresyonla gol sayısı tahmin etmeye ek olarak, geliştirilen bir diğer istatistiksel yöntem ise, doğrudan maç sonuçlarını değerlendirmektir. Bu yöntemler, maçları (galibiyet, beraberlik, mağlubiyet) olarak değerlendirip, skorları dikkate almamaktadır [8,9].

Yapılan çalışmalardan biri satranç oyuncularının değerlendirilmesi için geliştirilmiş olan ELO puanlama sistemini (Elo, 1978) futbol alanında uygulayarak bahis şirketlerinin oranlarıyla karşılaştırılmıştır [10].

[12] de İngiltere 1. Lig takımlarından Tottenham Hotspur'un 1995-1997 yılları arasındaki karşılaşmaları Bayes ağları kullanılarak modellenmiş ve konunun uzmanlarının da bilgisi ile oluşturulan uzman Bayes ağı ortalama %59 başarı yakalamıştır. Diğer Bayes ağları ise %40-50 başarıya ulaşabilmişlerdir.

Yapay sinir ağları kullanılarak 2006 FIFA Dünya Kupası veri kümesinde, sonucu berabere biten maçlar çıkarıldığında ev sahibi mi yoksa konuk takımın mı kazanacağı %76.9 başarı ile tahmin edilebilmiştir [11].

Rugby maçlarının yapay sinir ağları ile modellenmiş, ve dört farklı Rugby liginden oluşan veri kümeleri ile yapılan testlerde en yüksek %75, en düşük ise %52 başarı elde edilmiştir [13].

[14] de UEFA Şampiyonlar Ligi'nde oynanmış 96 futbol maçı, çeşitli makine öğrenmesi metodları ile modellenmiştir. Her maç için yaklaşık 30 özelliğin bulunduğu veri kümesi kullanılarak, yaklaşık %60'lık başarı sağlanabilmiştir.

## 3. Veri Kümesi

Çeşitli veri kaynakları içinde, bir çok yerli ve yabancı araştırmada da kullanılan en popüler futbol veri kaynağı football-data.co.uk isimli web sitesidir. Bu web sitesinde 1993-1994 futbol

sezonu başından itibaren geçtiğimiz sezon sonuna kadar 11 Avrupa ülkesinde yaklaşık 20 adet futbol liginde oynanan maçlar ile ilgili veri bulunmaktadır. Bu veriler; mücadele eden iki takımın ismi, maçın oynandığı tarih, maçın ilk yarısı süresince takımların attığı gol sayıları, maçın tamamında takımların attığı gol sayıları, toplam ve kaleyi bulan şut sayıları, korner sayıları, sarı ve kırmızı kart ile cezalandırılan oyuncu sayılarıdır. Ancak, atılan gol sayıları dışındaki istatistikler İngiltere dışındaki ülkelerin maç verilerinde bulunmamaktadır. Veri kümesinin büyük bir çoğunluğunda maç istatistikleri eksik olduğundan dolayı, bu özellikler veri kümesinden çıkarılmıştır. Ayrıca, veri kümesindeki maç tarihlerinin bir çoğunun yanlış olması sebebiyle, bu değerler maçın oynandığı hafta bilgisi ile değiştirilmiştir; çünkü önemli olan maç tarihleri değil, maçların oynandığı sıra bilgisidir.

Bu çalışmada Avrupa'da oynanan 16 futbol ligi değerlendirilme-ye alınmıştır. Makine öğrenmesi için düzenlenen veri kümesinde yalnızca üç adet özellik bulunmaktadır:

- *Maçın sonucu* (ev sahibi galibiyeti için H, beraberlik için D, ve deplasman için A)

- *Ev sahibi takımın atmasını beklediğimiz ortalama gol sayısı*

- *Konuk takımın atmasını beklediğimiz gol sayısı*

Takımların atmasını beklediğimiz ortalama gol sayılarının nasıl hesaplandığı bir sonraki bölümde detaylı açıklanmıştır.

