



Makine Öğrenmesi Yöntemleriyle Euroleague Basketbol Maç Sonuçlarının Tahmin Edilmesi ve Maç Sonuçları Üzerinde En Etkili Değişkenlerin Bulunması

Predicting Euroleague Basketball Match Outcomes With Machine Learning Techniques and Revealing the Most Important Game Related Variables

Erhan ÇENE¹

¹Yıldız Teknik Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 34210, İstanbul
• ecene@yildiz.edu.tr • ORCID > 0000-0001-5336-6004

Makale Bilgisi / Article Information

Makale Türü / Article Types: Arařtırma Makalesi / Research Article

Geliş Tarihi / Received: 6 Temmuz / July 2021

Kabul Tarihi / Accepted: 19 Ocak / January 2022

Yıl / Year: 2022 | **Cilt - Volume:** 13 | **Sayı - Issue:** 1 | **Sayfa / Pages:** 31-54

Atf/Cite as: Çene, E. "Makine Öğrenmesi Yöntemleriyle Euroleague Basketbol Maç Sonuçlarının Tahmin Edilmesi ve Maç Sonuçları Üzerinde En Etkili Değişkenlerin Bulunması - Predicting Euroleague Basketball Match Outcomes with Machine Learning Techniques and Revealing the Most Important Game Related Variables: Ondokuz Mayıs Üniversitesi Spor ve Performans Arařtırmaları Dergisi - Ondokuz Mayıs University Journal Of Sports And Performance Researches 13(1), April 2022: 31-54

MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE EUROLEAGUE BASKETBOL MAÇ SONUÇLARININ TAHMİN EDİLMESİ VE MAÇ SONUÇLARI ÜZERİNDE EN ETKİLİ DEĞİŞKENLERİN BULUNMASI

ÖZ

Bu çalışmada 2016-2017 ile 2020-2021 yılları arasında oynanan 1358 EuroLeague basketbol maçlarındaki takım istatistikleri göz önüne alınmış ve bu takım istatistiklerinden hangilerinin maçın galibi üzerinde en çok etkiye sahip olduğu belirlenmeye çalışılmıştır. Maçlar, k-ortalama kümeleme analizi sonucunun belirttiği skor farklarına göre yakın, dengeli ve dengeli olmayan olmak üzere üç gruba ayrılmıştır. Hem bu üç grup hem de tüm maçlara k en yakın komşuluk, naive bayes, lojistik regresyon, destek vektör makinaları, karar ağacı, rastgele orman ve yapay sinir ağları algoritmaları uygulanmış ve en etkili algoritmalar lojistik regresyon, destek vektör makineleri ve yapay sinir ağları olarak bulunmuştur. Bu üç algoritma maç sonucunu tüm maçlar için yaklaşık %84 oranında doğru bilmiştir. Yakın maçlarda bu oran %79 a düşmüş, dengeli maçlarda %97 e, dengeli olmayan maçlarda %100 e çıkmıştır. Maç sonucu üzerinde en çok etkili olan değişkenler savunma ribaundu, gerçek şut yüzdesi, top çalma, top kaybı, hücum ribaundu ve denenen serbest atıştır. Burada bulunan sonuçlar takımların maç içi stratejilerini belirlerken en çok odaklanmaları gereken konular üzerinde fikir vermekte ve bu konular üzerine yoğunlaşarak strateji belirlemelerine yardımcı olacağı düşünülmektedir.

Anahtar Kelimeler: Basketbol, makine öğrenmesi, maç sonucu tahmini, EuroLeague.

PREDICTING EUROLEAGUE BASKETBALL MATCH OUTCOMES WITH MACHINE LEARNING TECHNIQUES AND REVEALING THE MOST IMPORTANT GAME RELATED VARIABLES

ABSTRACT

In this study, team statistics in 1358 EuroLeague basketball matches played between 2016-2017 and 2020-2021 seasons were taken into account and it was tried to determine which of these team statistics had the most impact on the winner of the match. The matches were divided into three groups as close, balanced and unbalanced games according to the score differences indicated by the k-means cluster analysis result. K nearest neighbor, naive bayes, logistic regression, support vector machines, decision tree, random forest and artificial neural network algorithms were applied to both these three groups and all matches, and the most effective algorithms were found to be logistic regression, support vector machines and artificial neural networks. These three algorithms can predict correctly the match result for all matches with approximately 84% accuracy. This rate decreased

to 79% in close matches, increased to 97% in balanced matches and to 100% in unbalanced matches. The variables that have the most influence on the outcome of the match are defensive rebounds, true shooting percentage, steals, turnovers, offensive rebounds and free throw attempts. The results give an idea on the issues that the teams should focus on while determining their in-match strategies and help them determine their strategy by focusing on these issues.

Keywords: *Basketball, machine learning, match outcome prediction, EuroLeague.*

GİRİŞ

Verilerin toplanma, işleme ve analiz edilme hızı ve gücü gelişen bilgisayar sistemlerine bağlı olarak son yıllarda büyük bir artış göstermiştir. Bu artışa paralel olarak spor kulüpleri ellerindeki veriyi daha verimli bir şekilde kullanma arayışına girmiş, bu da istatistiğin sporla iç içe girmesini ve spor analitiği kavramının ortaya çıkmasını sağlamıştır. Spor analitiğinin en ilgi çekici konularından biri de maç sonuçlarının başarılı bir şekilde tahmin edilmesi ve maç sonuçlarını etkileyen maç içi etmenlerin doğru bir şekilde belirlenmesidir. Maç sonuçlarını etkileyen etmenlerin önceden belirlenebilmesi takımların buna göre hazırlık yapmasını ve buna göre taktik belirlemesini sağlayacaktır.

Çeşitli spor dalları için, maç sonucunu tahmin etmek için makine öğrenmesi yöntemlerinden sıklıkla faydalanılmaktadır. At yarışı sonuçlarını (Davoodi ve Khanteymooori, 2010), Amerikan futbolu, ragbi ve futbol sonuçlarını (McCabe ve Trevathan, 2008; Tax ve Joustra, 2015), ve golf sonuçlarını (Wiseman, 2016) makine öğrenmesi yöntemleriyle tahmin eden çalışmalar mevcuttur.

Benzer şekilde en çok ilgi çeken sporlardan biri olan basketbolda da (McComb, 2004) maç sonucunu etkileyen etmenleri belirlemeyi ya da maç sonucunu tahmin etmeyi amaçlayan çalışmalar mevcuttur. İspanya Basketbol Ligi'nde oynanan 870 maç için yapılan bir çalışmada (Ibáñez ve ark., 2008), kazanan ve kaybeden takımların maç içi istatistikleri diskriminant analizi ile karşılaştırılmış ve galibiyette en çok etkili olan etmenlerin asist, top çalma ve blok olduğu belirlenmiştir. Kurulan modelin maç sonuçlarını doğru sınıflama oranı %82,4 olsa da veri eğitim ve test verisi olarak ayrılmadan model uygulandığı için sonuç yanıltıcı olabilir.

Maçları kümeleme analizi yardımıyla maç skoru farklarına göre sınıflara ayıran pek çok çalışma mevcuttur (Csataljay ve ark., 2009; Lorenzo ve ark., 2010; Garcia ve ark., 2014; Çene, 2018). Bu çalışmalar her bir grup için maç sonucunu etkileyen etmenleri bağımsız örneklem t-testi, Wilcoxon sıra işaret testi, diskriminant analizi ve karar ağaçları gibi farklı istatistiksel tekniklerle açıklamaya çalışmışlardır. Her bir çalışma farklı maç grubu ve farklı örneklem büyüklükleriyle çalışmış ve maç sonucu üzerinde etkili olan etmenlerin farklılık gösterdiği gözlemlenmiştir.

2004-2016 yılları arasındaki 156 olimpiik basketbol ma verisi zerine yapılan bir alıřmada (Leicht ve ark., 2017), lojistik regresyon ve karar aaları yntemleri uygulanmıřtır. Lojistik regresyon asist, savunma ribaundu, řut yzdesi, yapılan ve alınan foul, top alma ve top kaybının ma sonucunu etkileyen etmenler olduėunu tahmin ederken, karar aacı ise řut yzdesi ve savunma ribauntlarının en etkili deėiřkenler olduėunu tahmin etmiřtir. Lojistik regresyon modeli %85.5, karar aacı modeli ise %81.4 oranında ma sonularını doėru sınıflama yzdesine sahiptir. Ancak bu alıřmada da veri seti eėitim ve test verisi olarak ikiye ayrılmamıř, modeller tm veriye uygulanmıřtır.

Bir yksek lisans tezinde (Jones, 2016), NBA malarını hangi takımın kazanacaėı iki takım arasındaki skor farkı tahmin edilerek bulunmaya alıřılmıřtır. Bunun iin 2008-2011 yılları arasında NBA de oynanmıř 144 ma iin skor farkı En Kk Kareler (EKK) regresyonu ve lojistik regresyon ile modellenmiřtir. Saha ii řut yzdesi, 3 sayı řut yzdesi, serbest atıř řut yzdesi, hcum ribaundu, asist, top kaybı ve isabetli serbest atıř sayısının takım skorları farkını modellemek iin etkili olduėu bulunmuř ve ardından etkili olan deėiřkenler gz nne alınarak 2013-2016 yılları arasındaki malar iin tahminde bulunulmuřtur. Modelin doėru sınıflandırma yzdesi %62 olarak elde edilmiřtir.

