



FUTBOL MAÇLARI İÇİN TAKTİK TABANLI TAHMİN SİSTEMİ

HARUN ÖZMEN

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Doç. Dr. Erdinç. UZUN

2022

T.C.
TEKİRDAĞ NAMIK KEMAL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



FUTBOL MAÇLARI İÇİN TAKTİK TABANLI TAHMİN SİSTEMİ

HARUN ÖZMEN

ORCID: 0000-0002-4943-2432

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

YÜKSEK LİSANS TEZİ

Danışman: Doç. Dr. Erdinç UZUN

ŞUBAT-2022

Her hakkı saklıdır.

ÖZET

FUTBOL MAÇLARI İÇİN TAKTİK TABANLI TAHMİN SİSTEMİ

Harun ÖZMEN

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Doç. Dr. Erdiñ UZUN

Spor müsabakaları organizasyonları insanların zaman geçirdikleri önemli faaliyetler arasında yer almaktadır. Spor müsabakaları arasında futbol, dünyadaki en yüksek popülerlik oranına sahiptir. Birçok futbol müsabakası hem bölgesel bazda hem de uluslararası bazda düzenlenmektedir. Futbol maçlarının sonuçları futbol takımı taraftarları için çok önemlidir ve günlük yaşamlarında destekledikleri takımın sonuçlarına göre mutlu ya da üzgün olabilirler. Yapay zekâ ve makine öğrenmesindeki gelişmeler sonucunda birçok alanda yapılan tahminler futbol sonuçlarının tahminini de içermektedir. Ekonomik getirileri de olan futbol bahis sistemi düşünüldüğünde, futbol maç sonuçlarının tahmin edilmesi önem kazanmıştır. Bu çalışma, İtalya-Serie A, İngiltere-Premier Lig ve Alman-Bundesliga liglerindeki sonuçları tahmin etmek için bilindiği kadarı ile literatürde ilk kez sunulan taktik tabanlı bir tahmin yaklaşımı sunmaktadır. Bu yaklaşımda oyuncuların sahadaki pozisyonları ve piyasa değerleri web sayfalarından otomatik olarak alınmakta ve veri madenciliği yöntemlerine açık hale getirilmektedir. Veri madenciliği aşamasında Torbalama, Yinelemeli Sınıflandırıcı İyileştirici ve Derin Öğrenme gibi farklı yöntemler karşılaştırılmakta ve üç yöntemin farklı liglerde benzer sonuçlar verdiği görülmektedir. Veri madenciliği yöntemlerinde yaklaşık %60 doğruluk elde edilmektedir. Her üç lig için makine öğrenmesi yöntemlerinde kullanılan özellikler incelendiğinde takım sıralamasının maç sonuçlarını etkileyen en önemli özellik olduğu görülmüştür. Serie A'da Stoper ve Merkez Orta Saha özellikleri öne çıkarken, Premier Lig'de Sağ Kanat ve Kaleci gibi farklı özelliklerin önemli olduğu görülmüştür. Bundesliga'da Stoper ve Santrafor özelliklerinin maç sonuçlarına daha çok etki ettiği görülmüştür.

Anahtar Kelimeler: Veri madenciliği, Özellik mühendisliği, Tahmine dayalı analiz, Derin Öğrenme

ABSTRACT

A TACTIC-BASED PREDICTION SYSTEM FOR FOOTBALL MATCHES

Harun ÖZMEN

Department of Computer Engineering

MSc. Thesis

Supervisor: Assoc. Prof. Dr. Erdinç UZUN

Organizations of sports competitions are among the important activities that people spend their time on. Among the sports competitions, football has the highest popularity rate in the world. Many football competitions are organized both on a regional basis and on an international basis. The results of football matches are very important for the fans of the football team, and they can be happy or sad according to the results of the team they support in their daily lives. As a result of developments in artificial intelligence and machine learning, predictions made in many fields include the prediction of football results. Considering the football betting system, which also has economic returns, the prediction of football match results has become important. This study presents a tactical-based forecasting approach to be presented for the first time in the literature to our knowledge, to predict results in Italy-Serie A, England-Premier League and Germany-Bundesliga leagues. In this approach, players' on-field placements and market values are automatically obtained from web pages and made available for data mining methods. The data mining phase compares different methods such as Bagging, Iterative Classifier Optimizer and Deep Learning, and it is seen that the three methods give similar results in different leagues. An accuracy of about 60% is achieved in data mining methods. When the features used in machine learning methods are examined for all three leagues, it is the most important feature in predicting the team ranking. While Center-Back and Central Midfielder features stand out in Serie A, different features such as Right Winger and Goalkeeper are seen to be important in Premier League. In the Bundesliga, the Centre-Back and Centre-Forward features have had a greater impact on the matches.

Keywords: Data mining, Feature engineering, Predictive analysis, Deep Learning

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
İÇİNDEKİLER	iii
ÇİZELGELER DİZİNİ	v
ŞEKİLLER DİZİNİ	vi
SİMGELER DİZİNİ	vii
KISALTMALAR DİZİNİ	viii
TEŞEKKÜR	ix
1. GİRİŞ	1
1.1 Literatür Özeti	3
1.2 Çalışmanın Amacı ve Kapsamı	5
2. TAKTİK-TEMELLİ YAKLAŞIM	6
2.1 Web Verisinin İşlenmesi	6
2.2 Özellik Mühendisliği.....	10
2.3 Eğitim ve Test Veri Kümesi.....	13
2.4 Tahmine Dayalı Modeller	14
2.4.1 Naive Bayes	16
2.4.2 J48 Sınıflandırıcı.....	16
2.4.3 Derin Öğrenme	17
2.4.4 Torbalama	19
2.4.5 Yinelemeli Sınıflandırıcı İyileştirici	19
3. DENEYSEL SONUÇLAR	21
3.1 Veri Kümeleri Hakkında İstatistiksel Bilgiler	21
3.1.1 Taktik İstatistikler	21
3.1.2 Transfer Piyasası Değerleri Açısından İstatistikler	24
3.2 Performans Metrikleri	26
3.3 Metodların Performans Değerlendirmesi.....	28
3.4 Test Veri Kümesinde Tahmin Performanslarının Karşılaştırılması.....	29
3.5 Alt Test Veri Kümesinde Tahmin Performanslarının Karşılaştırılması.....	30
3.6 Özelliklerin Bilgi Kazançları	32
4. SONUÇ VE ÖNERİLER	37
KAYNAKLAR	39

EK-1. Serie A – Ev Sahibi Taktik ve Yüzdeleri	45
EK-2. Serie A – Deplasman Taktik ve Yüzdeleri	46
EK-3. Premier Lig – Ev Sahibi Taktik ve Yüzdeleri.....	47
EK-4. Premier Lig – Deplasman Taktik ve Yüzdeleri	48
EK-5. Bundesliga – Ev Sahibi Taktik ve Yüzdeleri.....	50
EK-6. Bundesliga – Deplasman Taktik ve Yüzdeleri	51
EK-7. Derin Öğrenme için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı	53
EK-8. Yinelemeli Sınıflandırıcı İyileştirici için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı.....	54
EK-9. Torbalama için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı.....	55
EK-10. Navie Bayes için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı.....	56
EK-11. J48 Sınıflandırıcı için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı	57
EK-12. Derin Öğrenme için Alt Test Veri Setinde Liglerin Tahmin Performansı.....	58
EK-13. Yinelemeli Sınıflandırıcı İyileştirici için Alt Test Veri Setinde Liglerin Tahmin Performansı.....	59
EK-14. Torbalama için Alt Test Veri Setinde Liglerin Tahmin Performansı	60
EK-15. Navie Bayes için Alt Test Veri Setinde Liglerin Tahmin Performansı	61
EK-16. J48 Sınıflandırıcı için Alt Test Veri Setinde Liglerin Tahmin Performansı.....	62
ÖZGEÇMİŞ	Hata! Yer işareti tanımlanmamış.

ÇİZELGELER DİZİNİ

Çizelge 2.1. Bu çalışmada kullanılan özellikler	11
Çizelge 2.2. Şekil 2.5'te verilen eşleşme özelliklerindeki değerler	15
Çizelge 3.1. 2014-2020 Serie A'da ev sahibi ve deplasman takımlarının tercih ettiği taktikler	21
Çizelge 3.2. 2014-2020 Yılları arasında Premiere Lig'de ev sahibi ve deplasman takımlarının tercih ettiği taktikler.....	23
Çizelge 3.3. 2014-2020 Bundesliga'da ev sahibi ve deplasman takımlarının tercih ettiği taktikler	24
Çizelge 3.4. 2014-2020 Yılları arasında üç ligde oyuncuların bölgelere göre ortalama transfer piyasa değerleri.....	25
Çizelge 3.5. 2014-2020 Yılları arasında üç ligdeki takımların ortalama transfer piyasa değerleri	25
Çizelge 3.6. Bir takım için karışıklık matrisi	26
Çizelge 3.7. Serie A- Derin Öğrenme için performans ölçümleri	29
Çizelge 3.8. Test veri seti için üç yöntemin ve üç ligin tahmin performanslarının karşılaştırması	30
Çizelge 3.9. Alt test veri seti için üç yöntemin ve üç ligin tahmin performanslarının karşılaştırması.....	31

ŞEKİLLER DİZİNİ

Şekil 2.1. Taktik tabanlı yaklaşımın akış şeması.....	6
Şekil 2.2. Bir haftanın maçları ve bir maçın URL bilgisi.....	7
Şekil 2.3. Maçın kadrolarına ulaşma	8
Şekil 2.4. Maç kadrosu ve bir futbolcunun verilerini içeren HTML örneği.....	9
Şekil 2.5. Bir futbol maçında ev sahibi ve deplasman takımı oluşumu örneği	13
Şekil 2.6. Derin Öğrenme ve Makine Öğrenmesinin sınıflandırma yaklaşımlarının karşılaştırılması	17
Şekil 2.7. Derin Öğrenme ve Makine Öğrenmesinin sınıflandırma yaklaşımlarının karşılaştırılması.....	18
Şekil 2.8. Bagging (Torbalama) algoritması çalışma şekli.....	19
Şekil 3.1. Serie A'daki en iyi özelliklerin bilgi kazanımı	33
Şekil 3.2. Premier Lig'deki en iyi özelliklerin bilgi kazanımı	34
Şekil 3.3. Bundesliga'daki en iyi özelliklerin bilgi kazanımı	35

SİMGELER DİZİNİ

C	Sınıf
F	Özellik
ρ	Olasılık Fonksiyonu
TP	Gerçek doğru
TN	Gerçek yanlış
FP	Sahte doğru
FN	Sahte yanlış



KISALTMALAR DİZİNİ

DL	Derin Öğrenme algoritması
B	Torbalama algoritması
ICO	Yinelemeli Sınıflandırıcı İyileştirici algoritması
NB	Navie Bayes algoritması
J48	j48 Sınıflandırıcı algoritması
D	Deplasman
E	Ev sahibi
N	Çizelge 3.1'deki tabloda verilen bölgelerdeki oyuncu sayısı



TEŐEKKÜR

Bu tezin her aŐamasında yardımlarını esirgemeyen zaman mefhumuna dikkat etmeden bu alıŐmada ortaya ıkan pozitif sonuların heyecanını paylaşmak iin gecenin bir yarısı aradıđımda mutluluđumu paylaşan ve destek veren ok kıymetli ve sayđı deđer hocam Do. Dr. Erdin Uzun'a (Tekirdađ Namık Kemal Üniuersitesi orlu Mühendislik Fakültesi) sonsuz Őükran ve minnetlerimi sunarım. İerik ve konu olarak ok farklı olan bu alıŐmanın ortaya ıkıŐ hikayesi tarafımda özel olsa da bu alıŐmada en büyük motivasyonum biricik kızım Ada Aysima Özman'e ok teşekkür ederim. Son olarak hayatımın her anında maddi ve manevi desteđini esirgemeyen anne ve babama sayđı ve hürmetlerimi sunarım.

Harun ÖZMEN

Bilgisayar Mühendisi

1. GİRİŞ

Veri Madenciliği, büyük veri kümelerini inceleyerek faydalı veri kalıplarını keşfetme ve bu veriler üzerinde uygulama ve analiz süreçleri ile faydalı bilgileri ortaya çıkarma sürecidir (Hand, 2012). Literatürde veri madenciliğinin uygulama alanları olarak finansal dolandırıcılığın tespiti (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011) (Zhou & Kapoor, 2011) ile bilgi ve iletişim teknolojisi alanındaki analizlerde (Kaur, Lechman, & Marszk, 2017) kullanılmıştır. Spor alanında ise basketbol (Štrumbelj & Vračar, 2012) (Vračar, Štrumbelj, & Kononenko, 2016), kriket (Asif & McHale, 2016) ve at yarışında (Lessmann, Sung, & Johnson, 2010) veri madenciliği ve analiz yönetimleri kullanılmıştır.