## 4. Takımların Güçlerinin Hesaplanması ve Ortalama Gol Sayısının Belirlenmesi

Güçlü bir takımı, zayıf bir takımdan ayıran özellik daha çok gol atma ve kalesinde daha az görme kabiliyetidir [4]. Dolayısıyla bir futbol maçının sonucunu tahmin eden modeller, mücadele edecek iki takımın arasındaki bu kabiliyet farklarını dikkate almalıdır.

Eğer  $i$  ve  $j$  takımları arasında oynanan bir maçın skoru  $(X_{ij}, Y_{ij})$  ise,  $i$  takımının  $j$  takımına attığı gol sayısı olan  $X_{ij}$ ,  $i$  takımının ofansif gücüne ve  $j$  takımının defansif gücüne bağlıdır. Benzer şekilde,  $j$  takımının  $i$  takımına attığı gol sayısı olan  $Y_{ij}$ ,  $i$  takımının defansif gücüne ve  $j$  takımının ofansif gücüne bağlıdır. Takımların ofansif ve defansif güçlerinin dikkate alındığı fonksiyon ise Poisson dağılımı gösterir [1] ve örnek bir algoritma denklem (1)'de verilmiştir.

$$X_{ij} \sim \text{Poisson}(\text{LOAG} \cdot O_i \cdot D_j) \quad (1)$$

$$Y_{ij} \sim \text{Poisson}(\text{LOAG} \cdot D_i \cdot O_j)$$

LOAG - Ligde bir takımın attığı ortalama gol sayısı

$O_i$  -  $i$  takımının ofansif gücü

$D_i$  -  $i$  takımının defansif gücü

$O_j$  -  $j$  takımının ofansif gücü

$D_j$  -  $j$  takımının defansif gücü

Bir takımın ofansif gücü, mücadele ettiği ligde bir takımın maç başına attığı gol ortalamasından ne kadar az veya fazla gol attığını; defansif gücü ise ligde bir takımın maç başına yediği gol ortalamasından ne kadar az veya fazla gol yediğini gösterir. Takımların ofansif ve defansif güçleri hesaplanırken, öncelikle ligde bir takımın maç başına attığı ortalama gol, denklem (2) kullanılarak hesaplanır.

$$LOAG = TG \div (TS \cdot N) \quad (2)$$

LOAG - Ligde bir takımın attığı ortalama gol

TG - Ligde atılan toplam gol

TS - Takım Sayısı

N - Değerlendirmeye alınan hafta sayısı

Örnek olarak Türkiye Süper Ligi 2014/2015 Sezonu'nun 25. Haftası'nda oynanmış olan Fenerbahçe-Beşiktaş karşılaşmasının tahmin edileceği bir senaryo ele alınırsa, öncelikle ligdeki tüm takımların attığı gol sayıları hesaplanır. Takımların ofansif ve defansif güçleri hakkında güncel bir bilgi elde etmek açısından son 5 haftanın maçları dikkate alındığında (N = 5), ligin 20, 21, 22, 23 ve 24. haftaları boyunca ev sahibi takımlar 71, konuk takımlar ise 60 gol olmak üzere toplam 131 gol (TG) atmışlardır. Türkiye Süper Ligi'nde 18 takımın mücadele ettiği düşünüldüğünde (TS = 18), bir takımın maç başına attığı ortalama gol sayısı 2 numaralı denklem kullanılarak aşağıdaki gibi hesaplanır:

$$LOAG = 131 \div (18 \cdot 5) = 1.46$$

Bu süre zarfında Fenerbahçe kendi sahasında ve deplasmanda olmak üzere toplam 9, maç başına ise 1.8 gol atmıştır. Aynı dönemde Beşiktaş ise toplam 11, maç başına ise 2.2 gol atmıştır.  $i$  takımının ortalama attığı gole  $OAG_i$  denirse,  $i$  takımının ofansif gücü denklem (3) ile hesaplanır:

$$O_i = OAG_i \div LOAG \quad (3)$$

$$D_i = OYG_i \div LOAG$$

$O_i$  -  $i$  takımının ofansif gücü

$OAG_i$  -  $i$  takımının ortalama attığı gol sayısı

LOAG - Ligde bir takımın attığı/yediği ortalama gol

$D_i$  -  $i$  takımının defansif gücü

$OYG_i$  -  $i$  takımının ortalama yediği gol sayısı

Bu durumda;

$$O_{Fenerbahçe} = 1.8 \div 1.46 = 1.23$$

$$O_{Beşiktaş} = 2.2 \div 1.46 = 1.51$$

olmaktadır. Yani Fenerbahçe, ligdeki ortalama bir takımdan %23, Beşiktaş ise %51 daha fazla gol atma gücüne sahiptir.

Benzer bir yöntemle takımların defansif güçlerini de hesaplamak gerekir. Bir takımın, bir maçta attığı gol sayısı rakip takımın da yediği sayısı olduğundan dolayı, ligde atılan toplam gol sayısı ile yenilen toplam gol sayısı eşittir. Bu sebeple, bir takım, kalesinde ortalama 1.46 gol görmüş şeklinde de yorumlanabilir. Fenerbahçe bu 5 haftalık dönemde toplam 5, maç başına ortalama 1 gol kalesinde görmüştür. Beşiktaş ise toplam 6, maç başına ortalama 1.2 gol kalesinde görmüştür. Bu durumda Fenerbahçe ve Beşiktaş'ın defansif güçleri;

$$D_{Fenerbahçe} = 1 \div 1.46 = 0.68$$

$$D_{Beşiktaş} = 1.2 \div 1.46 = 0.82$$

olarak hesaplanır.

Yani son ligin 20. ve 25. haftaları arasında Fenerbahçe lig ortalamasından %32, Beşiktaş ise %18 daha az gol yemiştir.

Denklem (1)'de gösterilen, ve takımların atmasını beklediğimiz gol sayısını veren fonksiyona göre, Fenerbahçe'nin Beşiktaş'a atmasını beklediğimiz gol sayısı ( $X_{Fenerbahçe-Beşiktaş}$ ) ve

Beşiktaş'ın Fenerbahçe'ye atmasını beklediğimiz gol sayısı ( $X_{Beşiktaş-Fenerbahçe}$ ) aşağıdaki gibi hesaplanır:

$$X_{Fenerbahçe-Beşiktaş}$$

$$\sim \text{Poisson} (LOAG \cdot O_{Fenerbahçe} \cdot D_{Beşiktaş})$$

$$\sim \text{Poisson} (1.46 \cdot 1.23 \cdot 0.82)$$

$$\sim \text{Poisson} (1.473)$$

$$X_{Beşiktaş-Fenerbahçe}$$

$$\sim \text{Poisson} (LOAG \cdot O_{Beşiktaş} \cdot D_{Fenerbahçe})$$

$$\sim \text{Poisson} (1.46 \cdot 1.51 \cdot 0.68)$$

$$\sim \text{Poisson} (1.499)$$

Maç esnasında atılan gol sayıları tam sayı olduğundan dolayı bu Poisson ortalama beklenen değerlerinin takımların gol atama, 1 gol atma, 2 gol atma, vb. gibi tam sayı olasılıklarının hesaplanması gerekmektedir. Ancak, bu çalışmada maç skor olasılıklarını hesaplamayıp, bu elde edilen Poisson değişkenlerinin (1.473 ve 1.499) maç sonucunu makine öğrenmesi ile modellerken ne kadar başarılı olacağı test edilecektir.