Bazı alıřmalar ise hangi deėiřkenlerin ma sonucu zerinde etkili olduėuna odaklanmak yerine doėrudan ma sonucunu sınıflama bařarısı zerine yoėunlařmıřtır. Yapay sinir aėları (Loeffelholz ve ark., 2009; Valenzuela R., 2018; Thabtah ve ark., 2019; Horvat ve ark., 2020; Ozkan, 2020) doėrusal regresyon (Magel ve Unruh, 2013), lojistik regresyon (Magel ve Unruh, 2013; Valenzuela R., 2018), karar destek makineleri (Pai ve ark., 2017; Kaur ve Jain, 2017; Valenzuela R., 2018), karar aaları (Pai, ChangLiao ve Lin, 2017; Thabtah, Zhang ve Abdelhamid, 2019; Horvat, Havař ve Srpak, 2020), Naive Bayes (Valenzuela R., 2018; Thabtah, Zhang ve Abdelhamid, 2019; Horvat, Havař ve Srpak, 2020), Rastgele Orman (Valenzuela R., 2018; Horvat ve ark., 2020) ma sonucunu tahmin etmek iin kullanılan makine ėrenmesi tekniklerinden bazılarıdır.

Literatrdeki alıřmaların kullandıkları ynteme, rnekleme ve ulařtıkları sınıflama bařarısına dair daha detaylı bilgiye Tablo – 1 den ulařılabilir. Tablo – 1 den grldėu gibi bu alanda yapılan alıřmaların hemen hepsinin NBA, uluslararası basketbol turnuvaları ya da yerel ligler kaynaklı olduėu grlmektedir. Oysaki basketbolun NBA dıřındaki en rekabeti organizasyonu olan EuroLeague hakkında yapılmıř alıřmalara pek rastlanmamaktadır.

Bu alıřmanın amacı gemiřte oynanmıř EuroLeague malarındaki farklı takım istatistiklerini kullanarak hem ma sonucunu tahmin eden bir makine ėrenme modeli kurmak, hem de ma sonucunu tahmin etmede etkili olan ma ii istatistikleri belirlemektir.

YÖNTEM

Araştırma Grubu (Evren-Örneklem)

Bu çalışmanın verisini, 2016-2017 sezonundan 2020-2021 sezonunun sonuna kadar oynanan EuroLeague basketbol maçlarına ait takım istatistikleri oluşturmaktadır. Bu sezonların seçilme sebebi, EuroLeague'in 2016-2017 yılından itibaren kapalı bir lig haline gelmesidir. 2016-2017 sezonuna kadar takımlar gruplara ayrılarak maçlar yaparken, bu sezondan itibaren tüm takımlar tek bir grupta toplanmış ve her takımın birbiriyle maç yapmasına olanak sağlayan bir lig haline gelmiştir. Veri seti 5 sezona yayılmış toplam 1358 maçtan oluşmaktadır.

Takım istatistiklerinin elde edilmesi: 1358 maça ait veriler, EuroLeague resmi web sayfasında (www.euroleague.net), her bir maça ait bilgi sayfasından elde edilmiştir. Veriler elde edilirken Python programlama dilinin beautifulsoup paketinden faydalanılmıştır.

Verilerin Toplanması/İşlem Yolu: Resmi maç istatistikleri sınırlı sayıda maç içi istatistiği sağlamaktadır. Veriler R programlama diline aktarılarak yeni maç içi istatistikleri oluşturulmuştur. Bu çalışmada kullanılan maç içi istatistiklere, kısaltmalarına ve tanımlarına Tablo – 2 den ulaşılabilir.

Verilerin Analizi: Bu çalışmada maç sonuçlarını tahmin etmek ve maç sonuçlarını etkileyen etmenleri belirleyebilmek için birkaç farklı yöntem kullanılmıştır. İlk önce, çalışmada kullanılan maç içi istatistiklere ait betimsel istatistiklere yer verilmiştir. Ardından değişkenlerin normal dağılım gösterip göstermediğine Kolmogorov-Smirnov normallik testiyle bakılmış, değişkenlerin normal dağılmadığı gözlemlendiğinden, Mann-Whitney-U testi kullanılarak hangi değişkenlerin kazanma ve kaybetme üzerinde etkili olabilecekleri belirlenmiştir.

Bir sonraki aşamada farklı makine öğrenmesi yöntemleri veri seti üzerine uygulanarak maç sonucunu en iyi tahmin eden modeller belirlenmiştir. Bu çalışmada kullanılan makine öğrenmesi yöntemleri k-en yakın komşuluk, Naive Bayes, lojistik regresyon, destek vektör makineleri, karar ağaçları, rastgele orman ve yapay sinir ağlarıdır. Veri seti öncelikle %70 eğitim, %30 test verisine bölünmüş ve ardından maç sonucu bağımlı değişken olacak şekilde, her bir makine öğrenmesi yöntemi için model, eğitim verisi üzerine 5 katmanlı çapraz doğrulama içerecek şekilde 10 kez tekrarlanarak kurulmuş ve modelin başarısı bu yapı üzerinden ölçülmüştür. Oluşturulan modelin başarısı, test verisi üzerinde sınanmış ve her bir model için doğru sınıflama yüzdesi ve diğer başarı ölçütleri raporlanmıştır.

Tablo 1: Literatür Özeti

Çalışma	Veri Seti	n	Yöntemler	Başarı	En Etkili Değişkenler
Ibanez vd. (2008)	İspanyol Basketbol Ligi (2001-2006)	870	Diskriminant Analizi	%82,4	Asist; Top Çalma; Blok
Csataljay vd. (2009)	Avrupa Basketbol Şampiyonası (2007)	54	Kümeleme Analizi, Wilcoxon İşaret Sıra Testi	--	Başarılı 3 Sayılık Atış; Şut Yüzdesi; Başarılı Serbest Atış; Serbest Atış Yüzdesi; Savunma Ribaundu
Loeffelholz vd. (2009)	NB (2007-2008)	620	Yapay Sinir Ağları	%83,0	--
Lorenzo vd. (2010)	16 Yaş Altı Avrupa Şampiyonaları (2004-2005)	122	Kümeleme Analizi, Bağımsız Örneklem t-testi	--	Top Kaybı; Assist; 2 Sayılık Atış; Savunma Ribaundu
Magel ve Unruh (2013)	NCAA (2009-2012)	380	Doğrusal Regresyon; Lojistik Regresyon	%68,0	Serbest Atış Sayısı; Savunma Ribaundu; Assist; Top Kaybı
Garcia vd. (2014)	İspanyol Basketbol Ligi (2007-2008)	306	Kümeleme Analizi; Diskriminant Analizi	--	Asist; Başarılı 2 Sayılık Atış; Savunma Ribaundu; Top Çalma; 3 Sayılık Atış
Jones (2016)	NBA (2008-2011)	144	Doğrusal Regresyon; Lojistik Regresyon	%62,0	Şut Yüzdesi; 3 Sayı Şut Yüzdesi; Serbest Atış Şut Yüzdesi; Hücum Ribaundu; Asist; Top Kaybı; Başarılı Serbest Atış Sayısı
Pai vd. (2016)	NBA (2008-2010)	400	Karar Ağaçları; Yapay Sinir Ağları	%85,0	2 Sayılık Şut Yüzdesi; 3 Sayılık Şut Yüzdesi; Serbest Atış; Savunma Ribaundu; Top Çalma; Asist

Leicht vd. (2017)	Olimpiyat Oyunları (2004-2016)	156	Karar Ağaçları; Lojistik Regresyon	%85,5	Asist; Savunma Ribauntı; Şut Yüzdesi; Yapılan ve Alınan Faul; Top Çalma; Top Kaybı
Kaur ve Jain (2017)	NBA (2015-2016)	800	Karar Destek Makineleri	%89,0	Savunma ve Hücum Reytingi; Sayı; Gerçek Şut Yüzdesi; Efektif Şut Yüzdesi; Ribaunt Yüzdesi
Çene (2018)	EuroLeague (2016-2017)	259	Kümeleme Analizi; Bağımsız Örneklem t-testi; Karar Ağacı	--	Gerçek Şut Yüzdesi; Top Çalma; Yapılan Fauller; Başarılı 2 Sayılık Atış; Başarılı 3 Sayılık Atış; Savunma Ribaundu
Valenzuela (2018)	NBA (2006-2017)	14280	Lojistik Regresyon; Naive Bayes; Karar Destek Makinesi; Yapay Sinir Ağları; Rastgele Orman	%73,29	--
Thabtah vd. (2019)	NBA Final Serileri (1980-2017)	430	Naive Bayes; Yapay Sinir Ağları; Karar Ağaçları	%83,0	--
Horvat vd. (2020)	NBA (2009-2018)	11578	Lojistik Regresyon; Naive Bayes; Karar Ağaçları; Yapay Sinir Ağları; Rastgele Orman	%59,0	--
Huang ve Lin (2020)	NBA - Golden State Warriors (2017-2018)	82	Regresyon Ağacı	%87,0	--
Özkan (2020)	Turkish Basketball League (2015-2016)	240	Yapay Sinir Ağları	%79,2	--