Makine Öğrenimi ve Yapay Zekâ, birçok spor aktivitesinin sonuçlarını tahmin etmek için kullanılmaktadır. Futbol dünya çapında en popüler spor aktivitesidir. Birçok futbol kulübü yerel liglerde ve uluslararası futbol turnuvalarında yarışmaktadır. Her yıl dünya çapında düzenli olarak birçok futbol turnuvası düzenlenmektedir (Örneğin UEFA Şampiyonlar Ligi, Copa Libertadores, CONCACAF Şampiyonlar Ligi). Birçok ulusal futbol takımı hem ulusal liglerde hem de uluslararası turnuvalarda mücadele etmektedir. Çalışmamızda Avrupa'daki üç büyük lig ele alınmıştır. Bu liglerden ilki; 212 bölgede 643 milyon eve, 4,7 milyar potansiyel TV izleyicisine yayın yapan ve ortalama 36.000 izleyici sayısı ile Bundesliga'dan sonra en yüksek seyirci ortalamasına sahip Premier Lig'dir (https://en.wikipedia.org/wiki/Premier_League, 2021). İncelediğimiz ikinci lig ise; UEFA tarafından belirlenen lig katsayılarına göre Avrupa'da üçüncü sırada yer alan İtalya Serie A olup IFFHS tarafından 2020 yılı için Avrupa'nın en güçlü ulusal ligi seçilmiştir (https://en.wikipedia.org/wiki/Serie_A, 2021). Üzerinde çalıştığımız üçüncü ve son lig ise; Dünya çapında 200'den fazla ülkede televizyonlarda yayınlanan Alman Bundesliga olup 2011-2012 verilerine göre maç başına ortalama 45.134 taraftar ile en fazla seyirciye sahip ligdir (<https://en.wikipedia.org/wiki/Bundesliga>, 2021).

Literatürde özellik, araştırma sırasında gözlemlenen bir olgunun nicel kalitesi için tanımlanmaktadır (Bishop, 2006). Etkili ve verimli makine öğrenimi ve örüntü tanımlama süreçleri için özellikler ayırt edici, anlaşılır ve bağımsız değerler içermelidir. Daha önce yapılan çalışmalarda futbol maç sonuçlarının tahmini için; Takımların hücum, orta saha, savunma ve genel değerlendirme özellikleri (Rahul Baboota, 2019) ve sezon sayısı, maç sayısı, takım sayısı ve oyuncu sayısı (Rudrapal, Boro, Srivastava, & Shyamu, 2019) gibi istatistiksel özellikler kullanılmıştır. Forrest & Simmons (2000) ise yaptıkları çalışmada The Times, Daily Mail ve The Mirror gazetelerindeki futbol tahmincilerinin verilerini alarak çevrimiçi bahis oranlarının

maç sonuçlarının tahminine etkisini incelemiştir. Transfermarkt web sitesindeki kullanıcı yorumları (Peeters, 2018) maç sonuçlarının tahmininde veri seti olarak analiz edilmiştir. Geçmişte yapılan diğer çalışmaların çoğunda araştırmaya konu olan takımların geçmiş maçlarının sonuçları veri seti olarak kullanılarak tahmin yapan modeller oluşturulmaya çalışılmıştır. Bizim çalışmamız ise hem oyuncu tabanlı hem de taktik tabanlı veri seti kullanarak maç sonuçlarının tahmin edilmesinde literatüre katkı sağlamayı amaçlamaktadır.

Bir futbol maçında üç olası sonuç vardır. Bunlar ev sahibi galibiyeti, deplasman galibiyeti ve beraberliktir. Üçlü sınıflandırma probleminin konusu olan bir maç sonucunun olası üç sonucunu tahmin etmede veri seti çok önemlidir. Bu çalışmada şimdiye kadar literatürde tartışılmayan bir veri seti incelenmiştir. Veri setimiz, bir maçın sonucunu etkileyen oyuncu faktörünü ve teknik direktör etkisini yansıtan taktik faktörünü içermektedir. Transfermarkt web sitesinden hem oyuncu hem de taktik bilgilerini içeren veriler bir web tarayıcısı yardımıyla otomatik olarak elde edilmiştir.

Bu tezin geri kalan kısmı şu şekilde organize edilmiştir: Bölüm 1.1, literatürde yapılmış ilgili diğer çalışmaları gösterir. Bölüm 2, taktik tabanlı yaklaşımda verinin toplanması, özelliklerin çıkarılması, veri setlerinin oluşturulması ve makine öğrenmesi metotları hakkında bilgi içerir. Ayrıca, bu bölüm maç sonuçlarının tahmininde kullanılan veri setinin taktik ve oyuncu bazında nasıl oluşturulacağını anlatır. Bölüm 3, veri seti hakkında taktiksel istatistikler ve futbolcuların piyasa değerleri hakkında ön bilgi verdikten sonra makine öğrenmesi metotları karşılaştırılır. Ek olarak, bu çalışmada kullanılan özelliklerin makine öğrenme modellerine katkısı incelenerek sonuçların tahmininde ise oluşturulan modellerin yeni verilerle test edilmesi ve sonuçların açıklanmasını içerir. Bölüm 4, çalışmanın sonuçları özetlendikten sonra çalışma hakkında ileriye yönelik yapılabilecek katkılar ile bu çalışmanın başka hangi alanlara katkısının olabileceği ile ilgili projeksiyon sunulur.

1.1 Literatür Özeti

1 (Bir) trilyon dolara yaklaşan futbol bahis piyasasının değeri, futbol maç sonuçlarını hem ekonomik hem de akademik olarak incelenmesini ve sonuçlarını tahmin etmeyi amaçlayan bir sistemin geliştirilmesini sağlamıştır (Rahul Baboota, 2019). Bu nedenle futbol maç sonuçlarının tahmini sadece akademik bir ilgi konusu olmayıp, ekonomik getirisi nedeniyle bahis sektöründeki birçok kişinin uğraştığı bir alandır. Literatürde daha önce yapılan çalışmalar sırasıyla sonuç odaklı ve skor bazlı çalışmalar olmak üzere iki ana grupta incelenmiştir. Skora dayalı çalışmalar her maç için maç skorunu tahmin etmeyi amaçlarken, sonuç bazlı çalışmalar her maç için ilgili maçın sonuç sınıfını tahmin etmeyi amaçlamaktadır.

Sonuç odaklı yaklaşımla yapılan çalışmada (Joseph, Fenton, & Neil, 2006), İngiltere Premier Ligi takımlarından sadece Tottenham Hotspurs takımının 1995-1997 yılları arasındaki maç sonuçlarını tahmin eden bir sistem geliştirilmiştir. Bu çalışmada, Uzman Bayes Ağlarının K-En Yakın Komşular (KNN), Navie Bayesian ve MC4 karar ağaçları gibi makine öğrenme sistemlerine kıyasla daha iyi sonuçlar verdiğini gösterdiler. %59,21 doğruluk oranı elde etmelerine rağmen çalışmalarında belirli bir dönemde sadece bir takıma odaklanmışlardır. Bayes ağlarını kullanan başka bir çalışma (Owramipur, Eskandarian, & Mozneb, 2013) tarafından yapılmıştır. Yaptıkları çalışmada Bayes ağları sistemini kullanarak İspanyol futbol kulüplerinden FC Barcelona takımının maç sonuçlarını tahmin ettiler. Çalışmalarındaki veri setlerinde ilk 11'de oynayan bir oyuncunun sakatlığı, futbolcuların psikolojik seviyesi ve hava koşulları gibi farklı özellikleri değerlendirdiler. %92 gibi çok yüksek bir doğrulukla tahminler sunan bir model geliştirmiş olmalarına rağmen, modelleri bir takım için bir sezonda sadece 20 maçın gözlemlerine dayanmaktadır. Bayes Ağlarına ilişkin bir diğer çalışma da (Fenton & Constantinou, 2013) tarafından daha önceden yayınladıkları çalışmalarını (Constantinou, Fenton, & Neil, 2012) geliştirilerek yapılmıştır. Çalışmalarında 2011-2012 sezonu için ELP ligi için maç sonuçları tahminlerini maçlardan önce çevrim içi olarak yayınlamışlardır. Ekonomik getiri bazlı yapmış oldukları çalışmalarında bir önceki çalışmalarına göre daha az kompleks bir model kurarak daha iyi kazanç elde edilebileceğini göstermişlerdir.

Dixon & Coles (1997) tarafından yapılan çalışma ise skor temelli bir çalışma olup İngiltere Premier Ligine dair 3 sezonluk süreç incelenmiştir. 1995 -1996 yılları arası bahis şirketlerinin oranlarını kullanmışlar ve futbol bahis pazarındaki oranların tutarsızlığını değerlendirerek verimli bir model oluşturmuşlardır. Veri setlerinde geçmiş yıllara dönük verilerden yararlanmışlardır. Kurmuş oldukları sistem yarışmacı ekiplerin dinamik doğası ve

veri yapısı itibari ile karmaşık olan teknik bir Poisson Regresyon modelidir. Çalışmalarında bahisçilerin vermiş oldukları oranlara göre olumlu sonuçlar getirecek bir bahis stratejisi geliştirmeyi hedeflemişlerdir. Goddard (2005) yapmış olduğu çalışmada gol temelli ve sonuç temelli bir modeli beraber incelemiştir. İngiltere Premier Lig ve diğer 3 alt lige dair geçmişe yönelik 15 sezonluk veri setlerini kullanarak maçlardan önce sıralı bir Probit Regresyon modeli kullanarak sonuçları tahmin etmiştir. En iyi sonucu ise sonuç temelli ve gol temelli hibrit modelden elde etmiştir. Çalışmalarında veri setlerinde yarışan iki takım arasındaki coğrafi mesafe, maçlardaki seyirci sayısı gibi çok farklı özelliklerin maç sonuçlarını etkilediğini göstermişlerdir. Ancak çalışmalarında maç sonuçlarının tahmininden ziyade ekonomik getiri ve sabit oranlı bahis piyasasının etkinliği üzerine durmuşlardır. Karlis & Ntzoufras (2010) yapmış oldukları çalışmalarında gol temelli yaklaşımı benimsemişler ve ağırlıklı olasılık yaklaşımı ile klasik maksimum olasılık yaklaşımını karşılaştırmışlardır. Veri seti olarak UEFA Şampiyonlar Ligi'nin 2008-2009 periyodunu incelemişlerdir. Kurmuş oldukları model ile müsabakalarda oluşacak aykırı değerlere karşı koruma oluşturmayı hedeflemişlerdir. Boshnakov, Kharra, & McHale (2017) ise çalışmalarında futbol maçları için skor dağılımını tahmin etmeye çalışmış ve iki değişkenli sayım modelini değerlendirmişlerdir. İngiltere Premier Lig'ine dair 2006/2007 sezonu ile 2016/16 sezonları arasındaki verileri kullanmışlardır. Ev sahibi ve deplasman takımlarının atmış oldukları gol sayılarını ele alarak iki değişkenli bir dağılım modeli oluşturmuşlar ve bu modeli Weibull İnter-Arrival Times ve Copula temelinde gerçekleştirmişlerdir. İki değişkenli Weibull sayım modelinin çok iyi sonuçlar verdiğini göstermişlerdir. Rue & Salvesen (2000) ise yapmış oldukları çalışmada gol temelli bir yaklaşımı incelemişlerdir. Gol temelli çalışmalarını takımların atak ve defansif özelliklerini göz önüne alarak zaman temelli incelemişlerdir. Bayesian Dynamic Generalized Linear model geliştirerek istatistiksel bir yaklaşım sergilemişlerdir. Markov Chain Monte Carlo tekniği kullanarak bir sezonu geriye doğru analiz etmişlerdir. Çalışmalarında aykırı maç sonuçlarının tespiti, takımların son sıralaması, bahis karı gibi çeşitli durumları belirleyen uygulamalarla genişletmişlerdir. Ancak çalışmaları İngiltere Premier Lig ve 1. Lig için sadece 1997-1998 sezonu ile sınırlı kalmıştır dolayısı ile genelleştirme için uygun olmayabilir. Crowder, Dixon, Ledford, & Robinson (2002) ise yapmış oldukları çalışmada (Dixon & Coles, 1997) 'in orijinal Stochastic Process modeline yeni bir yakın benzerlik getirerek öngörü gücüne yoğunlaşmadan daha izlenebilir bir hesaplama sağlamışlardır. Koopman & Lit (2015) ise çalışmalarında (Maher, 1982)'in yapmış olduğu çalışmayı temel alarak model geliştirmişlerdir. İngiltere Premier Ligine ait 9 yıllık geçmiş verileri kullanarak 2010-2011 ve 2011-2012 sezonlarının sonuçlarını tahmin etmişlerdir. Modellerinde iki değişkenli Poisson dağılımı ele