## 5. Veri Kümesinin Modellenmesi ve Algoritmaların Karşılaştırılması

Makine öğrenmesi, veri kümesinde bulunan özelliklerin sonucu nasıl etkilediğinin anlaşılması ve henüz sonucu belli olmayan bir durumun sınıflandırılması için kullanılmaktadır. Makine öğrenmesi konusunda çok çeşitli algoritmalar geliştirilmiştir. Bu çalışmada veri kümesi Naive Bayes, BayesNet, Multilayer Perceptron, LogitBoost, DecisionTable, ZeroR ve C4.5 algoritmaları ile modellenmiş ve maç sonucunu ev sahibi takım galibiyeti (H), beraberlik (D) ve konuk takım galibiyeti (A) seçenekleri içerisinde doğru tahmin etme başarıları karşılaştırılmıştır.

Naive Bayes algoritması özelliklerin kendi aralarında bağımsız olduğunu varsayarak, hangi özelliğin hangi değere sahipken hangi sınıf değerinin seçildiğini analiz eder. Bayes Net, özellik ve sınıf kümelerini yönlendirilmiş çevrimsiz çizge ile ifade ederek, olası sınıf değerleri için olasılık hesaplar. Multilayer Perceptron bir grup özelliği bir grup sınıf değerine eşler. Girdiler ve çıktılar arasında birden çok katman bulunur ve veri bu ara katmanlarda işlenir. LogitBoost, tahmin başarısını düşüren verilerin veri kümesindeki önemini artırarak, tahmin başarısını iyileştirme amacı güder. DecisionTable, veri kümesinde bulunan özellikler hangi değerlere sahipken, o gözlemin nasıl sınıflandırıldığını anlayarak, henüz sınıflandırılmamış gözlemlere sınıf değeri atar. ZeroR, sadece sınıf değerine odaklanır ve diğer özellikleri dikkate almaz. Test kümesindeki tüm gözlemlere, eğitim kümesinde en çok bulunan sınıf değerini atar. Makine öğrenmesi algoritması olarak kullanılmaz, genelde diğer algoritmaların başarılarını karşılaştırmak için bir kriter olarak ele alınır. C4.5 eğitim kümesindeki gözlemleri kullanarak bir karar ağacı oluşturur.

Her ligin dinamikleri ve yapısı farklı olduğundan dolayı veri kümesi lig bazında bölünmüş olup, her ligin veri kümesi ayrı ayrı işleme alınmıştır. Algoritmalar WEKA uygulamasında standart değişken değerleri kullanılarak test edilmiştir. Takımların ofansif ve defansif güçleri son 5 hafta boyunca gösterdikleri performans dikkate alınarak hesaplanmıştır (N = 5). Avrupa'dan 16 farklı futbol ligi için yapılan testlerin sonuçları Tablo 1'de verilmiştir.

Tablo 1: Sonuçlar

	Naive Bayes	BayesNet	Multilayer Perceptron	LogitBoost	Decision Table	ZeroR	C4.5
İngiltere 1. Ligi	%48.29	%47.03	%49.77	%49.02	%48.18	%46.52	%47.88
İngiltere 2. Ligi	%43.21	%43.57	%42.30	%43.84	%43.94	%43.93	%41.82
İngiltere 3. Ligi	%43.25	%43.97	%43.88	%43.91	%43.97	%43.97	%42.24
İngiltere 4. Ligi	%42.17	%42.75	%41.69	%42.42	%42.75	%42.75	%41.67
Almanya 1. Ligi	%45.68	%45.03	%46.20	%46.92	%46.34	%46.52	%44.42
Almanya 2. Ligi	%44.19	%45.35	%45.89	%45.75	%46.08	%46.08	%44.21
İtalya 1. Ligi	%43.70	%40.70	%45.25	%45.82	%45.68	%46.00	%44.86
İtalya 2. Ligi	%43.73	%44.39	%41.84	%43.82	%44.39	%44.39	%42.81
İspanya 1. Ligi	%46.26	%47.18	%48.61	%48.48	%48.93	%47.82	%48.20
İspanya 2. Ligi	%44.00	%44.33	%43.99	%43.99	%44.17	%44.33	%44.00
Fransa 1. Ligi	%44.85	%46.21	%46.17	%45.71	%46.21	%46.21	%44.40
Fransa 2. Ligi	%45.13	%45.71	%45.71	%45.28	%45.71	%45.71	%44.27
Türkiye 1. Ligi	%47.26	%47.12	%50.22	%50.29	%50.86	%46.65	%50.25
Hollanda 1. Ligi	%48.18	%48.72	%49.65	%49.72	%50.47	%48.53	%49.49
Belçika 1. Ligi	%49.71	%49.32	%52.22	%51.67	%51.54	%47.44	%49.97
Portekiz 1. Ligi	%47.92	%47.58	%49.04	%48.72	%49.87	%45.90	%48.48