Tablo 2: Değişken Kısaltmaları, İsimleri ve Tanımları

Değişken Kısaltması	Değişken Adı	Tanımı
Maç_Sonuç	Maç Sonucu	Maç sonucu Galibiyet ya da Mağlubiyet değerlerinden birini alır. Çalışmanın bağımlı değişkenidir.
Ev_Depl	Ev Sahibi / Deplasman	Ev sahibi ya da Deplasman değerlerinden birini alır.
Hüc.Rib	Hücum Ribaundu	Kaçırılan bir şut sonucunda hücum yapan takımın aldığı ribaund sayısı.
Sav.Rib	Savunma Ribaundu	Kaçırılan bir şut sonucunda savunma yapan takımın aldığı ribaund sayısı.
Sayı.Pası	Sayı Pası	Sayı yapacak bir oyuncuya verilen pas sayısı.
Top.Çalma	Top Çalma	Bir pası keserek ya da rakip oyuncunun kontrolünde olan bir topu kaparak pozisyon kazanma
Top.Kaybı	Top Kaybı	Hatalı yürüme, hatalı pas, rakibe top çaldırma ya da hücum faul neticesinde pozisyonun rakibe geçmesi.
Yapılan.Blok	Yapılan Blok	Rakibin şutuna dokunarak şutun kaçmasına sebep olan müdahale.
Alınan.Blok	Alınan Blok	Kendi takımındaki bir oyuncunun şutuna yapılan blok sayısı
Yapılan.Faul	Yapılan Faul	Oyuncuların yaptığı toplam kural ihlali sayısı
Alınan.Faul	Alınan Faul	Rakip takımın yaptığı toplam kural ihlali sayısı
Sayı.2.Başarılı	Başarılı 2 Sayılık Atış Sayısı	Başarılı 2 Sayılık Atış Sayısı
Sayı.2.Denenen	Denenen 2 Sayılık Atış Sayısı	Başarılı ve Başarısız 2 Sayılık Atış Sayısının toplamı
Sayı.3.Başarılı	Başarılı 3 Sayılık Atış Sayısı	Başarılı 3 Sayılık Atış Sayısı
Sayı.3.Denenen	Denenen 3 Sayılık Atış Sayısı	Başarılı ve Başarısız 3 Sayılık Atış Sayısının toplamı
Serb.Atış.Başarılı	Başarılı Serbest Atış Sayısı	Başarılı Serbest Atış Sayısı
Serb.Atış.Denenen	Denenen Serbest Atış Sayısı	Başarılı ve Başarısız Serbest Atış Sayısının toplamı

Saha.İçi.Başarılı	Başarılı Saha İçi İsbet Sayısı	Başarılı 2 Sayılık Atış Sayısı + Başarılı 3 Sayılık Atış Sayısı
Saha.İçi.Denenen	Denenen Saha İçi İsbet Sayısı	Denenen 2 Sayılık Atış Sayısı + Denenen 3 Sayılık Atış Sayısı
Sayı.Pası.Top.Kaybı.Oranı	Sayı Pası Top Kaybı Oranı	Sayı Pası / Top Kaybı
Sayı.2.%	2 Sayılık Atış Yüzdesi	Başarılı 2 Sayılık Atış Sayısı / Denenen 2 Sayılık Atış Sayısı
Sayı.3.%	3 Sayılık Atış Yüzdesi	Başarılı 3 Sayılık Atış Sayısı / Denenen 3 Sayılık Atış Sayısı
Serb.Atış.%	Serbest Atış Yüzdesi	Başarılı Serbest Atış Sayısı / Denenen Serbest Atış Sayısı
Saha.İçi.%	Saha İçi Atış Yüzdesi	Başarılı Saha İçi İsbet Sayısı / Denenen Saha İçi İsbet Sayısı
Hüc.Rib.%	Hücum Ribaundu Yüzdesi	Bir takımın alabileceği bütün olası hücum ribauntlarının yüzde olarak ne kadarını aldığı
Sav.Rib.%	Savunma Ribaundu Yüzdesi	Bir takımın alabileceği bütün olası savunma ribauntlarının yüzde olarak ne kadarını aldığı
Rib.%	Ribaund Yüzdesi	Bir takımın alabileceği bütün olası bütün ribauntlarının yüzde olarak ne kadarını aldığı
Etkin.Şut.%	Etkin Şut Yüzdesi	(Sayı.2.Başarılı + 1.5 X Sayı.3.Başarılı) / Saha.İçi.Denenen
Gerçek.Şut.%	Gerçek Şut Yüzdesi	Atılan Sayı/(2 X (Saha.İçi.Denenen+ 0.44 X Serb.Atış.Denenen))
Serb.Atış.Oranı	Serbest Atış Oranı	Serb.Atış.Denenen/Saha.İçi.Denenen
Poz.Başına.Top.Kaybı	Pozisyon Başına Top Kaybı	Top Kaybı Sayısı / Pozisyon Sayısı

En başarılı bulunan ve başarı yüzdesi olarak diğerlerinden ayrılan üç model seçilmiş ve bu modeller kullanılarak maç sonucunu en çok etkileyen maç içi istatistikler belirlenmiştir. En önemli maç içi istatistiklerin belirlenmesinde, hesaplanan değişken önemi kullanılmıştır. Değişken önemi, farklı makine öğrenmesi yöntemleri için farklı şekillerde hesaplanmaktadır. Değişken önemleri grafik yardımıyla görselleştirilmiştir.

Çalışmanın son aşamasında maçlar, maç sonucundaki skor farklarına göre gruplara ayrılmışlardır. Bunun için skor farkları k-ortalama kümeleme analizi öklid uzaklığı kullanılarak kümelere ayrılmıştır ve en uygun küme sayısı üç olarak

belirlenmiştir. Maç sonucundaki fark için kümeleme analizi kritik değerlerin 10 ve 21 olduğunu göstermiştir. Maç sonucundaki farkın 10 ya da daha az olduğu maçlar “yakın maçlar”, 11 ile 21 arasında olan maçlar “dengeli maçlar” ve 21 in üzerinde olduğu maçlar “dengeli olmayan maçlar” olarak sınıflandırılmıştır. Makine öğrenmesi yöntemleri her üç sınıf üzerinde tekrarlanarak, bu gruptaki maçları tahmin etmedeki model başarısı ölçülmüş ve her bir gruptaki maç sonucunu etkileyen etmenler ortaya konulmuştur.

Makine öğrenmesi yöntemleriyle modelleme ve kümeleme analizi için R programlama dili kullanılmıştır. Makine öğrenmesi yöntemleri için caret (Kuhn, 2008) paketi yardımıyla, k en yakın komşuluk algoritması için class paketindeki knn() fonksiyonu, naive bayes algoritması için klaR paketindeki NaiveBayes() fonksiyonu, lojistik regresyon için glmnet paketi, destek karar makineleri için kernlab paketindeki ksvm() fonksiyonu, karar ağaçları için C50 paketindeki C50() fonksiyonu, rastgele orman için ranger paketindeki ranger() fonksiyonu, yapay sinir ağları için nnet paketindeki nnet() fonksiyonu ve kümeleme analizi için ise kmeans() fonksiyonu kullanılmıştır.

Kullanılan Makine Öğrenmesi Yöntemleri

K-En Yakın Komşuluk

K En Yakın Komşuluk algoritması, sınıflama problemleri için kullanılan en temel makine öğrenmesi yöntemlerindedir. Bu algoritma parametrik olmayan bir algoritma olup, sınıflama yaparken gözlemlerin Özellik benzerliğini kullanır. Algoritmanın çalışma mantığı temel olarak şöyledir (Hastie ve ark., 2021).

1. Birden büyük pozitif k sayısı belirlenir. k değeri sınıflama yapılırken en yakın kaç komşu değer göz önüne alınacağını belirtir. k değeri belirlenirken algoritma farklı k değerleriyle sınanarak en yüksek başarıya sahip k değeri seçilebilir.
2. Noktalar arasındaki uzaklığı ölçmek için bir ölçüt belirlenir. En sık kullanılan uzaklık ölçütü öklid uzaklık ölçütüdür.
3. Test verisindeki her bir gözlem için
 - a. Yeni sınıfa atanacak gözlem ile eğitim verisindeki her bir gözlem arasındaki uzaklık ölçülür.
 - b. Uzaklık değerleri küçükten büyüğe sıralanır.
 - c. Sıralanmış veride en küçük k adet değer göz önüne alınır.
 - d. k adet gözlemin frekansı hangi sınıfta daha fazlaysa yeni gözlem o sınıfa atanır.

Naive Bayes

Naive Bayes algoritması, Bayes teoremine dayanan bir sınıflama algoritmasıdır. Girdi değişkenleri birbirinden bağımsız olduğunda daha iyi sonuç verir. Bayes teoremi şu şekilde ifade edilir.

$$P(a|X) = \frac{P(X|a)P(a)}{P(X)}$$

Burada $P(a)$ sınıfın önsel olasılığını, $P(X|a)$ olabirlik fonksiyonunu; $P(X)$ ise tahmincinin önsel olasılığını gösterir. Bu değerler kullanılarak $P(a|X)$ sonsal olasılık; X Özellikleri verildiğinde gözlemin a sınıfında yer alma olasılığı olarak ifade edilir.

Bu denklem kullanılarak Naive Bayes algoritması basitçe şu şekilde ifade edilebilir (Hastie ve ark., 2021).

1. Veri setinden her bir sınıfın önsel olasılığı hesaplanır.
2. Her bir olası girdi değeri için olabirlik fonksiyonu hesaplanır.
3. Bayes formülü yardımıyla her bir çıktı sınıfı için olasılık hesaplanır.
4. Gözlem hangi sınıftaki olasılığı daha yüksekse o sınıfa atanır.