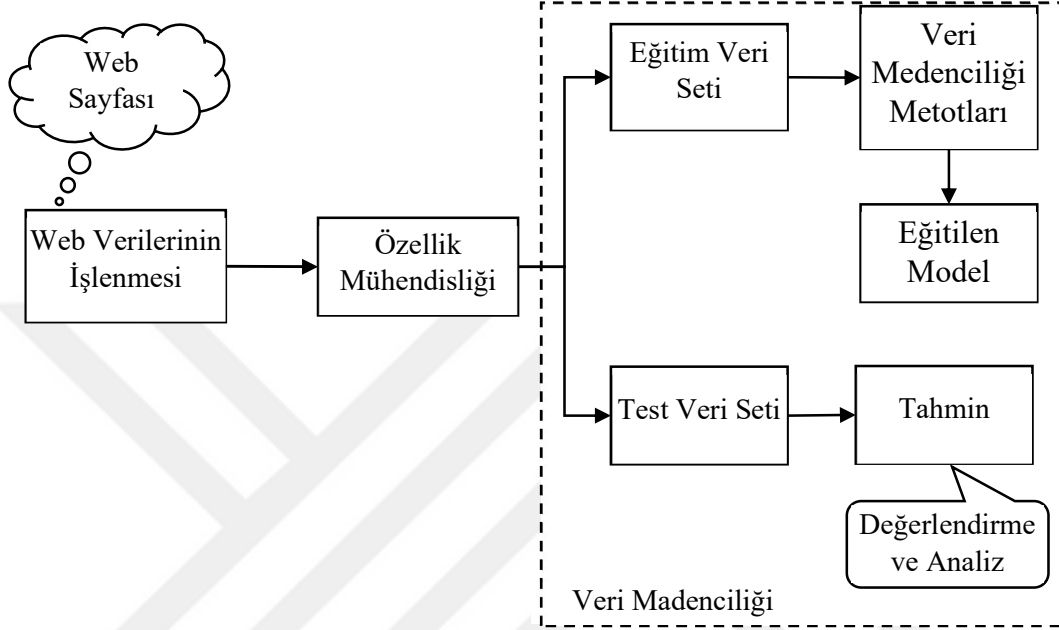
almışlar ve istatistiksel bir model önermişlerdir. Karşılaşan iki takım için atak ve defans güçlerine bağlı olarak zaman içerisinde değişen yoğunluk katsayılarını kullanmışlardır. Modellerinin bahisçilerin oranlarına göre dahi iyi getiri sağladığını göstermişlerdir.

1.2 Çalışmanın Amacı ve Kapsamı

Bu çalışma ile literatürde bilindiği kadarı ile hiç incelenmemiş olan taktik ve oyuncu bazlı bir veri seti oluşturularak maç sonuçlarının önceden tahmin edilmesi hedeflenmektedir. Çalışmamızda Avrupa’da ki üç büyük majör lig arasında yer alan Alman Bundesliga, İtalyan Serie-A ve İngiltere Premier Lig incelenmiştir. Her üç lig için birbirine yakın ve iyi sonuçlar elde edilerek çalışmamızın geliştirilebilir olduğu gösterilmiştir. Bu çalışma ile her ligin kendine has özelliklerinin ortaya çıkartılması hedeflenerek teknik direktörlere taktik seçiminde yardımcı olunması amaçlanmaktadır.

2. TAKTİK-TEMELLİ YAKLAŞIM

Bu yaklaşım web verisinin işlenmesi, özellik mühendisliği ve veri madenciliği olmak üzere üç ana parçadan oluşmaktadır. Şekil 2.1 bu parçaların ilişkilerini ve yaklaşımın akışını gösterir.



Şekil 2.1. Taktik tabanlı yaklaşımın akış şeması

Şekil 2.1'de, Web Veri İşleme modülü, transfermarkt.com'dan web sayfalarını indirir ve verileri çıkarmak için bunları ayrıştırır. Bu modül sayesinde istenilen ligi ve sezonunun bilgileri toplanmaktadır. Özellik Mühendisliği modülü, Veri Madenciliği modülünde kullanılmak üzere özellikleri hazırlar. Ayrıca bu modül bir eğitim veri seti ve test veri seti oluşturmak için kullanılır. Veri Madenciliği modülünün iki ana görevi vardır:

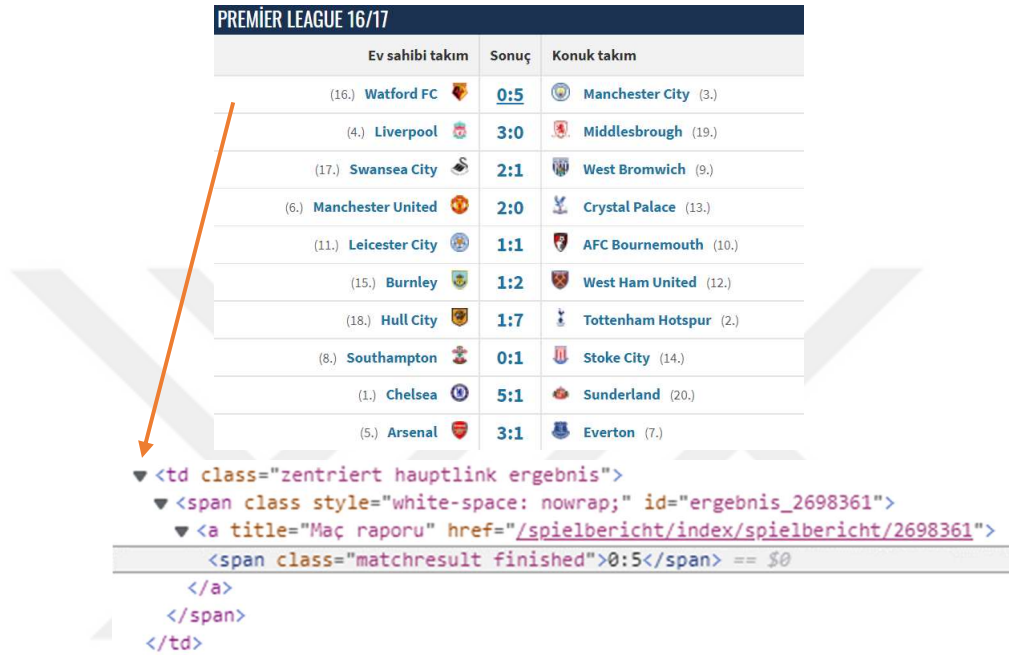
1. Bir eğitim veri seti kullanarak bir öğrenme modeli oluşturmak
2. Öğrenilen modeli kullanarak test verilerinden futbol maçlarını tahmin etmek.

2.1 Web Verisinin İşlenmesi

Web verilerini işlemek için bir Web Crawler geliştirilmiştir. Bu Web Crawler'a haftanın maçları bir URL (Uniform Resource Locator) girdisi olarak verilir. URL internette bir kurala uygun web sayfalarına ulaşmak için kullanılan bir karakter düzenidir. Örneğin, Premier Lig için URL girdisi:

https://www.transfermarkt.com.tr/premierleague/spieltagtabelle/wettbewerb/GB1?saison_id=2016&spieltag=38

şeklindedir. Burada “saison_id” maç dönemini verirken “spieltag” hafta bilgisini verir. Bu şekilde farklı sezonlara ve farklı dönemlere ulaşılabilir. “premier-league” ifadesi değiştirilerek farklı liglere ulaşılabilir. Bu URL ’nin döndürdüğü sonuç Şekil 2.2’de verilmiştir.



Ev sahibi takım	Sonuç	Konuk takım
(16.) Watford FC	0:5	Manchester City (3.)
(4.) Liverpool	3:0	Middlesbrough (19.)
(17.) Swansea City	2:1	West Bromwich (9.)
(6.) Manchester United	2:0	Crystal Palace (13.)
(11.) Leicester City	1:1	AFC Bournemouth (10.)
(15.) Burnley	1:2	West Ham United (12.)
(18.) Hull City	1:7	Tottenham Hotspur (2.)
(8.) Southampton	0:1	Stoke City (14.)
(1.) Chelsea	5:1	Sunderland (20.)
(5.) Arsenal	3:1	Everton (7.)

```
<td class="zentriert hauptlink ergebnis">
  <span class style="white-space: nowrap;" id="ergebnis_2698361">
    <a title="Maç raporu" href="/spielbericht/index/spielbericht/2698361">
      <span class="matchresult finished">0:5</span> == $0
    </a>
  </span>
</td>
```

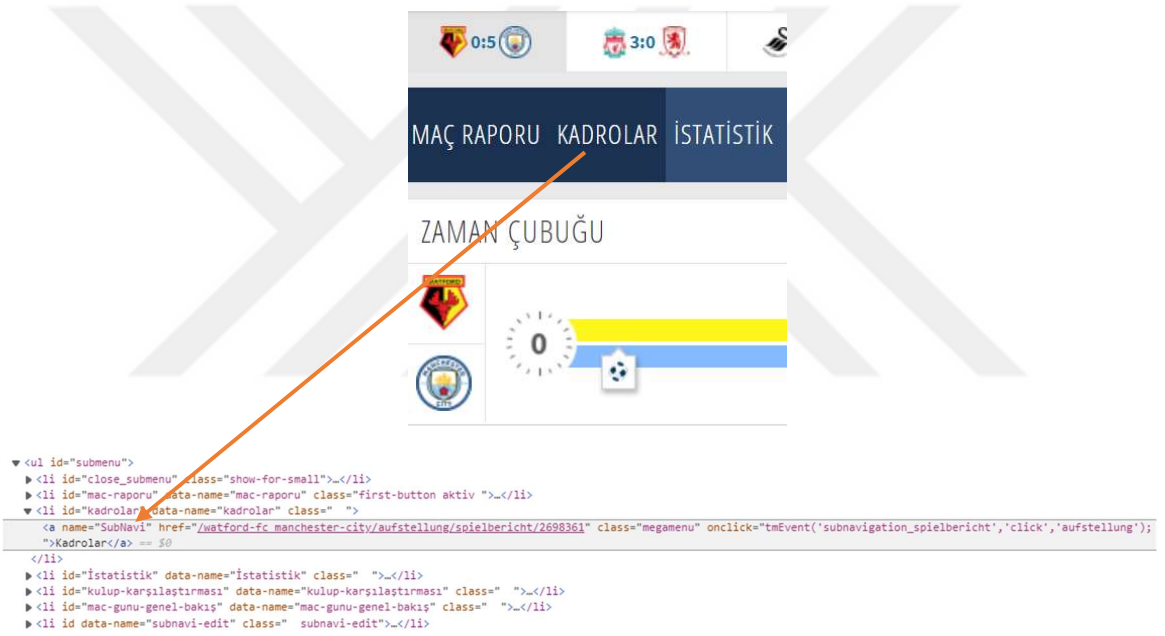
Şekil 2.2. Bir haftanın maçları ve bir maçın URL bilgisi

Bu aşamadan sonra maçlara ait URL verileri toplanır. Burada URL verileri toplanırken öncelikle DOM (Document Object Model) oluşturulur. DOM oluşturulduktan sonra, etiketler arasındaki içeriğe erişmek için CSS (Cascading Style Sheets) Seçicileri (Selectors) (UZUN E. , 2020)veya XPATH (YERLİKAYA, KIRAT, & UZUN, 2018) gibi çıkarma yöntemleri kullanılabilir. CSS Seçiciler, çıkarmak istediğiniz içeriği seçmek için kullanılan kalıplardır. XPATH, içeriğe erişmek için " path like" sözdizimini kullanır. Örneğin:

- td.zentriert.hauptlink.ergebnis a
- //td[@class="zentriert hauptlink ergebnis"]/span/a

CSS Seçiciler ve XPATH, HTML sayfasında bulunan bir elemente ulaşmak için kullanılan standartlardır. Her iki yönteminde birbirine benzerlikleri vardır. XPATH klasör mantığı gibi bir

erişim söz konusu iken CSS Seçiciler daha sade bir kullanıma sahiptir. Örneğin, maç raporuna ulaşmak için kullanılan yukarıdaki CSS Seçicide, td etiketinin class değeri “zentriert.hauptlink.ergebnis” üzerinden bu etikete ulaşılır. Class değeri CSS Seçicide “.” İle ifade edilir. “a” değeri ise “a” elementini döndürür. “a” elementi web sayfaları arasında geçişi sağlamak yani link vermek için kullanılır. Bu element içindeki “href” etiketi linkin kaynak verisini yani URL’yi verir. Böylelikle bu haftaya ait 10 maç verisinin linkine erişilmiş olur. “spielbericht/index/spielbericht/” içeren linke erişildiğinde futbolcuların tüm verileri gelmez. Bunun için “Kadrolar” linkine tıklanması gerekir. Geliştirdiğimiz Crawler, yine CSS Seçiciler kullanılarak buradaki URL verisine ulaşır.