Yapılan çalışma neticesinde, Decision Table algoritması 16 ligin 11'inde en yüksek başarı değerini elde etmiştir. ZeroR 8, BayesNet 6, Multilayer Perceptron 3 ve LogitBoost 1 ligde en başarılı sonucu vermiştir. NaiveBayes ve C4.5 algoritmaları ise test edilen 16 futbol liginden hiç birinde en başarılı olamamıştır.

Futbol karşılaşmalarının 3 sınıflı {H,D,A} bir sınıflandırma problemi olarak değerlendirilmesinden, rastgele bir sınıf seçimi ile %33 başarı sağlanabileceği varsayılabilir. Bu durumda, %33'ten daha iyi sınıflandırma yapan algoritmalar teorik olarak başarılı olarak değerlendirilebilir de her lig için ZeroR'dan daha yüksek skor elde eden algoritmaların başarılı kabul edilmesi daha doğru olacaktır. Bu çalışmada ZeroR'dan en az %1 daha iyi sonuç veren algoritmalar başarılı olarak değerlendirilmiştir. Buna göre Multilayer Perceptron İngiltere 1. Ligi'nde %3.25 ve Belçika 1. Ligi'nde %4.78, Decision Table ise İspanya, Türkiye, Hollanda ve Portekiz 1. Liglerinde sırasıyla %1.11, %4.21, %1.94, %3.97 daha başarılı olmuştur.

## 6. Sonuç

Geçmiş gözlemlerin makine öğrenmesi ile modellenmesi ve oluşturulan bu modelin yeni, sınıflandırılmamış gözlemlerin sınıflandırılmasında kullanılması bir çok alanda kullanılmaktadır. Spor sektörünün popülerliği göz önünde bulundurulunca, gelecek maçların sonuçlarının tahmin edilmesi bir çok istatistiksel ve makine öğrenmesi çalışmasına konu olmuştur. İnternet üzerinde kolaylıkla veri kümesi bulunabilmesi sebebiyle bu ça-

lışmada sadece futbol ele alınmıştır; ancak diğer bir çok takım ve bireysel sporlara da uygulanabilmektedir.

Bu çalışmada sadece takımların attıkları ve yedikleri gol ortalamaları üzerinden özellikler oluşturulurken, maç sonucunu etkileyen hava şartları, sakat / cezalı oyuncular, maçın önemi, vb. gibi özellikler eklendiği takdirde modelleme başarısının artabileceği öngörülmektedir. Kullanılan veri kümesinin çok kapsamlı olmasından ötürü, her futbol maçı için bu özelliklerin eklenmesi vakit alacak bir süreç olduğundan, bu çalışmada dikkate alınmayıp, sadece takımların ofansif ve defansif kabiliyetlerinin sonuca etkisinin modellenmesi amaçlanmıştır.