Lojistik Regresyon

Lojistik Regresyon bağımlı değişkenin iki sınıfa sahip olduğu durumlar için sıklıkla kullanılan bir sınıflama algoritmasıdır. Lojistik regresyon modeli, bağıntısıyla kurulur. Bu model sonunda yapılan hesaplamalar $[0-1]$ aralığında bir olasılık değeri üretilir. Bu olasılık değeri kullanılarak her bir gözlemi mevcut iki sınıftan birine atanır (Rasouliyan ve Miller, 2006).

$$\ln = \left(\frac{P(Y)}{1-P(Y)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Destek Vektör Makineleri

Destek Vektör Makineleri, hem sınıflama hem regresyon problemleri için kullanılan bir algoritmadır. Algoritma sınıflama işlemini gözlemleri birbirinden ayrık şekilde sınıflayabileceği N -boyutlu bir hiperdüzlem bularak yapar. Algoritma olası sonsuz sayıda hiperdüzlem içerisinde sınıflar arası uzaklığı maksimum olanı tercih eder (Smola ve Schölkopf, 2004).

Karar Ağaçları

Karar ağaçları hem regresyon hem de sınıflama problemleri için kullanılan bir algoritmadır. Algoritma verideki Özellikleri kullanarak belirli kurallar oluşturur ve bu kurallar yardımıyla veriyi daha küçük parçalara ayırır. Karar ağaçları düğüm ve dallar yardımıyla görsel olarak da sunulabilen bir algoritmadır.

Bu çalışmada C5.0 karar ağacı yöntemi kullanılmıştır (Kuhn ve Johnson, 2013). Bunun dışında farklı karar ağacı algoritmaları bulunsa da karar ağaçlarının temel çalışma prensibi şu şekildedir. Algoritma tek bir düğüm ile başlar ve bu düğüm üzerinde tekrar eden bir işleme başlar. Her adımda, bu düğümün her bir Özelliğe göre ayrıştırılmasıyla elde edilen “kazanç” metriği hesaplanır. Olası bütün ayrıştırmalardan elde edilen kazançlar karşılaştırılarak, düğümün dallara ayrılması ya da ayrılmamasına karar verilir. Yeni bir ayrışma olmayana kadar işlem devam eder.

Rastgele Orman

Rastgele Orman, birden fazla karar ağacının bir araya gelmesiyle oluşturulan bir algoritmadır. Tek bir karar ağacının verdiği sonucu kullanmaktansa çok daha fazla sayıda ve birbiriyle düşük korelasyona sahip olan karar ağacı üretilerek, tüm bu karar ağaçlarından elde edilen sonuçlara göre karar verilir (Kuhn ve Johnson, 2013; Navega ve ark., 2015; Hastie ve ark., 2021).

Yapay Sinir Ağları

Yapay sinir ağları hem sınıflama hem regresyon problemleri için kullanılan bir makine öğrenmesi algoritmasıdır. Algoritma girdiler, katmanlar ve çıktılar olmak üzere üç parçadan oluşur. Kullanılan veri seti, girdi kısmını oluştururken, veri katman kısmında işlenir ve çıktı kısmında sonuçlar elde edilir. Katman kısmında algoritma her bir değişkene bir ağırlık atayarak net girdi fonksiyonunu oluşturur, net girdi fonksiyonu bir aktivasyon fonksiyonu yardımıyla işlenir. Ardından bu işlemler geriye doğru çalıştırılarak ağırlıklar güncellenir ve hata istenilen düzeyden düşük olunca kadar güncelleme işlemi tekrar edilir (Bishop, 2006).

K-Ortalama kümeleme algoritması

Çalışmada, maç sonuçları oluşan farklara göre gruplara ayrılmıştır. Gruplara ayırmak için en uygun sayı farkları K-ortalama kümeleme algoritması kullanılarak belirlenmiştir. Bu algoritma şu şekilde çalışır (Hastie ve ark., 2021).

1. Veri setinin ayrılacağı küme sayısı – k belirlenir.
2. Veri her bir gözlemi rassal olarak bir kümeye atar.

3. Küme merkezleri belirlenir.
4. Her bir gözlem noktası küme merkezlerine olan uzaklıklarına bakılarak en yakın kümenin sınıfına atanır.
5. Herhangi bir gözlemin, kümesi değişmeyinceye kadar üçüncü ve dördüncü adımlar tekrarlanır.

MODELLERİN BAŞARI KRİTERLERİ

Sınıflama probleminde, model başarısı hata matrisi yardımıyla hesaplanır. Hata matrisi, sütunda gerçek değerlerin satırda ise tahmin değerlerinin bulunduğu bir matristir. Basketbol maç sonucu “Galibiyet” ve “Mağlubiyet” olarak iki sınıfa sahip olduğundan hata matrisi 2x2 boyutlu olacaktır. Tablo-3 de 2x2 boyutlu örnek bir hata matrisi verilmiştir.

Tablo 3: Örnek Bir Hata Matrisi

		GERÇEK DEĞERLER	
		POZİTİF	NEGATİF
TAHMİN	POZİTİF	DP	YP
DEĞERLERİ	NEGATİF	YN	DN

Bu matriste köşegen elemanlar sırasıyla pozitif ve negatif değerlerin doğru tahmin edildiği Doğru Pozitif (DP) ve Doğru Negatif (DN) sayısını verirken; ters köşegen elemanları ise yanlış tahmin edilen Yanlış Pozitif (YP) ve Yanlış Negatif (YN) değerlerini verir.

Bu dört değer kullanılarak hesaplanan aşağıdaki değerler model performansının ölçülmesinde kullanılır (Gorunescu, 2011).

Doğruluk= $(DP+DN)/(DP+DN+YP+YN)$, Modelin doğru sınıflama yüzdesini verir.

Hata Oranı= $1-\text{Doğruluk}$, Modelin yanlış sınıflama yüzdesini verir.

Duyarlılık= $DP/(DP+YN)$, Modelin Pozitif sınıfı bilme konusundaki başarısını verir.

Seçicilik= $DN/(DN+YP)$, Modelin Negatif sınıfı bilme konusundaki başarısını verir.

Kesinlik= $DP/(DP+YP)$, Modelin pozitif olarak sınıfladığı tüm gözlemlerin içinden gerçekten pozitif olanların oranını verir.

F1 Skoru= $(2 \times \text{Duyarlılık} \times \text{Kesinlik}) / (\text{Duyarlılık} + \text{Kesinlik})$, F1 Skoru, duyarlılık ve kesinliği bir arada göz önüne alarak hesaplandığından farklı modellerin birbiriyle karşılaştırılabilmesini sağlar.

En iyi modele, bu ölçütler birlikte değerlendirilerek karar verilmelidir.

DEĞİŞKEN ÖNEMİ

Maç sonucunda etkili olan değişkenler, caret (Kuhn, 2008) paketinin sonucunda verilen değişken önemi değerleriyle belirlenmiştir. Değişken önemleri, hem tüm maçlar için hem de her bir maç grubu için, en başarılı üç makine öğrenmesi sonucu için ayrı ayrı raporlanmıştır. En başarılı üç makine öğrenmesi yöntemi lojistik regresyon, destek vektör makineleri ve yapay sinir ağları olarak bulunmuştur. Değişken önemi, lojistik regresyon için son modelin standardize edilmiş katsayılarının mutlak değerine göre belirlenirken, destek vektör makinelerinde ROC eğrisinin altında kalan alana göre belirlenir. Yapay sinir ağlarında ise değişken önemi, ağırlıkların mutlak değerinin kombinasyonu kullanılarak bulunur (Gevrey ve ark., 2003).

BULGULAR

K-ortalama kümeleme analizi, maç sonucu farkları için kritik değerlerin 10 ve 21 olduğu görülmüştür. Buna göre maç sonucu farkları [1 – 10] arasındaki maçlar “yakın maçlar”, (10 – 21) arasındaki maçlar “dengeli maçlar” ve 22 ve üzerindeki farkla biten maçlar ise “dengeli olmayan maçlar” olarak adlandırılmıştır. 1358 adet maçın 762 tanesi yakın maç, 451 tanesi dengeli maç ve 145 tanesi ise dengeli olmayan maç olarak sınıflandırılmıştır.

Her bir grupta, maç içi istatistiklerinin galibiyet ve mağlubiyet durumunda değişiklik gösterip göstermediği, Tablo-4’te belirtilmiştir. Mann-Whitney-U testinin sonucuna göre hemen hemen her maç içi istatistiklerinin galibiyet ve mağlubiyette anlamlı farklılık gösterdiği gözlemlenmiştir. Farklılık göstermeyen değişkenler 2 Sayılık Atış Denemesi, 3 Sayılık Atış Denemesi, Saha İçi Şut Denemesi ve Hücum Ribaundudur. Yapılan ve alınan bloklar yakın maçlar için farklılık göstermemiş, alınan ve yapılan fauller ise dengeli olmayan maçlar için farklılık göstermemiştir.

Her bir makine öğrenmesi yöntemi, maç sonucunun (Galibiyet ya da Mağlubiyet) bağımlı değişken, diğer maç içi istatistiklerinin bağımsız değişken olduğu 4 veri seti için uygulanmış (Tüm Maçlar, Yakın Maçlar, Dengeli Maçlar ve Dengeli Olmayan Maçlar) ve model başarıları Tablo-5’te Özetlenmiştir.

Tablo-5’e göre hem tüm maçlar için, hem de gruplara göre ayrılmış veri setleri için makine öğrenmesi yöntemlerine bakıldığında en başarılı makine

öğrenmesi algoritmaları lojistik regresyon, karar destek makinaları ve yapay sinir ağlarıdır. Tüm modellerde F1 skorunun, doğruluk değerine yakın değerler aldığı görülmektedir. Bulunan sonuçların bu kapsamda tutarlı olduğu söylenebilir.

Tüm maçlar göz önüne alındığında en iyi modeller maçların sonucunu %84 oranında doğru bilmektedir. Bu maç grubu için en düşük başarıya sahip olan algoritma %73 ile naive bayes algoritmasıdır.