Şekil 2.3. Maçın kadrolarına ulaşma

“Kadrolar” içindeki linke ulaşmak için

- li#kadrolar a.megamenu

CSS Seçicisi kullanılır. Burada “li” “ul” etiketinin bir ögesidir. Menu oluşturmak için kullanılmıştır. “#” id özelliği üzerinden bu element ulaşmamızı sağlar. Burada id değeri “kadrolar”dır. “a” link verisini tutan elementin “class” değeri “megamenu” olan elementine ulaşılır. Buradan kadrolara ait bilgileri veren link verisine ulaşılabilir. Erişilen yeni linkte hem ev sahibi hem de deplasman takımının kadrolarına elde edilebilir.

- Ev sahibi takım verisi: div.sb-team.sb-heim
- Deplasman takımı verisi: div.sb-team.sb-gast

Şekil 2.4, kadroları ve bir futbolcunun HTML verisini sunar.

İLK ONBİR		İLK ONBİR	
1	Heurelho Gomes (36 yaşında) Kaleci, 1,00 mil. €	13	Willy Caballero (35 yaşında) Kaleci, 2,00 mil. €
25	Jose Cholevas (32 yaşında) Stoper, 2,00 mil. €	4	Vincent Kompany (31 yaşında) Stoper, 25,00 mil. €
11	Valon Behrami (32 yaşında) Stoper, 2,50 mil. €	30	Nicolás Otamendi (29 yaşında) Stoper, 30,00 mil. €
32	Brandon Mason (19 yaşında) Sol Bek, -	22	Gaël Clichy (31 yaşında) Sol Bek, 8,00 mil. €
22	Daryl Janmaat (27 yaşında) Sağ Bek, 7,50 mil. €	25	Fernandinho (32 yaşında) Sağ Bek, 25,00 mil. €
14	Abdoulaye Doucouré (24 yaşında) Önlibero, 6,00 mil. €	42	Yaya Touré (34 yaşında) Önlibero, 10,00 mil. €

```

<table class="inline-table">
  <tbody>
    <tr>
      <td rowspan="2"></td>
      <td>
        <a title="Heurelho Gomes" class="wichtig" href="/heurelho-gomes/leistungsdatendetails/spieler/19059/saison/2016/wettbewerb/G81">Heurelho Gomes</a> == 58
        <span title="Takım kaptanı" class="kapitaenicon-table icons_sprite">&nbsp;</span>
        " (36 yaşında ) "
      </td>
      <td>Kaleci , 1.00 mil. €</td>
    </tr>
  </tbody>
</table>

```

Şekil 2.4. Maç kadrosu ve bir futbolcunun verilerini içeren HTML örneği

Şekil 2.4, tablo etiketi bir HTML tablosunu tanımlar. Bu etiket, bir tablo satırını belirten bir veya daha fazla tr etiketinden ve bir tablo hücrelerini belirten bir veya daha fazla td etiketinden oluşur. Ne yazık ki buradaki yapı CSS Seçici veya XPATH için uygun değildir. Örneğin “Mevki (market değeri)” şeklinde bir desen vardır. Kaleci için bu desen "Kaleci, 1,00 mil. €" şeklindedir. Burada çözüm, Düzenli İfadeler (Regular Expressions) kullanmaktır. Düzenli ifade, bir arama düzenini belirten bir dizi karakterdir. Örneğin:

- Regular Expression: (.*)€.€(.*)m

Yukarıdaki ifadede normal parantezler, nokta (.), soru işareti (?) ve yıldız işareti (*) özel karakterlerdir. Parantezler yakalama grubunuzu oluşturur. Bu örnekte, gruba, bir futbolcunun konumuna ve bir virgül karakterine kadar işaret etmektedir. Grupta herhangi bir karakterle eşleşmek için nokta (.) kullanılır. Yıldız işareti (*) sıfır veya daha fazla oluşumla eşleşir. Son olarak, soru işareti (?) sıfır veya bir örneği eşleştirmek anlamına gelmektedir. İkinci grup, bir oyuncunun piyasa değeri olan € karakteri ile başlar ve m karakteri ile biter.

Bu şekilde 5 sezonluk veriler Premier Ligi, Serie A ve Bundesliga için toplanmıştır. Toplanan verileri özellik çıkarımları yapıp makine öğrenmesi metotları için uygun hale getirilmiştir.

2.2 Özellik Mühendisliği

Literatürdeki önceki çalışmalar, maç sonuçlarının tahmininde kullanılan modelin başarısının verilerden elde edilen özelliklere bağlı olduğunu kanıtlamakta olup (Joseph, Marco , Nelson, Sears, & Tygar, 2006) (Owramipur, Eskandarian, & Mozneb, 2013). Rue & Salvesen (1997) bir maçın sonucunu etkileyen faktörler arasında ev sahibi takım ve deplasman takımı için hücum ve savunma gücünün öneminden bahseder. Aynı şekilde (Goddard, 2005) ise yapmış olduğu çalışmada bir futbol maçının sonucuna karşılaşan iki takımın geçmişe yönelik defans ve atak güçlerinin etki ettiğini göstermiştir. Rahul Baboota (2019) ise karşılaşan iki takımından ev sahibi ve deplasman takımları için atak, orta saha, defans ve genel gibi özellikleri baz alarak bir model kurmuştur. Peeters (2018) ise Transfer Market verileri ile bahisçilerin vermiş oldukları oranları baz alarak bir model kurmuştur.

Ancak bu istatistiki veriler bir futbol karşılaşmasında yarışan iki takım için realisttik bir bakış açısı değildir. Çünkü ev sahibi takım ile deplasman takımlarının geçmişe yönelik atak, defans, orta saha ve genel gibi özellikleri tahmini yapılan maç için gerçekçi değildir. Ya da Transfer Market veya Twitter gibi kaynaklardan elde edilen kullanıcıların yorumları baz alınarak kurulan modeller ilgili maç için gerçekçi bir özellik olamaz.

Bu çalışma, karşı karşıya gelen iki takımın (ev ve deplasman) maç sırasındaki futbol oyununun mevcut özelliklerine dayanmaktadır. Bir futbol maçında her takımın sahada 11 oyuncusu vardır. Her bir oyuncunun maça veya takıma olan etkisi farklıdır. Yaklaşımımızda, bir oyuncunun takımı adına maç sonuçları üzerindeki etkisi, oyuncuların tahmini oyuncu değerleri ile temsil edilir. Bu tahmini oyuncu değerleri, oyuncunun performansına göre genellikle birkaç ayda bir yenilenir (Simmons, Bryson, & Frick, 2013). Örneğin, bir orta saha oyuncusu sezon başında 6,00 milyon Euro'luk bir transfer değerine sahipken, bu oyuncunun ilerleyen haftalarda takımı adına gösterdiği yüksek performans, oyuncunun değerini 10,00 milyon Euro'ya yükseltebilir. Aynı şekilde sezon başında 10,00 milyon Euro değerinde olan bir forvetin transfer değeri de kötü performansı nedeniyle ilerleyen haftalarda 7,50 milyon Euro'ya düşebilir. Bu nedenle modelimizde tahmin edilecek bir futbol maçının sonucunu etkileyecek

faktörlerden biri olan oyuncu kalitesi, transfermarkt.com web sitesinden elde edilen mevcut oyuncu transfer değeri ile belirlenmektedir.

Çizelge 2.1. Bu çalışmada kullanılan özellikler

Grup İsmi	Özellik İsmi	Özellik (Kısaltma)	Açıklama
Geleneksel Özellikler	Team ID	Team_Id	Benzersiz Takım Numarası
	Ranking	Rank	Bir Takımın Ligindeki Sıralaması
	MatchDay	MDay	Hafta Takım Sayısı (1-38/1-34)
Alandaki Bölgeler	Rol ismi	Deplasman/Ev Özellikler	
Bölge 0	Kaleci	D ¹ / E ² GK	N ³ = 1
	Stoper	D ¹ / E ² CB N	N ³ = 1..5
Bölge 1	Sağbek	D ¹ / E ² RB N	N ³ = 1..4
	Solbek	D ¹ / E ² LB N	N ³ = 1..2
	Libero	D ¹ / E ² FB N	N ³ = 1..1
Bölge 2	Önlibero	D ¹ / E ² DM N	N ³ = 1..4
	Merkez Orta Saha	D ¹ / E ² CM N	N ³ = 1..4
Bölge 3	Orta Saha Sağ	D ¹ / E ² RM N	N ³ = 1..3
	Orta Saha Sol	D ¹ / E ² RM N	N ³ = 1..3
	Orta Saha	D ¹ / E ² FM N	N ³ = 1..1
Bölge 4	On Numara	D ¹ / E ² AM N	N ³ = 1..3
	Sağ Kanat	D ¹ / E ² RW N	N ³ = 1..4
	Sol Kanat	D ¹ / E ² LW N	N ³ = 1..4
Bölge 5	Santrafor	D ¹ / E ² CF N	N ³ = 1..5
	Forvet Arkası	D ¹ / E ² SS N	N ³ = 1..2

¹Deplasman, ²Evsahibi,

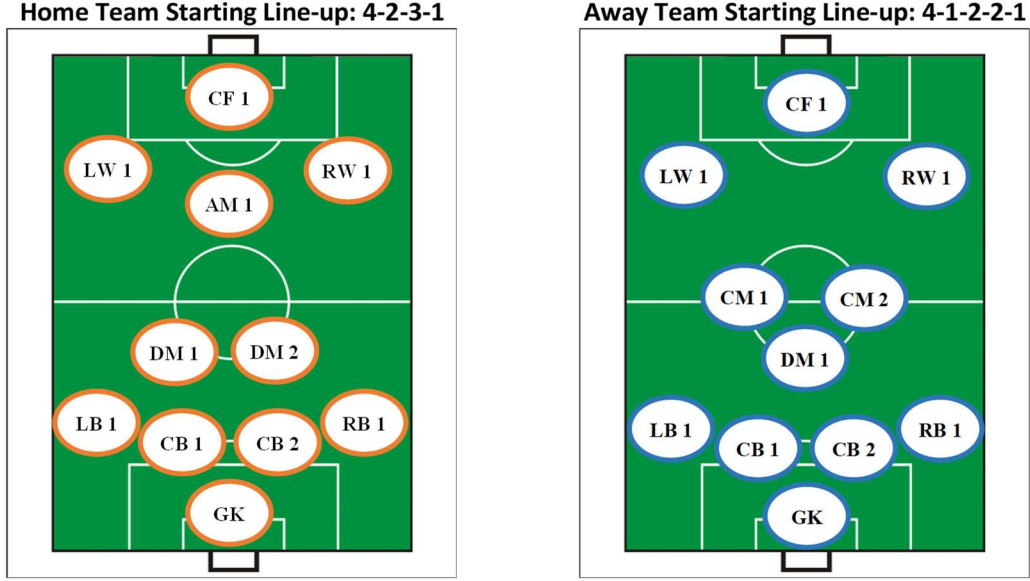
³Bu roldeki oyuncu sayısı

Veri kümemizde üç geleneksel özellik vardır. Team_Id, her takım için verilen benzersiz bir değerdir. Matchday özelliği, maçların planlandığı gün anlamına gelir. Ranking özelliği, belirli bir maç gününde takımların sıralamasıdır. Çalışmamızda yer alan diğer özellikler ise futbolcuların pozisyon ve piyasa değerlerini içeren özelliklerdir. Tüm özellikler her iki takım için ayrı ayrı hazırlanmıştır. Çizelge 2.1, çalışmamızdaki tüm özellikleri tanıtmaktadır.