Modelleme sonuçlarına göre, seçilen algoritmalar sınıflandırma başarısı açısından çok büyük farklılık göstermemiştir. En iyi sonuç gösteren ve en kötü sonuç gösteren algoritmalar arasında en fazla %4'lük bir fark görülmüştür. En yüksek sonucun %50-52 seviyesinde olduğu bu durumda, varolan sınıflandırma algoritmalarının futbol veri kümesi için çok uygun olmadığı gibi bir değerlendirme de yapılabilir.

Bir diğer çıkarılabilecek sonuç ise, daha yüksek modelleme başarısı elde edilen liglerin, diğerlerine göre daha tahmin edilebilir olduğudur. İngiltere 1. Ligi'nde futbol maçlarının sonucu %49.77 oranda başarılı tahmin edilebilmişken, İngiltere 2. Ligi'nde bu oran %43.94, İngiltere 3. Ligi'nde %43.97, İngiltere 4. Ligi'nde ise %42.75 olarak hesaplanmıştır. Dolayısıyla, İngiltere 1. Ligi'nde, İngiltere'nin alt liglerine kıyasla daha az sürpriz sonuç çıktığı şeklinde düşünülebilir.

## 7. Kaynaklar

- [1] Maher, M. J. (1982), Modelling association football scores. *Statistica Neerlandica*, 36: 109-118.
- [2] 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)*, 46: 265-280.
- [3] 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)*, 51: 157-168.
- [4] Karlis, D. and Ntzoufras, I. (2011). Robust fitting of football prediction models. *IMA Journal of Management Mathematics*, 22, 171-182.
- [5] 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)*, 49: 399-418.
- [6] Karlis D. and Ntzoufras J. (2000). On modelling soccer data. *Student 3*, 229-245.
- [7] Karlis, D. and Ntzoufras, J. (2003) Analysis of sports data Using bivariate Poisson models. *The Statistician* 52, 381-393.
- [8] Knorr-Held (2000) Dynamic Rating of Sports Teams. *The Statistician* 49, 261-276.
- [9] Koning, R. H. (2000), Balance in Competition in Dutch Soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49: 419-431.
- [10] Hvattum, L.M, Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of forecasting*, Volume 26, Issue 3, 460-470.
- [11] Joseph, A., Fenton, N.E., Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems*. 19, 544-553.
- [12] Huang, K.Y. (2010). A Neural Network Method for Prediction of 2006 World Cup Football Game. *The 2010 International Joint Conference on Neural Networks*, 1-8.
- [13] McCabe, A., Trevathan, J. (2008). Artificial Intelligence in Sports Prediction. *Fifth International Conference on Information Technology: New Generations*, 1194-1197.
- [14] Hucaljuk, J., Rakipovic, A. (2011). Predicting football scores using machine learning techniques. *MIPRO 2011*, 1623-1627.



### **Berk Karaoğlu**

1990 yılında Ankara’da doğmuştur. İlk, orta ve lise öğrenimini Türkiye Eğitim Derneği (TED) Ankara Koleji’nde tamamladıktan sonra, 2011 yılında İzmir Ekonomi Üniversitesi Yazılım Mühendisliği Bölümü’nden şeref öğrencisi olarak mezun olmuştur. 2010 yılından beri farklı sektörlerde faaliyet gösteren bir çok özel şirket için VoIP, VOD, e-ticaret, vb. projelerde uzman yazılım mühendisi ve mobil uygulama geliştiricisi olarak görev almıştır. Ayrıca, Türkiye’de bir çok devlet kurumunun mobil uygulamalarının geliştirilmesinde katkıda bulunmuştur. 2011’den beri Bilkent Üniversitesi Bilgisayar Mühendisliği Bölümü’nde yüksek lisans yapmakta olup, başlıca ilgi alanları; makine öğrenmesi, bilgi erişim sistemleri, veri madenciliğidir. Halen veri madenciliği ve makine öğrenmesi ile çeşitli rüzgâr ve güneş enerjisi üretim sistemlerinin, elektrik üretim tahmini konusu üzerinde çalışmaktadır.