Yakın maçlar göz önüne alındığında tüm algoritmaların başarı oranının düştüğü görülmektedir. En başarılı modelin doğru sınıflama yüzdesi %81 ile lojistik regresyon iken en düşük başarı %65 ile naive bayes algoritmasıdır.

Dengeli maçlara bakıldığında doğru sınıflama yüzdesinin %97 ye kadar çıktığı görülmektedir. Bu maç grubunda en başarısız model %87 başarıyla k en yakın komşuluk yöntemidir.

Dengeli olmayan maçlar için bakıldığında pek çok modelin %100 başarıya ulaştığı görülmektedir. En düşük başarı ise %98 ile k en yakın komşuluk yöntemidir. Her ne kadar bu veri seti için mükemmel başarıya ulaşılmışsa da dengeli olmayan maç verisinde diğer gruplara göre daha az sayıda maç olduğu ve bu grupta maç sayısının artması halinde model başarısının az da olsa düşme ihtimalinin bulunduğu göz önüne alınmalıdır.

Tüm bu sonuçlar göz önüne alındığında maç sonucunu tahmin etmede en etkili değişkenleri bulabilmek için lojistik regresyon, karar destek makinaları ve yapay sinir ağları modelleri için oluşturulan değişken önemi sonuçlarına bakılmalıdır. Bu amaçla her üç model için oluşturulan değişken önemi grafiği Şekil-1 den görülebilir.

Tüm maçlar için maç sonucu üzerinde en çok etkili olan değişkenler lojistik regresyon algoritması için savunma ribaundu, gerçek şut yüzdesi, top çalma ve top kaybı iken; destek vektör makinalarında gerçek şut yüzdesi, etkin şut yüzdesi, saha içi yüzdesi, saha içinde başarılı olan şut adedinin en etkili değişkenler olduğu görülmektedir. Bu algoritma için, savunma ribaundu, sayı pası, 3 sayı yüzdesi, 2 sayı yüzdesi ve sayı pası top kaybı oranının ikinci dereceden önemli değişkenler olduğu görülmektedir. Yapay sinir ağları için ise savunma ribaundunun en önemli değişken olduğu, top çalma, top kaybı, hücum ribaundu, serbest atış denemesi, gerçek şut yüzdesi, saha içi denemesi ve 2 sayılık atış denemesi sayısının etkili olduğu bulunmuştur.

Tablo 4: Maç Türleri ve Galibiyet ve Mağlubiyete göre Maç İçi İstatistiklere Ait Betimsel İstatistikler ve Mann-Whitney U Testi Sonucu

TÜM MAÇLAR (n=1358)												
Değişken	GALİBİYET					MAĞLUBİYET					MW -U Testi	
	Ort.	St. Sap.	Med.	Mak.	Min.	Ort.	St. Sap.	Med.	Mak.	Min.	p	A.D.
HücRib	10.04	3.59	10	30	1	10.25	3.62	10	23	1	0.115	ÖD
SavRib	24.81	4.24	25	40	10	21.62	4.16	21	37	8	0.000	***
SayıPası	18.91	4.40	19	34	4	16.11	3.84	16	28	4	0.000	***
TopÇalma	6.90	2.72	7	18	0	6.03	2.41	6	17	0	0.000	***
TopKaybı	12.10	3.34	12	25	3	13.14	3.70	13	25	2	0.000	***
Blok_Yap	2.74	1.82	3	10	0	2.32	1.63	2	10	0	0.000	***
Blok_Alı	2.32	1.63	2	10	0	2.74	1.82	3	10	0	0.000	***
Faul_Yap	20.10	3.55	20	35	9	21.01	3.72	21	36	11	0.000	***
Faul_Alı	20.81	3.68	21	34	11	19.93	3.49	20	34	9	0.000	***
Sayı2_Başarılı	21.07	4.29	21	38	8	18.91	3.95	19	31	7	0.000	***
Sayı2_Denenen	37.89	6.33	38	66	15	37.60	6.30	37	59	19	0.277	ÖD
Sayı3_Başarılı	9.51	3.14	9	21	2	8.03	2.82	8	19	1	0.000	***
Sayı3_Denenen	23.31	5.32	23	40	8	23.59	5.37	23	41	6	0.167	ÖD
SerbAtış_Başarılı	14.63	5.68	14	37	0	12.86	5.21	12	38	0	0.000	***
SerbAtış_Denenen	18.70	6.73	18	44	0	16.92	6.30	17	44	0	0.000	***
Sahaİçi_Başarılı	30.58	4.07	31	45	19	26.94	3.76	27	39	15	0.000	***
Sahaİçi_Denenen	61.20	5.94	61	84	45	61.19	5.98	61	84	41	0.803	ÖD
SayıPası_TopKaybı	1.71	0.74	1.56	7.25	0.38	1.35	0.64	1.24	10.5	0.29	0.000	***
Sayı2_Yüzde	0.56	0.08	0.56	0.85	0.31	0.51	0.08	0.5	0.76	0.23	0.000	***
Sayı3_Yüzde	0.41	0.10	0.41	0.83	0.11	0.34	0.10	0.34	0.65	0.05	0.000	***
SerbAtış_Yüzde	0.78	0.11	0.79	1	0.25	0.76	0.13	0.78	1	0	0.000	***
Sahaİçi_Yüzde	0.50	0.06	0.5	0.7	0.32	0.44	0.06	0.44	0.61	0.28	0.000	***
HücRib_Yüzde	0.31	0.09	0.31	0.62	0.05	0.29	0.08	0.29	0.66	0.04	0.000	***
SavRib_Yüzde	0.71	0.08	0.71	0.96	0.34	0.69	0.09	0.69	0.95	0.38	0.000	***
Rib_Yüzde	0.51	0.06	0.51	0.71	0.34	0.49	0.06	0.49	0.66	0.29	0.000	***
EtkŞut_Yüzde	0.58	0.07	0.58	0.84	0.38	0.51	0.07	0.51	0.74	0.31	0.000	***
GerçekŞut_Yüzde	0.62	0.07	0.61	0.86	0.43	0.55	0.06	0.54	0.76	0.35	0.000	***
SerbAtış_Oran	0.24	0.10	0.23	0.76	0	0.21	0.10	0.2	0.93	0	0.000	***
TopKaybı_Pozisyon	0.17	0.05	0.17	0.34	0.04	0.19	0.05	0.19	0.36	0.03	0.000	***

YAKIN MAÇLAR (n=762)												
Değişken	GALİBİYET					MAĞLUBİYET					MW -U Testi	
	Ort.	St. Sap.	Med.	Mak.	Min.	Ort.	St. Sap.	Med.	Mak.	Min.	p	A.D.
HücRib	10.12	3.56	10	30	1	10.41	3.56	10	23	1	0.045	*
SavRib	24.10	4.08	24	39	10	22.40	4.17	22	37	12	0.000	***
SayıPası	17.70	4.08	18	34	4	16.56	3.88	17	28	6	0.000	***
TopÇalma	6.51	2.56	6	18	0	6.07	2.44	6	15	0	0.001	**
TopKaybı	12.23	3.35	12	25	3	12.75	3.56	13	23	2	0.003	**
Blok_Yap	2.59	1.73	2	9	0	2.41	1.66	2	10	0	0.061	ÖD
Blok_Alı	2.41	1.66	2	10	0	2.59	1.73	2	9	0	0.061	ÖD
Faul_Yap	20.54	3.49	20	35	9	21.65	3.64	21	36	11	0.000	***
Faul_Alı	21.47	3.61	21	34	11	20.37	3.44	20	34	9	0.000	***
Sayı2_Başarılı	20.52	4.02	21	34	8	19.38	3.94	19	31	7	0.000	***
Sayı2_Denenen	37.83	6.14	38	65	15	37.62	6.32	37	59	19	0.595	ÖD
Sayı3_Başarılı	8.84	2.82	9	18	2	8.47	2.83	8	19	1	0.023	*
Sayı3_Denenen	22.87	5.36	23	40	8	24.04	5.40	24	40	8	0.000	***
SerbAtış_Başarılı	15.30	5.76	15	37	2	13.69	5.22	13	38	2	0.000	***
SerbAtış_Denenen	19.61	6.75	19	44	2	17.74	6.30	18	44	2	0.000	***
Sahalçı_Başarılı	29.36	3.66	29	42	19	27.85	3.58	28	39	17	0.000	***
Sahalçı_Denenen	60.70	5.96	60	84	47	61.66	5.95	62	84	41	0.000	***
SayıPası_TopKaybı	1.57	0.62	1.45	5.33	0.38	1.43	0.70	1.3	10.5	0.43	0.000	***
Sayı2_Yüzde	0.54	0.08	0.54	0.77	0.31	0.52	0.08	0.52	0.76	0.29	0.000	***
Sayı3_Yüzde	0.39	0.10	0.39	0.68	0.11	0.35	0.09	0.36	0.65	0.05	0.000	***
SerbAtış_Yüzde	0.78	0.11	0.78	1	0.41	0.77	0.11	0.79	1	0.25	0.584	ÖD
Sahalçı_Yüzde	0.49	0.06	0.48	0.67	0.32	0.45	0.05	0.45	0.61	0.28	0.000	***
HücRib_Yüzde	0.31	0.08	0.3	0.62	0.05	0.30	0.08	0.3	0.66	0.04	0.024	*
SavRib_Yüzde	0.70	0.08	0.7	0.96	0.34	0.69	0.08	0.7	0.95	0.38	0.024	*
Rib_Yüzde	0.50	0.06	0.5	0.69	0.35	0.50	0.06	0.5	0.65	0.31	0.007	**
EtkŞut_Yüzde	0.56	0.07	0.56	0.79	0.38	0.52	0.06	0.52	0.74	0.31	0.000	***
GerçekŞut_Yüzde	0.60	0.06	0.6	0.81	0.43	0.56	0.06	0.56	0.76	0.35	0.000	***
SerbAtış_Oran	0.26	0.11	0.25	0.76	0.03	0.23	0.10	0.22	0.93	0.03	0.000	***
TopKaybı_Pozisyon	0.18	0.05	0.17	0.32	0.05	0.18	0.05	0.18	0.32	0.03	0.008	**