Çizelge 2.1'de, futbol sahası, kaleci ve forvet oyuncularını arasında, santrafor/ lar ve/veya forvet arkası/ ları dahil olmak üzere 6 bölgeye ayrılmıştır. Bölge 1, sağda, solda ve ortada oynayan defans oyuncularını kapsar. Bu oyuncuların görevi, oyun sırasında diğer takımın

ataklarını durdurmaktadır. Genelde stoper pozisyonunda iki oyuncu kullanılırken, bazı takımların 5 oyuncuyla sahaya çıktığı bile görülmektedir. Bu nedenle yaklaşımımız stoper pozisyonu için 5 özelliğe (CB 1, CB 2, CB 3, CB 4, CB 5) sahiptir. Bölge 2, daha çok defans sağlayan aynı zamanda da atağı destekleyen ön libero/lar pozisyonundan oluşmaktadır. Bölge 3, merkez orta saha oyuncusu, orta saha sağ oyuncusu, orta saha sol oyuncusu ve serbest orta saha oyuncusu dahil olmak üzere dört pozisyon içerir ve pozisyonları atak ve defans oyuncularının ortasındadır. Bölge 4, diğer oyuncular için puan şansı yaratmak amacıyla özellikle Bölge 5'i destekler. Bölge 4, on numara oyuncusu, sağ kanat oyuncusu, sol kanat oyuncusu olmak üzere üç pozisyonu kapsar. Son olarak, Bölge 5'in santrafor ve forvet arkası olmak üzere iki pozisyonu vardır. Çizelge 2.1 bölgeleri göstermekte ve bu pozisyonlarda oynayacak minimum ve maksimum oyuncu sayısını vermektedir. Yani Çizelge 2.1'de hem ev sahibi hem de deplasman takımlarının verilerini içeren 98 özellik vardır. Bir pozisyonda oyuncu varsa oyuncunun piyasa değeri bu özelliğe veri olarak yazılır. Aksi takdirde, bu özelliğe 0 değeri verilir.

Maç sonucu açısından futbolcunun kalitesi en önemli faktörlerden biri olsa da menajerin etkisi de önemli faktörlerden biridir. Bir menajerin futbol maçlarının sonuçlarına etkisini özellik olarak göstermek çok zordur. Menajer, maçlar için planlamanın kritikliğini yansıtan bir taktik verir. Taktik, bir takımın oluşumunu birbirine bağlayan bir yöneticinin sistemini gösterir. Çizelge 3.1'de verilen 1-2-3-4-5 bölgeleri düşününce takım oluşumunda kaleci dikkate alınmaz. Daha önceki çalışmalarda bir futbol maçının sonucunu etkilediği düşünülen, atak ve savunma gücü özellikleri, aslında bir menajerin etkisini ortaya çıkarmayı amaçlamaktadır. Ancak literatürde bu özellikler ilgili takımların maç başına atılan gol sayısı ve beraberlik oranları gibi tarihsel istatistik verileri kullanılarak hesaplanmaya çalışılmıştır. Bu çalışmada, belki de en yenilikçi yaklaşım, ekiplerin sahada sıralanma şeklini gerçekçi bir şekilde yansıtan bir model sunmasıdır. Bir futbol maçında 11 kişilik kadroların her biri için oynanabilecek pozisyonların dağılımı her takım için nettir. Ancak bu dağılım menajerin seçtiği taktiğe göre değişir. Bir futbol maçında takımların diziliş farklılıkları dikkate alınır. Örneğin hücum futbolu için 4-2-3-1 (kaleci hariç) veya 4-3-3 varyasyonları kullanılır (TIGHE, 2012). 4-2-3-1'in oluşumunda Bölge 1, Bölge 2, Bölge 4 ve Bölge 5 için sırasıyla bir futbol sahasında 4, 2, 3 ve 1 oyuncudan oluşması anlamına gelmektedir. Şekil 2.5, iki karşıt takım için iki farklı dizilişi gösterir. Sahadaki oyuncu sayısı ve oynayabilecekleri pozisyon sayısı sabit iken, menajer etkisi olan varyasyon özelliğini veri setine yansıtmanın çok zor olduğunu daha önce belirtilmiştir.



Şekil 2.5. Bir futbol maçında ev sahibi ve deplasman takımı oluşumu örneği

Bu duruma dair yeni çözümümüz Çizelge 2.1'de gösterilmiştir. Çizelge 2.1'de her iki takımın oluşumu için 92 özellik olmasına rağmen, 22 özelliğin oyuncular için transfer piyasa değeri vardır. Diğer tüm özellikler veri olarak sıfır içermektedir.

2.3 Eğitim ve Test Veri Kümesi

Eğitim ve test veri setleri¹ bir Web Veri İşleme modülü ile transfermarkt.com'dan alınmıştır. Eğitim veri seti 2014-2019 arasındaki yılları içerir ve her lig için 5 sezonu kapsamaktadır. Test veri seti için, eğitim veri setinden elde edilen modelin değerlendirilmesi amacıyla her lig için 2019-2020 sezonunun verileri kullanılmıştır. Çizelge 2.1'de gösterilen özellik veri setine dayanarak, bir maçta ev sahibi ve deplasman takımlarının her birinin ilk 11 oyuncusu ve bu oyuncuların pozisyon bilgilerine göre mevcut oyuncular için transfermarkt.com web sitesinde güncellenen bilgi değerleri Euro para birimi olarak veri setine aktarılmıştır. Veri kümesine dahil olmayan pozisyonlar için 0 değeri atanmıştır. Başka bir deyişle ev sahibi ve deplasman takımlarının ilk 11 pozisyonu Transfemarket sitesindeki güncel para birimi değerleri ile veri setine yansıtılırken, veri setinde gösterilen ancak ilk 11'de yer almayan pozisyonlara 0 değeri atanmaktadır. Bu şekilde, her pozisyonu kapsayacak bir veri seti oluşturulmuş ve sistem optimal hale getirilmiştir. Bir futbol maçında oyuncu kalitesi çok önemli olmakla birlikte takımların oyun stratejisini belirleyen menajer farkı da çok önemlidir. Literatürde yeterince

¹ Bu veri setine <https://adys.nku.edu.tr/Datasets/> adresinden ulaşılabilir.

bahsedilmeyen bu farklılık bu çalışmada kendini göstermektedir. Örneğin, oyuncu kalitesi açısından hemen hemen aynı değerlere sahip iki takım arasında oynanan bir futbol maçında, oyunun sonucunu belirleyen en önemli faktör, takımların seçtikleri taktik anlayışıdır. Bu taktik anlayış oyuncuların sahadaki pozisyonları ile ortaya çıkmaktadır. Bu çalışmada geliştirilen sistem ekiplerin saha dizilişini içermektedir. Çizelge 2.2, Şekil 2.5'te var olan oyuncu pozisyonları değerlerini içermektedir. Şekil 2.5 ve Çizelge 2.2 beraber incelendiğinde bir futbol maçının sonucunu doğrudan etkileyen iki önemli faktörü bu çalışmada kurulan sistemle birlikte veri setimizde ele aldığımız görülmektedir. Bu faktörler sırasıyla:

- Oyuncu kalitesi: Bir maçtaki oyuncuların kalitesi, maçın sonucunu doğrudan etkileyen en önemli faktörlerden biridir. Bu çalışmadaki yaklaşımda, bir oyuncunun kalitesi, Transfermarkt web sitesindeki Euro para birimi olan mevcut değeri veri setimize yansıtılmıştır. Çünkü bir futbolcunun değeri performansına göre değişmektedir.
- Takımların Taktik Oluşumu: Futbol maçlarının sonucunu etkileyen diğer en önemli faktör menajerler tarafından seçilen oyun planıdır. Menajerlerin bir futbol maçının sonucuna etkisi, takımlarını sahada hangi taktik yaklaşımına sıraya koydukları ile ortaya çıkmaktadır. Ancak maç sonuçlarını tahmin eden bir sistemde futbol taktik anlayışını göstermek çok zordur. Bu çalışma ile Şekil 2.5'te gösterilen takımların saha oluşumları, diğer bir deyişle taktikleri, Çizelge 2.2'de gösterildiği gibi veri setimize yansıtılmaktadır. Bu çalışmadaki yaklaşım bir futbol takımının taktik saha oluşumu ile ilgili literatüre katkı sağlamayı amaçlamaktadır.

2.4 Tahmine Dayalı Modeller

Literatürdeki önceki çalışmaların çoğu, üçlü sınıflandırma problemlerini ikili sınıflandırmaya dönüştürerek sonuçları tahmin etmeye odaklanmıştır (Baboota & Kaur, 2019). Bu çalışmalarla kurulan sistemler için genelleme yapılamadığından gerçek futbol maç sonucu tahminlerinde her lig için kullanılamaz. Üçlü sınıflandırma konusunu ele alan çalışmamız gerçek dünyaya uyarlanabilir şekilde tasarlanmış olup 2015-2019 yılları arasında İngiltere Premier Ligi, İtalya Serie A ve Almanya Bundesliga ulusal liglerini içeren eğitim veri setinden oluşmaktadır.

Çizelge 2.2. Şekil 2.5'te verilen eşleşme özelliklerindeki değerler

Mday=6	
E Team Id=5, D Team Id=3	
E Rank=19, D Rank=1	
Bölge Adı	Özellikler ve Değerleri
Bölge 0	E GK = 20m, D GK = 65m
	E CB 1 = 5m, D CB 1 = 40m
	E CB 2 = 7m, D CB 2 = 27m
	E CB 3 = 0, D CB 3 = 0
Bölge 1	...
	E RB 1 = 3m, D RB 1 = 12m
	E LB 1 = 10m, D RB 2 = 8m
	...
	E FB = 0, D FB = 0
	E DM 1 = 5m, D DM 1 = 8m
Bölge 2	H DM 2 = 12m, A DM 2 = 0
	...
Bölge 3	E CM 1 = 0, D CM 1 = 17m
	E CM 2 = 0, D CM 2 = 8.5m
	...
	E RM 1 = 0, D RM 1 = 0
	...
	E LM 1 = 0, D LM 1 = 0
Bölge 4	...
	E FM = 0, D FM = 0
	E AM 1 = 15m, D AM 1 = 0
	...
	E RW 1 = 4m, D RW 1 = 13m
	...
Bölge 5	E LW 1 = 14m, D LW 1 = 18m
	...
	E CF 1 = 4m, D CF 1 = 50m
Bölge 5	...
	E SS 1 = 0, D SS 1 = 0
	...

Bu çalışmadaki yaklaşım, her takım için üç olası sonucu (Kazanma, Beraberlik, Kaybetme) öngörür. Çalışmadaki yaklaşımda Derin Öğrenme, Torbalama ve Yinelemeli Sınıflandırıcı İyileştirici dahil olmak üzere üç popüler veri madenciliği modeli ile Navie Bayes ve J48 Sınıflandırıcı kullanılmıştır.

2.4.1 Naive Bayes

Naive Bayes sınıflandırıcı algoritması sisteme verilen özelliklerin sistemde var olan sınıflardan bağımsız olduğunu varsayımı üzerine hareket ederek öğrenmeyi büyük ölçüde kolaylaştırmakta ve bağımsızlık bakış açısı genellikle zayıf varsayım olsa dahi uygulamada Naive Bayes genellikle daha karmaşık sınıflandırıcılara nazaran iyi sonuçlar verebilmektedir (Rish, 2001).

Naive Bayes Sınıflandırıcı algoritması birçok özellik kullanarak bir adet hedef değişken yani sınıf bulmayı amaçlar.

Naive Bayes sınıflandırıcı algoritması formülü:

$$\rho(C/F_1, \dots, F_n) = \frac{\rho(C)\rho(F_1, \dots, F_n/C)}{\rho(F_1, \dots, F_n)} \quad (1.1)$$

şeklindedir.

Yukarıda verilen formülde C verilen hedef yani sınıfı temsil ederken F özelliklerimizi temsil eder. Naive Bayes sınıflandırıcı verilen tüm koşullu olasılıkların çarpımından elde edilen sonuçtur.

Bizim modelimizde Naive Bayes Algoritması için batchsize: 380 seçilirken diğer özellikler default özellikler olarak kalmıştır

2.4.2 J48 Sınıflandırıcı

C4.5 karar ağacı algoritmasının bir uygulaması da J48 sınıflandırıcı algoritmasıdır (Quinlan J. , 1993). J48, eğitim setindeki öznitelik değerlerini kullanarak bir karar ağacı oluşturur ve sisteme yeni bir örnek verildiğinde sınıflandırır. Eğitim sırasında algoritma çeşitli örnekleri en iyi şekilde sınıflandıracak niteliği belirler. Özellik değerlerinin belirsizliği olmadığı durumlarda algoritma ilgili dalı sonlandırarak atama işlemini gerçekleştirir (Panigrahi & Borah, 2018).

Bizim modelimizde J48 Sınıflandırıcı Algoritması için batchsize: 380 seçilirken diğer özellikler default özellikler olarak kalmıştır

2.4.3 Derin Öğrenme

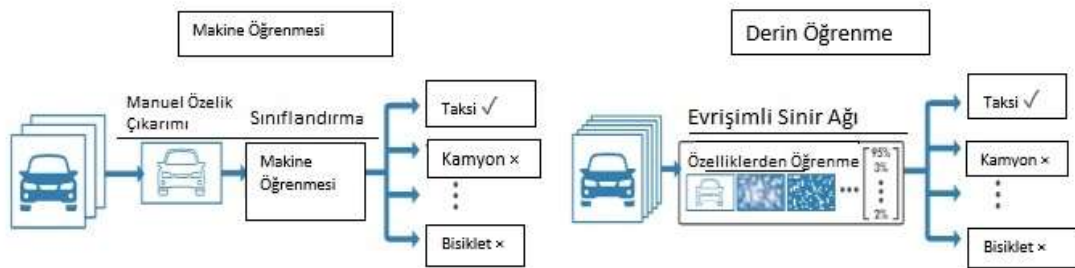
Veri kümesindeki eşleşme sonuçlarını tahmin etmek için kullanılan diğer model ise Derin Öğrenme modelidir. Literatürde genellikle görüntü analizinde (Schmidhuber, Meier, & Cireşan, 2012) (Krizhevsky, Sutskever, & Hinton, 2017) (Litjens, ve diğerleri, 2017) kullanılan Derin Öğrenme modeli, yapay sinir ağlarına dayanmaktadır. Bu çalışmada, Deeplearning4j (Gibson, ve diğerleri, 2016) Java kütüphanesine bağlı olan WekaDeeplearning4j (Lang, Bravo-Marquez, Beckham, Hall, & Frank, 2019) adlı bir derin öğrenme paketini kullanılmıştır.

Derin Öğrenme Makine Öğrenmesinin bir alt dalıdır. Makine Öğrenmesi ise Yapay Zekanın bir alt dalıdır. Dolayısı ile Derin Öğrenme Yapay Zekanın bir alt dalıdır. Literatürde genelde görüntü analizinde kullanılan Deep Learning modeli bizim çalışmamızda bir futbol müsabakasının olası üç sonucunu (galibiyet, beraberlik, mağlubiyet) tahmin etmek için kullanılmıştır.

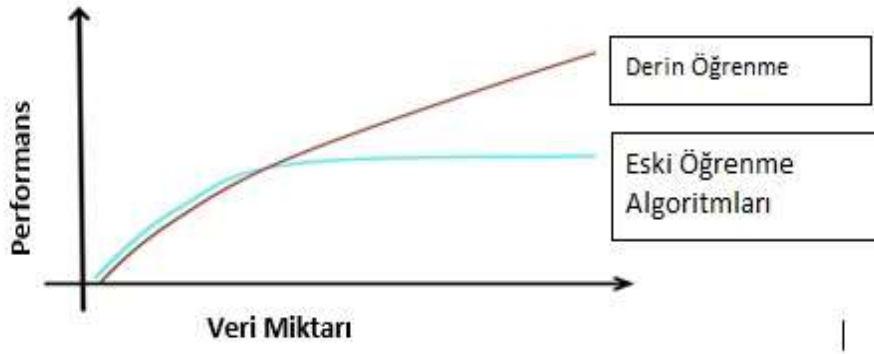
İnsan beyninin çalışma şeklini temel alan Yapay Sinir Ağları düğüm olarak adlandırılan nöronlardan oluşur. Üç katman halinde bu düğümler yan yana sıralanmışlardır. Bu düğümler sırasıyla;

- Giriş Katmanı
- Gizli Katman(lar)
- Çıkış Katmanı

Şeklindedir.



Şekil 2.6. Derin Öğrenme ve Makine Öğrenmesinin sınıflandırma yaklaşımlarının karşılaştırılması



Şekil 2.7. Derin Öğrenme ve Makine Öğrenmesinin sınıflandırma yaklaşımlarının karşılaştırılması

Derin öğrenmeyi makine öğrenmesinden ayıran iki önemli fark vardır bunlar sırası ile;

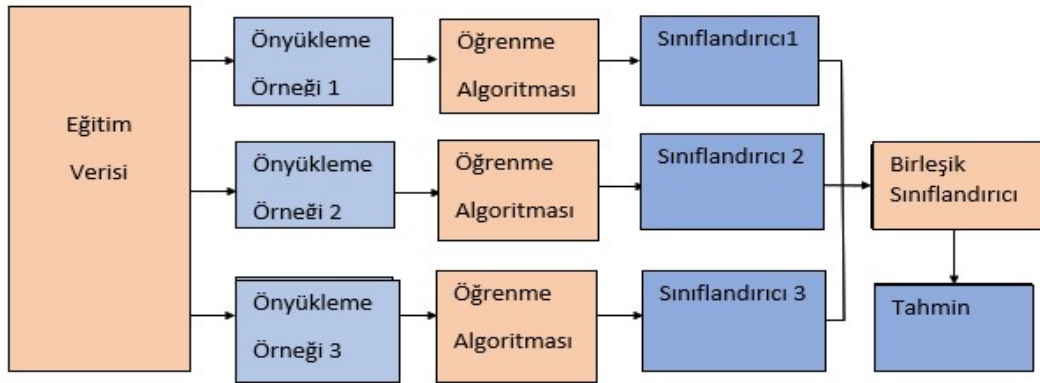
- A) Şekil 2.6 incelediğinde; Makine öğrenmesi öğrenme işlemi özelliklerden manuel olarak elde ederek öğrendiği model üzerinden sınıfları belirlerken Derin öğrenme ise öğrenme işlemi otomatik olarak verilerden elde eder (<https://www.mathworks.com/>, 2021). Diğer bir deyişle Derin Öğrenme, öğrenme işlemi “uçtan uca” öğrenme şekliyle gerçekleştirip bunu otomatik olarak nasıl yapacağını belirler.
- B) Şekil 2.7 incelediğinde; Makine öğrenmesi ile Derin öğrenme arasındaki diğer bir fark ise klasik makine öğrenme yöntemleri eğitim veri setindeki veri miktarı arttığında belli bir noktadan sonra plato oluştururken Derin öğrenmenin eğitim veri setindeki veri miktarı arttıkça daha iyi sonuçlar vermesidir (<https://www.mathworks.com/>, 2021).

Derin Öğrenme algoritmaları kendi kendine öğrenme yeteneğine sahiptir ancak çalışma şekli olarak Yapay Sinir Ağlarına bağlıdır. Eğitim sürecinde Derin öğrenme özellik çıkarımı, nesnelere gruplama ve faydalı veri kalıbı desenlerinin keşfedilmesinde girdilerdeki bilinmeyen öğeleri kullanmaktadır.

Derin öğrenme Modelleri birçok algoritmayı kullanmaktadır. Belirli görevler için bazı algoritmalar daha iyi sonuç vermektedir. Bizim çalışmamızda Weka yazılımı için geliştirilmiş WekaDeeplearning4j paketi kullanılmıştır. Model oluşturulurken ZoomModel olarak CustomNet, Instance İterator için DefaultInstanceIterator, EarlyStopping ve the İteratır Listener için EpochListener özellikleri seçilmiştir.

2.4.4 Torbalama

Bagging (Torbalama) algoritmasında temel amaç aynı görevi yerine getirmek için birden fazla zayıf model birleştirilerek kurulacak modelin etkin öğrenmesini sağlamaktır. Bagging terimi “bootstrap (önyükleme) + aggregating (toplama)” kelimelerinden türetilmiştir.



Şekil 2.8. Bagging (Torbalama) algoritması çalışma şekli

Torbalama algoritmasının temel amacı, aynı görevi yerine getirmek için birden fazla zayıf modeli birleştirerek kurulacak modelin etkin bir şekilde öğrenilmesini sağlamaktır (Breiman, 1996). Torbalama algoritması karar ağaçları ile birlikte kullanılmakta olup böylece doğruluk değeri artırılırken varyans değeri azaltılarak modelin kararlılığı önemli ölçüde artırılır ve aşırı öğrenme sorunundan kaçınılmış olunur (corporatefinanceinstitute, 2021). Bu algorithmada, her zayıf model, ikame (önyükleme) kullanılarak veri kümesindeki örneklerden rastgele bir alt örnek seçilerek eğitilir. Nihai sonuç, en son zayıf modellerin tahminlerinin toplanmasıyla elde edilir.

Bizim modelimizde Bagging Algoritması için RepTree algoritması seçilmiştir. Sistem parametreleri; bagsizepercen: 100, batchsize: 380, classifier: RepTree, numiterator: 2 olarak ayarlanmıştır.

2.4.5 Yinelemeli Sınıflandırıcı İyileştirici

Yinelemeli Sınıflandırıcı İyileştirici algoritması, seçilen sınıflandırıcı için çapraz doğrulama kullanarak yineleme sayısını optimize etmeyi sağlar ve bu sayede sınıflar ve özellikler iyileştirilir (Meshram, Safari, Khosravi, & Meshram, 2021). Bu algorithmada veriler birer birer sisteme verilir. Tüm girdiler sisteme verildikten sonra işlem tekrarlanır bu nedenle

bu algoritma yapay sinir ağlarının önemli bir unsurunu içermekte olup eğitim aşamasında başlangıç ağırlıkları rastgele olarak seçilir (Manikandan, Vasudev, & Balasubramanian, 2018).

Bizim sistemimizde IterativeClassifierOptimizer algoritmasının parametreleri sırası ile; batchsize: 380, classifier: Logiboost, evaluationMetric :precision seçilmiştir.



3. DENEYSEL SONUÇLAR

Bu bölümde öncelikle veri seti ile ilgili istatistiksel bilgiler verilmektedir. Daha sonra veri madenciliği sonuçlarının analizinde kullanılan performans ölçütleri tanımlanmıştır. Sonraki bölümler, tahmin sonuçlarını ve sonuçlardaki iyileştirmeleri kapsar. Son bölüm, tahmin işlemi sırasında özelliklerden elde edilen bilgileri içerir.

Çizelge 3.1. 2014-2020 Serie A'da ev sahibi ve deplasman takımlarının tercih ettiği taktikler

Serie A							
Ev Sahibi Takım							
Taktik	Toplam	Galibiyet	Berberlik	Mağlubiyet	Galibiyet%	Berberlik%	Mağlubiyet%
41221	593	259	146	188	43.68%	24.62%	31.70%
31402	434	173	103	158	39.86%	23.73%	36.41%
41212	325	137	84	104	42.15%	25.85%	32.00%
42031	104	50	23	31	48.08%	22.12%	29.81%
41131	92	45	28	19	48.91%	30.43%	20.65%
30412	91	45	18	28	49.45%	19.78%	30.77%
	2280	1003	547	730	43.99%	23.99%	32.02%
Deplasman Takım							
41221	552	188	135	229	34.06%	24.46%	41.49%
31402	459	141	101	217	30.72%	22.00%	47.28%
41212	306	74	67	165	24.18%	21.90%	53.92%
42031	113	53	25	35	46.90%	22.12%	30.97%
41131	92	38	25	29	41.30%	27.17%	31.52%
	2280	730	547	1003	33.02%	23.99%	43.99%

3.1 Veri Kümeleri Hakkında İstatistiksel Bilgiler

Bu çalışmada geliştirilen tarayıcı vasıtası ile Serie A, Premier Lig ve Bundesliga olmak üzere üç büyük lig için 2014-2020 yılları arasındaki 6 sezon veri seti olarak kullanılmıştır. Bu liglerden oluşan veri setinin ilk 5 sezonu eğitim veri seti olarak, son sezon (2019- 2020) ise tahminler için test veri seti olarak kullanılmıştır. Bu bölüm, çalışma için çok önemli olan taktikler ve piyasa değerleri hakkında istatistiksel bilgiler içermektedir.

3.1.1 Taktik İstatistikler

Bir takımın oluşumu, menajer tarafından belirlenen bir futbol taktiğidir. Bu bölümde teknik direktörlerin 2014-2020 yılları arasında her üç lig için ürettikleri dizilişlerin galibiyet, beraberlik ve mağlubiyet durumları incelenmiştir. Çizelge 3.1, İtalya için bu durumları ev sahibi ve deplasman takımları açısından göstermektedir.

Veri setinde Serie A'da ev sahibi takımların 51 farklı diziliş sahadayken, deplasman takımları 56 farklı diziliş kullandığı görülmüştür. Bu dizilişlerden 80'den fazla maç Çizelge 3.1 'te sunulmuştur. Çizelge 4.1'e göre 41221 taktiği hem ev sahibi hem de deplasman takımları tarafından en çok tercih edilen taktiktir. İkinci sırada ise üçlü savunmayı içeren 31402 taktiği dikkat çekmektedir. Üçlü savunma, son yıllarda İtalyan milli takımı tarafından da tercih edilmektedir. Ancak kazanma oranlarını incelediğimizde bu iki taktiğin sonuç açısından pek iyi olmadığı görülmektedir. Ev sahibi takım dizilişlerine bakıldığında 42031, 41131 ve 30412 taktikleri sırasıyla %48.08, %48,91 ve %49,45 ile daha iyi sonuçlar verdiği görülmüştür. Deplasmanda orta sahayı kullanmayan ve defansa odaklanan 42031 taktiği diğer durumlara göre %46,90 galibiyet oranı ile çok daha iyi sonuçlar vermiştir. Üstelik deplasmanda defans yerine orta saha kullanan 41131 taktiği ilginç bir şekilde galibiyet yüzdesini (%41,30) düşürürken beraberlik yüzdesini (%27,17) artırmaktadır.