MW-U: Mann-Whitney U **Ort.:** Ortalama, **St. Sap.:** Standart Sapma, **Med.:** Medyan, **Mak.:** Maksimum, **Min.:** Minimum, **A.D.:** Anlamlılık Düzeyi * 0.05, **0.01 ***0.001
ÖD: Önemli Değil

Tablo 4 (devam): Maç Türleri ve Galibiyet ve Mağlubiyete göre Maç İçi İstatistiklere Ait Betimsel İstatistikler ve Mann-Whitney U Testi Sonucu

DENGELİ MAÇLAR (n=451)												
Değişken	GALİBİYET					MAĞLUBİYET					MW -U Testi	
	Ort.	St. Sap.	Med.	Mak.	Min.	Ort.	St. Sap.	Med.	Mak.	Min.	p	A.D.
HücRib	9.95	3.74	10	21	1	9.98	3.66	10	21	2	0.972	ÖD
SavRib	25.33	4.14	26	40	15	21.10	3.76	21	31	12	0.000	***
SayıPası	19.84	4.26	20	32	11	15.89	3.59	16	27	5	0.000	***
TopÇalma	7.31	2.90	7	16	0	5.96	2.34	6	17	0	0.000	***
TopKaybı	12.05	3.35	12	24	3	13.38	3.76	13	24	4	0.000	***
Blok_Yap	2.98	1.94	3	10	0	2.21	1.60	2	9	0	0.000	***
Blok_Alı	2.21	1.60	2	9	0	2.98	1.94	3	10	0	0.000	***
Faul_Yap	19.69	3.49	19	33	11	20.29	3.66	20	31	11	0.004	**
Faul_Alı	20.08	3.61	20	30	11	19.51	3.39	19	33	11	0.007	**
Sayı2_Başarılı	21.40	4.50	21	37	9	18.42	3.89	18	31	7	0.000	***
Sayı2_Denenen	37.84	6.47	38	64	20	37.41	6.14	37	58	20	0.366	ÖD
Sayı3_Başarılı	10.04	3.18	10	21	2	7.73	2.63	8	16	1	0.000	***
Sayı3_Denenen	23.67	5.19	24	40	9	23.16	5.22	23	40	6	0.137	ÖD
SerbAtış_Başarılı	13.64	5.42	13	31	0	12.08	5.05	12	31	1	0.000	***
SerbAtış_Denenen	17.55	6.55	17	38	0	16.06	6.08	16	38	2	0.000	***
Sahaİçi_Başarılı	31.43	3.86	31	42	20	26.14	3.59	26	38	16	0.000	***
Sahaİçi_Denenen	61.52	5.90	62	78	45	60.57	5.87	60	82	43	0.010	*
SayıPası_TopKaybı	1.82	0.82	1.65	7.25	0.61	1.30	0.54	1.2	4.2	0.29	0.000	***
Sayı2_Yüzde	0.57	0.08	0.56	0.77	0.32	0.49	0.08	0.49	0.74	0.23	0.000	***
Sayı3_Yüzde	0.42	0.10	0.42	0.7	0.11	0.33	0.09	0.33	0.64	0.06	0.000	***
SerbAtış_Yüzde	0.78	0.12	0.79	1	0.25	0.75	0.13	0.76	1	0.33	0.002	**
Sahaİçi_Yüzde	0.51	0.06	0.51	0.7	0.34	0.43	0.06	0.43	0.6	0.29	0.000	***
HücRib_Yüzde	0.32	0.09	0.31	0.62	0.05	0.28	0.08	0.28	0.51	0.06	0.000	***
SavRib_Yüzde	0.72	0.08	0.72	0.94	0.49	0.68	0.09	0.69	0.95	0.38	0.000	***
Rib_Yüzde	0.52	0.06	0.52	0.71	0.34	0.48	0.06	0.48	0.66	0.29	0.000	***
EtkŞut_Yüzde	0.59	0.07	0.59	0.84	0.39	0.50	0.06	0.49	0.7	0.33	0.000	***
GerçekŞut_Yüzde	0.63	0.06	0.62	0.86	0.45	0.53	0.06	0.53	0.72	0.37	0.000	***
SerbAtış_Oran	0.23	0.10	0.22	0.62	0	0.20	0.09	0.19	0.58	0.01	0.000	***
TopKaybı_Pozisyon	0.17	0.05	0.17	0.34	0.04	0.19	0.05	0.19	0.34	0.06	0.000	***

DENGELİ OLMAYAN MAÇLAR (n=145)

Değişken	GALİBİYET					MAĞLUBİYET					MW -U Testi	
	Ort.	St. Sap.	Med.	Mak.	Min.	Ort.	St. Sap.	Med.	Mak.	Min.	p	A.D.
HücRib	9.91	3.24	10	20	2	10.23	3.81	10	21	3	0.697	ÖD
SavRib	26.95	4.48	27	38	15	19.10	4.06	19	32	8	0.000	***
SayıPası	22.31	4.01	22	31	11	14.40	3.86	15	24	4	0.000	***
TopÇalma	7.65	2.61	7	15	2	6.03	2.51	6	13	1	0.000	***
TopKaybı	11.54	3.24	11	20	4	14.39	3.90	14	25	5	0.000	***
Blok_Yap	2.83	1.80	3	9	0	2.21	1.51	2	7	0	0.003	**
Blok_Alı	2.21	1.51	2	7	0	2.83	1.80	3	9	0	0.003	**
Faul_Yap	19.02	3.74	19	28	11	19.87	3.69	20	30	11	0.059	ÖD
Faul_Alı	19.68	3.62	19	30	11	18.92	3.68	19	28	11	0.080	ÖD
Sayı2_Başarılı	22.96	4.40	22	38	12	17.95	3.82	18	30	10	0.000	***
Sayı2_Denenen	38.32	6.81	38	66	23	38.02	6.67	37	56	23	0.632	ÖD
Sayı3_Başarılı	11.44	3.47	12	19	2	6.63	2.76	6	13	1	0.000	***
Sayı3_Denenen	24.49	5.24	25	38	11	22.64	5.47	22	41	11	0.003	**
SerbAtış_Başarılı	14.18	5.60	14	31	2	10.93	4.75	11	26	0	0.000	***
SerbAtış_Denenen	17.50	6.44	17	37	2	15.32	6.35	15	36	0	0.007	**
Sahaİçİ_Başarılı	34.40	3.68	34	45	26	24.59	3.65	24	32	15	0.000	***
Sahaİçİ_Denenen	62.81	5.67	63	77	48	60.66	6.28	61	76	45	0.004	**
SayıPası_TopKaybı	2.11	0.81	1.92	5.75	0.82	1.09	0.49	1	4.4	0.3	0.000	***
Sayı2_Yüzde	0.60	0.08	0.59	0.85	0.41	0.47	0.08	0.47	0.73	0.28	0.000	***
Sayı3_Yüzde	0.47	0.11	0.46	0.83	0.18	0.29	0.10	0.29	0.53	0.07	0.000	***
SerbAtış_Yüzde	0.81	0.11	0.83	1	0.53	0.72	0.15	0.72	1	0	0.000	***
Sahaİçİ_Yüzde	0.55	0.05	0.55	0.7	0.44	0.41	0.06	0.4	0.54	0.28	0.000	***
HücRib_Yüzde	0.34	0.09	0.34	0.61	0.07	0.27	0.08	0.26	0.5	0.08	0.000	***
SavRib_Yüzde	0.73	0.08	0.74	0.92	0.5	0.66	0.09	0.66	0.93	0.39	0.000	***
Rib_Yüzde	0.53	0.06	0.53	0.67	0.4	0.47	0.06	0.47	0.6	0.33	0.000	***
EtkŞut_Yüzde	0.64	0.07	0.64	0.79	0.48	0.46	0.07	0.46	0.61	0.31	0.000	***
GerçekŞut_Yüzde	0.67	0.07	0.67	0.82	0.51	0.50	0.06	0.5	0.64	0.36	0.000	***
SerbAtış_Oran	0.23	0.10	0.22	0.54	0.03	0.18	0.09	0.18	0.46	0	0.000	***
TopKaybı_Pozisyon	0.16	0.04	0.17	0.27	0.06	0.21	0.06	0.21	0.36	0.08	0.000	***

MW-U: Mann-Whitney U **Ort.:** Ortalama, **St. Sap.:** Standart Sapma, **Med.:** Medyan, **Mak.:** Maksimum, **Min.:** Minimum, **A.D.:** Anlamlılık Düzeyi * 0.05, **0.01 ***0.001
ÖD: Önemli Değil

Tablo 5: Makine Öğrenmesi Algoritmalarının Performansı

	TÜM MAÇLAR							DENGELİ MAÇLAR						
	KEYK	NB	LR	DVM	KA	RO	YSA	KEYK	NB	LR	DVM	KA	RO	YSA
Doğruluk	0.763	0.736	0.844	0.842	0.807	0.790	0.845	0.867	0.878	0.970	0.970	0.919	0.896	0.974
Hata_Oranı	0.237	0.264	0.156	0.159	0.193	0.210	0.155	0.133	0.122	0.030	0.030	0.082	0.104	0.026
Duyarlılık	0.745	0.745	0.848	0.843	0.801	0.781	0.853	0.859	0.874	0.970	0.956	0.882	0.904	0.970
Seçicilik	0.754	0.740	0.847	0.842	0.803	0.785	0.850	0.861	0.875	0.970	0.957	0.890	0.902	0.971
Kesinlik	0.773	0.731	0.842	0.841	0.811	0.795	0.840	0.872	0.881	0.970	0.985	0.952	0.891	0.978
F1_Skoru	0.767	0.734	0.843	0.841	0.808	0.792	0.844	0.868	0.878	0.970	0.971	0.921	0.896	0.974

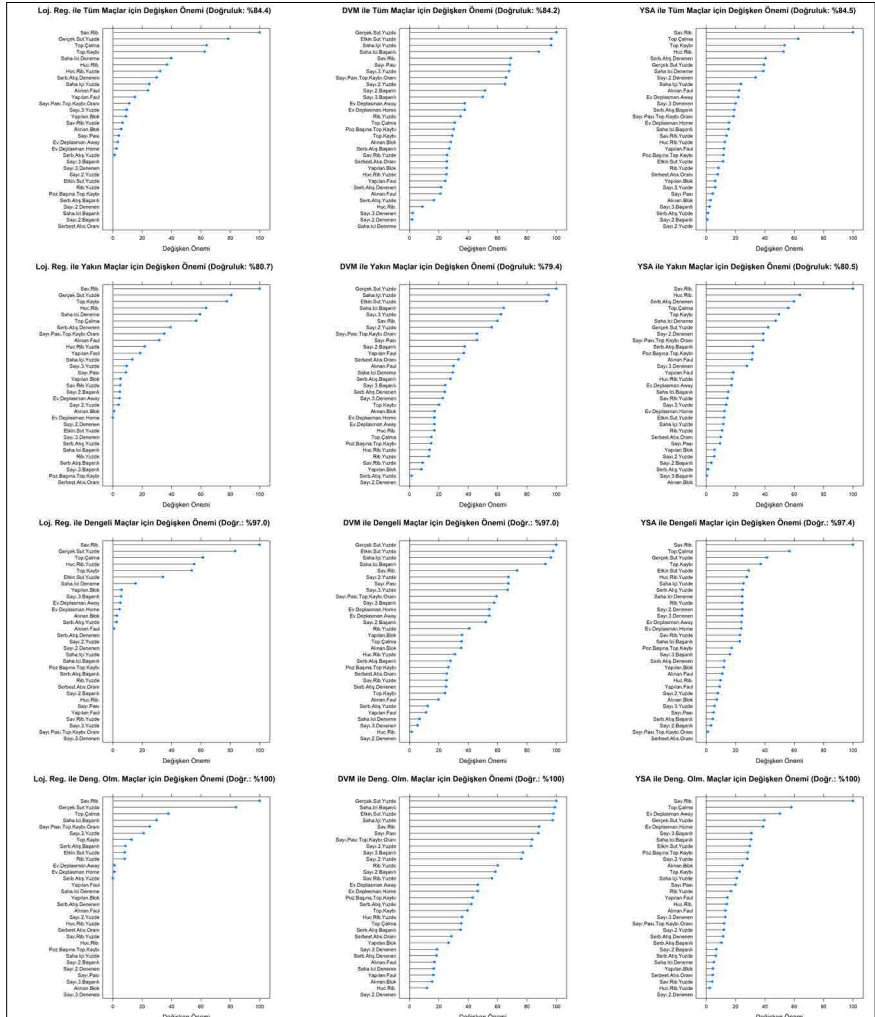
	YAKIN MAÇLAR							DENGELİ OLMAYAN MAÇLAR						
	KEYK	NB	LR	DVM	KA	RO	YSA	KEYK	NB	LR	DVM	KA	RO	YSA
Doğruluk	0.689	0.651	0.807	0.794	0.737	0.715	0.805	0.977	1.000	1.000	1.000	0.988	0.988	1.000
Hata_Oranı	0.311	0.349	0.193	0.206	0.263	0.285	0.195	0.023	0.000	0.000	0.000	0.012	0.012	0.000
Duyarlılık	0.667	0.662	0.829	0.794	0.781	0.746	0.833	0.954	1.000	1.000	1.000	1.000	0.977	1.000
Seçicilik	0.681	0.655	0.821	0.794	0.760	0.729	0.823	0.956	1.000	1.000	1.000	1.000	0.977	1.000
Kesinlik	0.697	0.648	0.794	0.794	0.718	0.703	0.788	1.000	1.000	1.000	1.000	0.977	1.000	1.000
F1_Skoru	0.695	0.648	0.803	0.794	0.725	0.706	0.799	0.977	1.000	1.000	1.000	0.988	0.989	1.000

KEYK: K-En Yakın Komşuluk, **NB:** Naive Bayes, **LR:** Lojistik Regresyon, **DVM:** Destek Vektör Makineleri, **KA:** Karar Ağacı, **RO:** Rastgele Orman, **YSA:** Yapay Sinir Ağları

Yakın giden maçlarda, lojistik regresyon savunma ribaundu, gerçek şut yüzdesi, top kaybı, hücum ribaundu, saha içi deneme sayısı ve top çalma istatistiklerinin maç sonucunu belirlemede etkili olduğunu göstermiştir. Destek vektör makineleri ise gerçek şut yüzdesinin, saha içi yüzdesinin, etkin şut yüzdesinin, saha içinde başarılı olan şut sayısı, 3 sayı yüzdesi, 2 sayı yüzdesi ve savunma ribaundunun yakın giden maçlarda maç sonucunu belirlemede en etkili değişkenler olduğunu ortaya koymuştur. Yapay sinir ağları, yakın giden maçlar için savunma ribaundunun en etkili değişken olduğunu ortaya koymuştur. Savunma ribaundundan sonra, hücum ribaundu, serbest atış denemesi, top çalma, top kaybı, saha içinden denenen şut sayısı ve gerçek şut yüzdesi en önemli değişkenlerdir.

Dengeli maçlara bakıldığında, lojistik regresyon algoritması çok daha az sayıda değişkenin maç sonucu üzerinde etkili olduğunu göstermiştir. En etkili değişkenler savunma ribaundu, gerçek şut yüzdesi, top çalma, hücum ribaundu yüzdesi ve top kaybıdır. Destek vektör makineleri ise dengeli maçlarda gerçek şut yüzdesinin, etkin şut yüzdesinin, saha içi yüzdesinin, saha içinde başarılı olan şut sayısının, savunma ribaundunun, 2 sayı yüzdesinin, sayı pasının ve 3 sayı yüzdesinin en etkili değişkenler olduğu görülmektedir. Yapay sinir ağları ise dengeli maçlar için savunma ribaundu, top çalma, gerçek şut yüzdesi ve top kaybının en etkili değişkenler olduğunu belirtmektedir.

Dengeli olmayan maçlarda, lojistik regresyon en etkili değişken olarak savunma ribaundu, gerçek şut yüzdesi, top çalmayı bulurken, destek vektör makinaları gerçek şut yüzdesi, saha içinden başarılı olan şut sayısı, etkin şut yüzdesi, saha içi yüzdesi, savunma ribaundu, sayı pası, top kaybı oranı, 3 sayı yüzdesi, başarılı 3 sayılık atış adedi, 2 sayılık yüzdesini en etkili değişkenler olarak bulmuştur. Yapay sinir ağları ise savunma ribaundu, top çalma, ev sahibi ya da misafir takım olmak ve gerçek şut yüzdesinin en etkili değişkenler olduğunu ortaya koymuştur.



Şekil 1. Maç tipine göre maç sonucunu en çok etkileyen maç içi istatistikler

TARTIŞMA

Basketbolda spor analitiğinin temelini oluşturan çalışmaların birinde maç sonucunu etkileyen en önemli dört etmenin olduğu belirtilmiştir (Oliver, 2004). Bu dört etmen, şut, top kaybı, ribaunt ve serbest atıştır. Literatürdeki benzer çalışmalara bakıldığında hem bu dört etmenin etkin olduğu, hem de başka etmenlerin de rol oynadığı görülmektedir. Ancak genel olarak bu dört etmenin birden etkili olduğunu bulan çalışma sayısı azdır. Bu dört etmenin birden maç sonuçları üzerinde etkili olduğunu belirten çalışmalar bulunsa da (Jones, 2016), bazı çalışmalar şut, ribaunt ve serbest atışın etkili olduğunu (Csataljay ve ark., 2009; Pai ve ark., 2017), bazı çalışmalar ise şut, ribaunt ve top kaybının etkili olduğunu (Lorenzo ve ark., 2010; Leicht ve ark., 2017) söylemektedir. Pek çok çalışma da şut ve ribaundun maç sonucunu belirlemede etkili olduğunu bulmuştur (Garcia ve ark., 2014; Çene, 2018; Kaur ve Jain, 2017).

Bu dört etmenin dışında sayı pasının etkili olduğunu söyleyen pek çok çalışma da mevcuttur (Ibáñez ve ark., 2008; Lorenzo ve ark., 2010; Magel ve Unruh, 2013; Garcia ve ark., 2014; Jones, 2016; Leicht ve ark., 2017; Pai ve ark., 2017). Bazı çalışmalarda ise top kaybı yerini top çalmaya bırakmıştır (Ibáñez ve ark., 2008; Garcia ve ark., 2014; Leicht ve ark., 2017; Pai ve ark., 2017; Çene, 2018).