İngiltere Premier Ligi'nde 80'in üzerinde maçta kullanılan taktik diziliş sayısı, İtalya Ligi'nden çok daha fazladır. Bu ligde hem ev sahibi hem de deplasman takımları 48 farklı dizilişte oynamıştır. Çizelge 3.2, 9 farklı dizilişte ev sahibi ve deplasman takımları için maç sonuç oranlarını vermektedir.

Premier Lig'de 41221 dizilişi çok tercih edilen bir taktiktir. Çizelge 3.2. 'e göre, Bölge 2'de dört oyuncu öne çıkmaktadır. Ancak %29,77 oranla deplasmanda çok başarısız bir taktik olduğu görülmektedir. Bu taktik yerine, 4. bölgede bir oyuncu daha kullanan 41131 taktiği, galibiyet oranını iç sahada ve deplasmanda sırasıyla %49.03 ve %41,31 oranında arttırmaktadır. Ev sahibi ve deplasman takımları 31402 taktiğini kullandığında kazanma oranı sırasıyla %30.00 ve %22,22'ye düşmektedir.

Çok kullanılmasa da üçlü defans kullanan ve 31402 taktiğine göre 2 hücum oyuncusu daha bulunan 32221 taktiği, iç sahada ve deplasmanda sırasıyla %52,88 ve %49,19 galibiyet oranı ile en iyi taktiklerden biri olduğu görülmüştür. İngiliz klasiği 442 olarak bilinen 40402 taktiği, bölge1, bölge3 ve bölge 5'i kullanmaktadır. Çizelge 3.2, bu basit taktiğin birkaç maçta kullanıldığını ve deplasmanda çok yetersiz kaldığını göstermektedir.

Bundesliga'da 18 takım olduğu için maç sayısı diğer liglere göre daha azdır. Ancak tercih edilen diziliş açısından ev sahibi takımlar 62 farklı taktik denerken, deplasman takımları sahada 58 farklı taktikle oynamıştır. Çizelge 3.3. öne çıkan taktikleri ve maç sonuçlarını göstermektedir. Çizelge 3.3'e göre diğer iki ligde de iyi sonuçlar veren 41131 dizilişinin hem iç

sahada hem de deplasmanda her zaman kullanıldığı görülmektedir. Diğer liglerde olduğu gibi 41221 taktiği de bu taktiğe göre kötü istatistiklere sahiptir.

Çizelge 3.2. 2014-2020 Yılları arasında Premiere Lig'de ev sahibi ve deplasman takımlarının tercih ettiği taktikler

Premier Lig							
Ev Sahibi Takım							
Formasyon	Toplam	Galibiyet	Berberlik	Mağlubiyet	Galibiyet%	Berberlik%	Mağlubiyet%
41221	508	240	120	148	47.24	23.62	29.13
41131	359	176	83	100	49.03	23.12	27.86
42031	159	74	30	55	46.54	18.87	34.59
41212	145	63	41	41	43.45	28.28	28.28
32221	104	55	22	27	52.88	21.15	25.96
41122	94	39	23	32	41.49	24.47	34.04
31402	90	27	19	44	30.00	21.11	48.89
40402	84	39	15	30	46.43	17.86	35.71
42022	83	40	24	19	48.19	28.92	22.89
...							
	2278	1013	536	729	0.4447	0.2353	0.32
Deplasman Takım							
41221	514	153	126	235	29.77	24.51	45.72
41131	305	126	65	114	41.31	21.31	37.38
42031	152	46	44	62	30.26	28.95	40.79
41212	140	46	28	66	32.86	20.00	47.14
32221	124	61	21	42	49.19	16.94	33.87
41311	95	23	26	46	24.21	27.37	48.42
31402	90	20	26	44	22.22	28.89	48.89
41122	89	30	20	39	33.71	22.47	43.82
40402	80	17	23	40	21.25	28.75	50.00
...							
	2278	729	536	1013	0.32	0.2353	0.4447

Ancak, diğer liglerden farklı olarak, Almanya'daki takımlar genellikle 41131 taktiğini tercih etmişlerdir. Çizelge 3.3, 41212 taktiğinin hem ev sahibi hem de deplasman maçlarında en iyi sonuçları verdiğini göstermektedir. Çizelge 3.3'ten 42022 dizilişinin ne ev sahibi ne de deplasman maçlarında tercih edilmemesi gereken bir taktik olduğu tespit edilmiştir. 42031 taktiği, yalnızca bir forvet bir bölge "Ofansif Orta Saha Oyuncusu" konumuna geri getirildiğinde hem deplasmanda hem de evde maç sonuçlarını iyileştirir.

Çizelge 3.3. 2014-2020 Yılları arasında Bundesliga'da ev sahibi ve deplasman takımlarının tercih ettiği taktikler

Bundesliga							
Ev Sahibi Takım							
Formasyon	Toplam	Galibiyet	Beraberlik	Mağlubiyet	Galibiyet%	Beraberlik%	Mağlubiyet%
41131	274	138	62	74	50.36%	22.63	27.01
41221	239	108	54	77	45.19%	22.59	32.22
41122	182	101	38	43	55.49%	20.88	23.63
41212	154	64	36	54	41.56%	23.38	35.06
42022	106	39	28	39	36.79%	26.42	36.7
42031	98	49	19	30	50.00%	19.39	30.61
...							
	1836	814	450	572	0.4435	0.2452	0.3111
Deplasman Takım							
41131	256	99	63	94	38.67	24.61	36.72
41221	224	61	59	104	27.23	26.34	46.43
41122	163	68	35	60	41.72	21.47	36.81
41212	137	46	31	60	33.58	22.63	43.80
42022	107	27	31	49	25.23	28.97	45.79
42031	103	26	27	50	25.24	26.21	48.54
31402	80	25	24	31	31.25	30.00	38.75
...							
	1836	572	450	814	0.3111	0.2452	0.4435

Çizelge 3.1, 3.2 ve 3.3 olmak üzere üç çizelgede sunulan taktiklere göre her ligin kendine has özellikleri olduğu görülmektedir. Almanya Ligi dışında daha az kullanılan 42031 taktiğinin her ligde başarılı sonuçlar verdiği belirlenmiştir. Bu çizelgelerde ortalama değerler incelendiğinde ev sahibi takımın galibiyet yüzdesi beklendiği gibi daha yüksektir. Bu yüzdeler Serie A, Premier Lig ve Bundesliga için sırasıyla %43,49, %44,47 ve %44,36 galibiyet oranlarına sahiptir. Sonuçlar birbirine çok yakındır. Her üç ligin de beraberlik ve mağlubiyet açısından benzer sonuçlara sahip olduğu görülmektedir. Taktik, bir menajer tarafından tasarlanmış bir şablon iken, bir sonraki bölümde veri seti, transfer piyasa değerleri açısından incelenmektedir.

3.1.2 Transfer Piyasası Değerleri Açısından İstatistikler

Transfer piyasa değeri, bir oyuncunun becerilerinin bir göstergesi olarak kullanılan sayısal bir değerdir. Bu değer, oyuncunun iyi veya kötü oynamasına bağlı olarak zamanla değişmektedir. Çizelge 4.4, her üç lig için bölgelerin ortalama futbolcu piyasa değerlerini ve bu

bölgelerin standart sapmalarını göstermektedir. Bir takımda yalnızca bir kaleci olduğundan Çizelge 3.4'ün Bölge 0 için standart sapması yoktur.

Çizelge 3.4. 2014-2020 Yılları arasında üç ligde oyuncuların bölgelere göre ortalama transfer piyasa değerleri

	Serie A		Premier League		Bundesliga	
	1 ¹	2 ²	1 ¹	2 ²	1 ¹	2 ²
Bölge 0	1	8.19	1	14.57	1	6.58
Bölge 1	3.63	8.01±3.91	3.91	13.33±6.72	3.82	9.3±5.56
Bölge 2	0.97	9.96±5.04	1.16	18.13±7.59	1.15	9.35±4.13
Bölge 3	2.53	10.54±5.59	1.79	17.53±8.56	1.7	10.63±4.63
Bölge 4	1.31	13.72±7.07	1.78	25.78±12.24	1.85	12.46±6.97
Bölge 5	1.55	14.90±7.57	1.35	26.09±9.11	1.48	14.41±6.11
İlk 11	11	10.45±7.69	11	17.98±11.38	11	10.34±6.99

¹Standart sapma ile ortalama oyuncu sayısı, ²Standart sapma ile ortalama piyasa değerleri (Milyon Euro)

Çizelge 4.4'e göre transfer piyasa değerleri açısından en değerli lig Premier Lig'dir. Her lig, atak oyuncuları içeren Bölge 4 ve Bölge 5'te en pahalı piyasa değerlerine sahiptir. İngiltere ligi kalecilere (Bölge 0) önem vermekte olup bu pozisyondaki oyunculara daha fazla bütçe ayırmaktadır. Çizelge 3.5, galibiyet, beraberlik ve mağlubiyet açısından ev sahibi ve deplasman takımlarının ortalama transfer piyasa değerlerinin analizini içermektedir.

Çizelge 3.5. 2014-2020 Yılları arasında üç ligdeki takımların ortalama transfer piyasa değerleri

	Serie A ¹	Premier League ¹	Bundeslig ¹
	Piyasa Değeri	Piyasa Değeri	Piyasa Değeri
EV Sahibi Galibiyet	13.34±9.64	22.19±13.97	12.79±8.56
Deplasman Mağlubiyet	7.52±5.75	14.3±9.32	8.03±5.61
Deplasman Galibiyet	14.57±10.41	23.83±14.68	14.33±9.38
Ev Sahibi Mağlubiyet	7.22±5.56	13.43±8.67	7.66±5.28
Ev Sahibi Beraberlik	9.42±7.09	16.43±10.29	9.38±6.41
Deplasman Beraberlik	10.36±7.51	16.74±10.68	9.36±6.33

¹Takımların Piyasa Değerleri
(Milyon Euro)

Çizelge 3.5'deki galibiyet istatistikleri, beklendiği gibi, piyasa değeri daha yüksek olan hem ev sahibi hem de deplasman takımlarının maçlarını kazandığını göstermektedir. Özellikle deplasman takımlarının deplasmanda kazanmak için daha fazla piyasa değerine sahip olduğu görülmektedir. Ev sahibi takımın ortalama piyasa değeri beraberlik açısından daha az olsa da ortalama piyasa değerlerinin birbirine yakın olduğu görülmüştür. Beraberlikte ev sahibi takımın ortalama piyasa değeri daha az olmasında taraftar ve ev sahibi takım faktörlerinin etkili olduğunu göstermektedir. Çizelge 4.5, her üç lig için de benzer sonuçlar olduğunu göstermektedir.

3.2 Performans Metrikleri

Bir futbol maçında bir takım için üç sınıf vardır: Galibiyet, Beraberlik, Mağlubiyet. Veri madenciliğinde, veri madenciliği yöntemleri kullanılarak eğitim veri setinden elde edilen bir öğrenme modeli ile bu üç sınıfın doğru bir şekilde tahmin edilmesi amaçlanmaktadır.

- True Positive (TP): Sınıf için bir takımın sonucunun doğru tahmini.
- True Negative (TN): Sınıf dışındaki diğer negatif sınıfların doğru tahmini.
- False Positive (FP): Takımın doğru sonucunun yanlış tahmini.
- False Negative (FN): Sınıf dışındaki diğer negatif sınıfların yanlış tahmini.

İkili sınıflandırmadan farklı olarak bu değerlerin hesaplanmasında durumlara dikkat etmek gerekmektedir. Karışıklık matrisi (Stehman, 1997), en uygun modelin sonuçlarının görsel olarak görüntülendiği yerdir. Çizelge 4.6, bu matrisi tanıtmaktadır.

Çizelge 3.6'da matrisin her satırı gerçek sınıfı ve her sütunu tahmin sınıfını göstermektedir.