Bu çalışmanın sonucuna bakıldığında tüm maçlar göz önüne alındığında uygulanan üç makine öğrenmesi algoritmasına bakıldığında savunma ribaundunun, gerçek şut yüzdesinin en etkili değişkenler arasında olduğu görülmektedir. Top kaybı ve top çalma lojistik regresyon ve yapay sinir ağları sonuçlarına göre en etkili değişkenler arasındayken, destek vektör makineleri için sayı pası top kaybı oranı, top kaybı ve top çalmaya göre daha etkili bulunmuştur. Serbest atış ile ilgili istatistikler doğrudan en etkili değişkenler arasında yer almazken, denenen serbest atış sayısı kısmen etkili bulunmuştur. Bu durum, serbest atış etkisinin bir kısmının gerçek şut yüzdesinin içinde yer almasıyla açıklanabilir. Bu dört etmenin dışında etkin bulunan değişkenler ise lojistik regresyon için hücum ribaundu yüzdesi ve alınan faul; destek vektör makineleri için sayı pası, 2 sayılıklı, 3 sayılıklı ve saha içi şut yüzdeleridir.

Maç türü değiştikçe makine öğrenmesi algoritmalarının bulduğu etkili değişkenler arasında çok ciddi farklılıklar oluşmamaktadır. Burada görülen en ciddi fark dengeli olmayan maçlarda yapay sinir ağlarının ev sahibi olmayı etkili olarak bulmasıdır.

Bu bilgiler ışığında bir maçı kazanmak için en etkili iki etmenin savunma ribaundu almak ve etkin bir şut yüzdesine sahip olmak olduğu söylenebilir. Savunma ribaundunu etkin bir şekilde almak hem rakibe verilecek hücum ribauntlarını ve ikinci şans sayılarını sınırlayacak hem de takımın hızlı hücumu çıkmasının önü-

nü açarak şut yüzdesini olumlu şekilde etkileyecektir. Bu iki değişkenin en etkili değişkenler olması bu yüzden şaşırtıcı değildir. Top kaybı ve top çalma Özellikle lojistik regresyon ve yapay sinir ağları modellerinde oldukça önemli bulunmuştur. Top kaybı yapmak takımın sayı atmasını engellemekte ve topun rakibe geçmesine sebep olmakta, top çaldırma ise rakip takıma doğrudan kolay bir sayı şansı sağlamaktadır. Çalışmada serbest atışla ilintili değişkenler çok etkili çıkmamıştır. Sayı pası da yine en önemli değişkenlerden değildir. Her ne kadar bu değişkenler kazanan ve kaybeden takımlar arasında farklılık gösterse de bu farklılık maç sonucunu belirleyecek bir yapı oluşturmamaktadır. Bu durum maçı kazanmak için illa top paylaşımının zorunlu olmadığını, daha çok sayı pası veren takımın mutlaka galip geleceği anlamına gelmediğini belirtmektedir. Benzer şekilde daha çok serbest atış çizgisine gelen takım illa maç kazanacak diye bir sonuca ulaşamamıştır.

SONUÇ VE ÖNERİLER

Bu çalışmada 2016-2017 ile 2020-2021 yılları arasında oynanan 1358 EuroLeague basketbol maçlarındaki takım istatistikleri göz önüne alınmış ve bu takım istatistiklerinden hangilerinin maçın galibi üzerinde en çok etkiye sahip olduğu belirlenmeye çalışılmıştır. Maçlar, kümeleme analizi sonucunun belirttiği skor farklarına göre yakın, dengeli ve dengeli olmayan olmak üzere üç gruba ayrılmıştır. Hem bu üç grup hem de tüm maçlara yedi farklı makine öğrenmesi algoritması uygulanmış, en etkili algoritmalar lojistik regresyon, destek vektör makineleri ve yapay sinir ağları olarak bulunmuştur. Bu üç algoritma maç sonucunu tüm maçlar için yaklaşık %84 oranında doğru bilmiştir. Yakın maçlarda bu oran %79 a düşmüş, dengeli maçlarda %97 'ye, dengeli olmayan maçlarda %100 e çıkmıştır. Maç sonucu üzerinde en çok etkili olan değişkenler savunma ribaundu, gerçek şut yüzdesi, top çalma, top kaybı, hücum ribaundu ve denenen serbest atıştır. Burada bulunan sonuçlar takımların maç içi stratejilerini belirlerken en çok odaklanmaları gereken konular üzerinde fikir vermekte ve bu konular üzerine yoğunlaşarak strateji belirlemelerine yardımcı olmaktadır.

KAYNAKLAR

- Bishop C.M. (2006). *Pattern Recognition and Machine Learning*, Springer.
- Çene E. (2018). What is the difference between a winning and a losing team: insights from Euroleague basketball. *International Journal of Performance Analysis in Sport*, 18(1), 55-68.
- Csataljay G, O'Donoghue P, Hughes M, & Dancs H. (2009). Performance indicators that distinguish winning and losing teams in basketball. *International Journal of Performance Analysis in Sport*, 9(1), 60-66.
- Davoodi E & Khanteymooori A.R. (2010). "Horse racing prediction using Artificial Neural Networks". *Proc. of the 11th WSEAS Int. Conf. on Neural Networks, NN '10, Proceedings of the 11th WSEAS Int. Conf. on Evolutionary Computing, EC '10, Proc. of the 11th WSEAS Int. Conf. on Fuzzy Systems, FS '10 içinde (ss. 155-160)*.
- Garcia J, Ibanez S.J., Gomez M.A., & Sampaio J. (2014). Basketball Game-related statistics discriminating ACB league teams according to game location, game outcome and final score differences. *International Journal of Performance Analysis in Sport*, 14(2), 443-452.

- Gevrey M, Dimopoulos I, & Lek S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* 160 (249-264).
- Gorunescu F. (2011). Classification performance evaluation. *Data Mining. Intelligent Systems Reference Library*, 12(319-330). Springer, Berlin, Heidelberg.
- Hastie T, Tibshirani R, James G, & Witten D. (2021). *An Introduction to Statistical Learning (2nd Edition)*. Springer Texts. New York.
- Horvat T, Havaš L, & Srpak D. (2020). The impact of selecting a validation method in machine learning on predicting basketball game outcomes. *Symmetry*, 12(3), 431.
- Ibáñez S, Sampaio J, Feu S, Lorenzo A, Gomez M, & Ortega E. (2008). Basketball game-related statistics that discriminate between teams' season-long success. *European Journal of Sport Science*, 8(6), 369-372.
- Jones E.S. (2016). Predicting outcomes of NBA basketball games. *Yayımlanmamış Yüksek Lisans Tezi*, North Dakota State University.
- Kaur H & Jain S. (2017). "Machine learning approaches to predict basketball game outcome". 3rd International Conference on Advances in Computing, Communication and Automation (Fall), ICACCA 2017.
- Kuhn M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1-26.
- Kuhn M & Johnson K. (2013). *Applied predictive modeling*. Vol 26 Springer, New York.
- Leicht A.S., Gómez M.A., & Woods C.T. (2017). Explaining match outcome during the men's basketball tournament at the olympic games. *Journal of Sports Science and Medicine*, 16(4), 468-473.
- Loeffelholz B, Bednar E, & Bauer K.W. (2009). Predicting NBA games using neural networks. *Journal of Quantitative Analysis in Sports*, 5(1).
- Lorenzo A, Gómez M.Á., Ortega E, Ibáñez S.J., & Sampaio J. (2010). Game related statistics which discriminate between winning and losing under-16 male basketball games. *Journal of Sports Science and Medicine*, 9(4), 664-668.
- Magel R & Unruh S. (2013). Determining factors influencing the outcome of college basketball games. *Open Journal of Statistics*, 3(4), 225-230.
- McCabe A & Trevathan J. (2008). "Artificial intelligence in sports prediction". *Fifth International Conference on Information Technology: New Generations (ITNG 2008)* (1194-1197).
- McComb D.G. (2004). *Sports in world history*. Routledge Taylor & Francis.
- Navega D, Coelho C, Vicente R, Ferreira M.T., Wasterlain S, & Cunha E. (2015). Ancestrees: ancestry estimation with randomized decision trees. *International Journal of Legal Medicine*, 129(5), 1145-1153.
- Oliver D. (2004). *Basketball on paper: rules and tools for performance analysis*. Potomac Books, Inc.
- Ozkan I.A. (2020). A novel basketball result prediction model using a concurrent neuro-fuzzy system. *Applied Artificial Intelligence*, 34(13), 1038-1054.
- Pai P.F., Chang Liao L.H., & Lin K.P. (2017). Analyzing basketball games by a support vector machines with decision tree model. *Neural Computing and Applications*, 28(12), 4159-4167.
- Rasouliyan L & Miller DP. (2006). *The logic and logistics of logistic regression*. Ovation Research Group, San Fransisco California, 1-14.
- Valenzuela R. (2018). Predicting national basketball association game outcomes using ensemble learning Techniques. *Yayımlanmamış Yüksek Lisans Tezi*, California State University.
- Smola A.J. & Schölkopf B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222.
- Tax N & Joustra Y. (2015). Predicting the Dutch football competition using public data: a machine learning approach. *Transactions on Knowledge and Data Engineering*, 10 (10), 1-13.
- Thabtah F, Zhang L, & Abdelhamid N. (2019). NBA game result prediction using feature analysis and machine learning. *Annals of Data Science*, 6(1), 103-116.
- Wiseman O. (2016). Using Machine Learning to Predict the Winning Score of Professional Golf Events on the PGA Tour. *Yayımlanmamış Yüksek Lisans Tezi*, National College of Ireland.