Çizelge 3.6. Bir takım için karışıklık matrisi

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	a	b	c
	Mağlubiyet	d	e	f
	Beraberlik	g	h	i

Her sınıf için yukarıdaki dört durumun hesaplanması ayrı ayrı yapılmaktadır. Üç sınıf için:

$$TP_{galibiyet} = a \quad (4.1)$$

$$TN_{galibiyet} = e + f + h + i \quad (4.2)$$

$$FP_{galibiyet} = d + g \quad (4.3)$$

$$FN_{galibiyet} = b + c \quad (4.4)$$

$$TP_{mağlubiyet} = e \quad (4.5)$$

$$TN_{mağlubiyet} = a + c + g + h \quad (4.6)$$

$$FP_{mağlubiyet} = b + h \quad (4.7)$$

$$FN_{mağlubiyet} = d + f \quad (4.8)$$

$$TP_{beraberlik} = i \quad (4.9)$$

$$TN_{beraberlik} = a + b + d + e \quad (4.10)$$

$$FP_{beraberlik} = c + f \quad (4.11)$$

$$FN_{beraberlik} = g + h \quad (4.12)$$

Literatürde bu durumlardan elde edilen birçok performans ölçütü bulunmaktadır. Doğruluk, doğru tahminlerin sayısının tüm tahminlere oranı olan en çok kullanılan ölçümdür. Ancak bu metrik, dengesiz bir veri kümesi için güvenilir bir metrik değildir (Uzun & Özhan, 2018). Bu nedenle bu makalede precision, recall, ve f -Measure dahil üç performans ölçütü

(Powers, 2011) kullanılmıştır. Bu metrikler, bir tahminin başarısının yararlı ölçümleridir. Bu metrikler her sınıf için ayrı ayrı hesaplanmaktadır.

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \quad (4.13)$$

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \quad (4.14)$$

$$f\text{-Measure}_c = \frac{2 * TP_c}{2 * TP_c + FP_c + FN_c} \quad (4.15)$$

Burada c sınıfıdır (galibiyet, beraberlik, mağlubiyet). Precision, bir sınıf için doğru tahmin edilen eşleşmelerin toplam tahmin edilen değerlere oranıdır. Recall, bir sınıf için doğru tahmin edilen eşleşmelerin gerçek değerlere oranıdır. f-Measure, Precision ve Recall değerlerinin ağırlıklı ortalamasıdır.

3.3 Metotların Performans Değerlendirmesi

Bu çalışmada, tahmin için gerekli öğrenme modelini oluşturmak için 2014-2019 yılları arasındaki 5 sezonluk eğitim verileri kullanılmıştır. Serie A, Premier Lig ve Bundesliga dahil olmak üzere her eğitim verisi o lig için özel olarak oluşturulmuştur. Sonuç olarak, her lig için bir öğrenme modeli oluşturulmuştur. Bu öğrenme modelleri için, tahmin modellerini oluşturmak için Derin öğrenme, Torbalama ve Yinelemeli sınıflandırıcı iyileştiriciyi içeren 3 farklı yöntem uygulanmıştır. Eğitim veri kümelerinden oluşturulan modeller, 2019-2020 sezonu maçlarını tahmin etmek için kullanılmıştır. Çizelge 3.7, Serie A için Derin öğrenme yöntemi kullanılarak oluşturulan öğrenme modeli için test verilerinden elde edilen performans ölçümlerini sunmaktadır.

Çizelge 3.7'ye göre 118 galibiyet, 84 mağlubiyet ve 11 beraberlik doğru tahmin edilmiştir. Bu tahminde, ev sahibi takım galibiyeti için negatif durumların 0.756 (Recall) ile başarılı bir şekilde tahmin edildiği görülmektedir. Pozitif durumlarda, 0,611 (Precision) değeri vardır. Sonuç olarak bu iki değer harmonik ortalamasını temsil eden f-Measure değeri ev sahibi takımın maçları için 0,676 (f-Measure) ile başarılı bir şekilde tahmin edilebilmektedir. Burada ev sahibi takımın maç kaybının f-Measure değeri 0,602'ye düşmektedir. Kazanılan sınıf

ile karşılaştırıldığında bu durum için öğrenme modelinin başarısız olduğu görülmektedir. Yani deplasman takımının galibiyetini bilmek daha zordur. Beraberlik durumunun tahmin edilmesi en problemli sınıftır ve başarı oranı diğer iki sınıfa göre çok düşüktür. Diğer çalışmalarda sıklıkla kullanılan 0,561 doğruluk değeri, derin öğrenme yöntemi ile elde edilen öğrenme modelinin genel başarısını göstermektedir. Ancak, sınıfların durumu bu metrikte görülemez. Çizelge 3.7'de gösterildiği gibi, sonuçları sınıf bazında daha iyi yorumlamak için f-Measure değeri daha güvenilirdir.

Çizelge 3.7. Serie A- Derin Öğrenme için performans ölçümleri

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	118	32	6
	Mağlubiyet	25	84	19
	Beraberlik	50	35	11
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.611	0.756	0.676
	Mağlubiyet	0.556	0.656	0.602
	Beraberlik	0.306	0.115	0.167
Accuracy(Doğruluk)		0.561		

3.4 Test Veri Kümesinde Tahmin Performanslarının Karşılaştırılması

Performans kıyaslaması için veri madenciliği yöntemleri uygulanabilir. Bu çalışma, karşılaştırma için Derin öğrenme, Torbalama ve Yinelemeli sınıflandırıcı iyileştirici dahil olmak üzere üç popüler yöntem kullanmaktadır. Çizelge 3.8, sınıfların f -Measure değerlerini ve her lig ve yöntem için her modelin accuracy değerlerini vermektedir.

Her sınıfın ve ligin f -Measure değerleri benzer performans sonuçlarını içermektedir. Ev sahibi galibiyetlerinin sonuçlarını tahmin etmek diğer sınıflara göre daha basitken, beraberlikleri bilmenin araştırma konusu üç lig için zor olduğu görülmüştür. Derin Öğrenme, beraberlik maçlarını tahmin etmede diğer yöntemlerle karşılaştırıldığında en iyi yöntemdir. Yinelemeli Sınıflandırıcı İyileştirici yöntemi, Premier Lig sonuçları hariç ortalama sonuçlarda en başarılı yöntemdir. Premier Lig'de torbalama bu lig için en uygun yöntemdir. Bundesliga'nın, tahmin görevindeki en zorlu lig olduğu görülmüştür. Bu sonuçlardan sonra her lig için haftalık tahminler analiz edilmiştir. Ayrıca tahmin görevinde sorunlu haftalar ve nedenleri araştırılmıştır.

Çizelge 3.8. Test veri seti için üç yöntemin ve üç ligin tahmin performanslarının karşılaştırması

Serie A						
f-Measure						
		DL ¹	B ²	ICO ³	NB ⁴	J48 ⁵
Sınıf	Galibiyet	0.676	0.697	0.683	0.642	0.638
	Mağlubiyet	0.602	0.635	0.674	0.616	0.506
	Beraberlik	0.167	0.226	0.194	0.318	0.269
	Accuracy(Doğruluk)	0.561	0.592	0.6	0.558	0.505
Premier Lig						
		DL	B	ICO	NB	J48
Sınıf	Galibiyet	0.676	0.693	0.679	0.551	0.558
	Mağlubiyet	0.613	0.637	0.58	0.542	0.543
	Beraberlik	0.132	0.195	0.11	0.29	0.279
	Accuracy(Doğruluk)	0.576	0.584	0.561	0.479	0.487
Bundesliga						
		DL	B	ICO	NB	J48
Sınıf	Galibiyet	0.646	0.635	0.654	0.556	0.549
	Mağlubiyet	0.613	0.573	0.597	0.503	0.481
	Beraberlik	0.089	0.2	0.175	0.328	0.342
	Accuracy(Doğruluk)	0.552	0.529	0.555	0.471	0.477

¹Derin Öğrenme, ²Navie Bayes, ³Torbalama, ⁴J48 Sınıflandırıcı, ⁵Yinelemeli Sınıflandırıcı Optimize Edici

Çizelge 3.8 incelendiğinde kurmuş olduğumuz modelin tüm haftaların dahil olduğu test veri setinde klasik veri madenciliği yöntemlerinden olan Navie Bayes ve J48 Sınıflandırıcı algoritmalarından daha iyi sonuç verdiği görülmektedir. Gerek doğruluk oranları gerekse de F-Measure değerleri açısından kurmuş olduğumuz model daha iyi tahmin sonuçları vermektedir.

3.5 Alt Test Veri Kümesinde Tahmin Performanslarının Karşılaştırılması

Test verilerinden tahmin sonuçları elde edildikten sonra tahmin hatalarında üç ana başlık ortaya çıkmaktadır: (1) Şampiyonlar Ligi ve UEFA ligi grup maçları, (2) kış transfer sezonu ve (3) son 2 maç günü. Örneğin, bu çalışmada üç büyük ligin her birinden 7-8 takım UEFA'nın iki büyük organizasyonu olan Şampiyonlar Ligi ve Avrupa Ligi'ne katılmaktadır. Takımlar bu organizasyonlara önem vermektedir. Bu nedenle bu organizasyonlardaki maç öncesi veya sonrasında takımların ulusal liglerindeki maçlarda beklenmeyen sonuçlar ortaya çıkabilmektedir. Ayrıca bu organizasyonlarda oyuncular hem fiziksel hem de zihinsel olarak

yorulmaktadır. Transfer sezonunda bazı oyuncular futbol oynamaya isteksiz veya çok istekli olabilmektedirler. Son iki hafta için takımların lig sıralamaları netleşirken oyuncuların maçlara konsantrasyonu da azalmaktadır. Test veri seti bu üç başlıkta futbol maç sonuçlarını olumsuz etkilemekte ve maç tahminlerini zorlaştırmaktadır. Bu nedenle Premier Lig için 13, 14, 15, 16, 17, 18, 19, 20, 21, 24, 25, 26, 27, 28, 37, 38 Serie A için 12,13, 14, 15, 16, 17, 18, 19, 20, 21,23, 24, 25, 26,27, 37, 38 ve Bundesliga için 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 33, 34 maç günleri alt test veri seti olarak adlandırılan yeni bir test veri seti için dahil edilmemiştir. Başka bir deyişle, Serie A, Premier Lig ve Bundesliga için 380, 380 ve 306 maç tahmin etmek yerine, alt test veri seti olarak sırasıyla 210, 220 ve 171 maç tahmin edilmiştir. Çizelge 3.9, bu maçlardan elde edilen performans sonuçlarını göstermektedir. Ayrıca bu görev için üç farklı veri madenciliği yöntemi test edilmiştir.

Çizelge 3.9. Alt test veri seti için üç yöntemin ve üç ligin tahmin performanslarının karşılaştırması

Serie A						
f -Measure						
		DL ¹	B ²	ICO ³	NB ⁴	J48 ⁵
Sınıf	Galibiyet	0.704	0.708	0.7	0.674	0.643
	Mağlubiyet	0.662	0.688	0.709	0.649	0.545
	Berberlik	0.149	0.225	0.102	0.295	0.283
Accuracy(Doğruluk)		0.6	0.619	0.619	0.586	0.519
Premier Lig						
		DL	B	ICO	NB	J48
Sınıf	Galibiyet	0.712	0.706	0.708	0.555	0.562
	Mağlubiyet	0.644	0.644	0.629	0.552	0.576
	Berberlik	0.098	0.211	0.143	0.198	0.239
Accuracy(Doğruluk)		0.6	0.595	0.604	0.464	0.5
Bundesliga						
		DL	B	ICO	NB	J48
Sınıf	Galibiyet	0.696	0.658	0.686	0.592	0.571
	Mağlubiyet	0.641	0.61	0.624	0.538	0.509
	Berberlik	0.113	0.203	0.172	0.372	0.341
Accuracy(Doğruluk)		0.584	0.549	0.578	0.503	0.497

¹Derin Öğrenme, ²Navie Bayes, ³Torbalama, ⁴J48 Sınıflandırıcı, ⁵Yinelemeli Sınıflandırıcı Optimize Edici

Çizelge 3.9'da her sınıf ve yöntemin sonuçlarında iyileştirmeler bulunmaktadır. 0,619 doğrulukla Serie A için Torbalama yöntemi Çizelge 3.9'un en iyi değeridir. Derin öğrenme yöntemi hem Premier Lig'de hem de Bundesliga'da en iyi performans sonuçlarını vermiştir. Bu sonuçlar, bazı haftaları istisna olarak tutmanın tahmin başarısını artırdığını göstermektedir.

Çizelge 3.9 incelendiğinde maç sonuçlarına negatif yönde etki eden haftaların çıkartıldığı alt veri setinde test sonuçları için kurmuş olduğumuz modelin hem Navie Bayes hem de J48 Sınıflandırıcı algoritmasına nazaran dahi iyi sonuçlar verdiği hem de kendi içinde daha tutarlı olduğu görülmektedir. Bu da kurmuş olduğumuz modelin inceleme konusu üç lig için seçmiş olduğumuz veri madenciliği yöntemleri ile elde edilen başarının klasik veri madenciliği yöntemleri (Navie Byeas, J48) ile elde etmenin mümkün olmadığını göstermektedir.

3.6 Özelliklerin Bilgi Kazançları

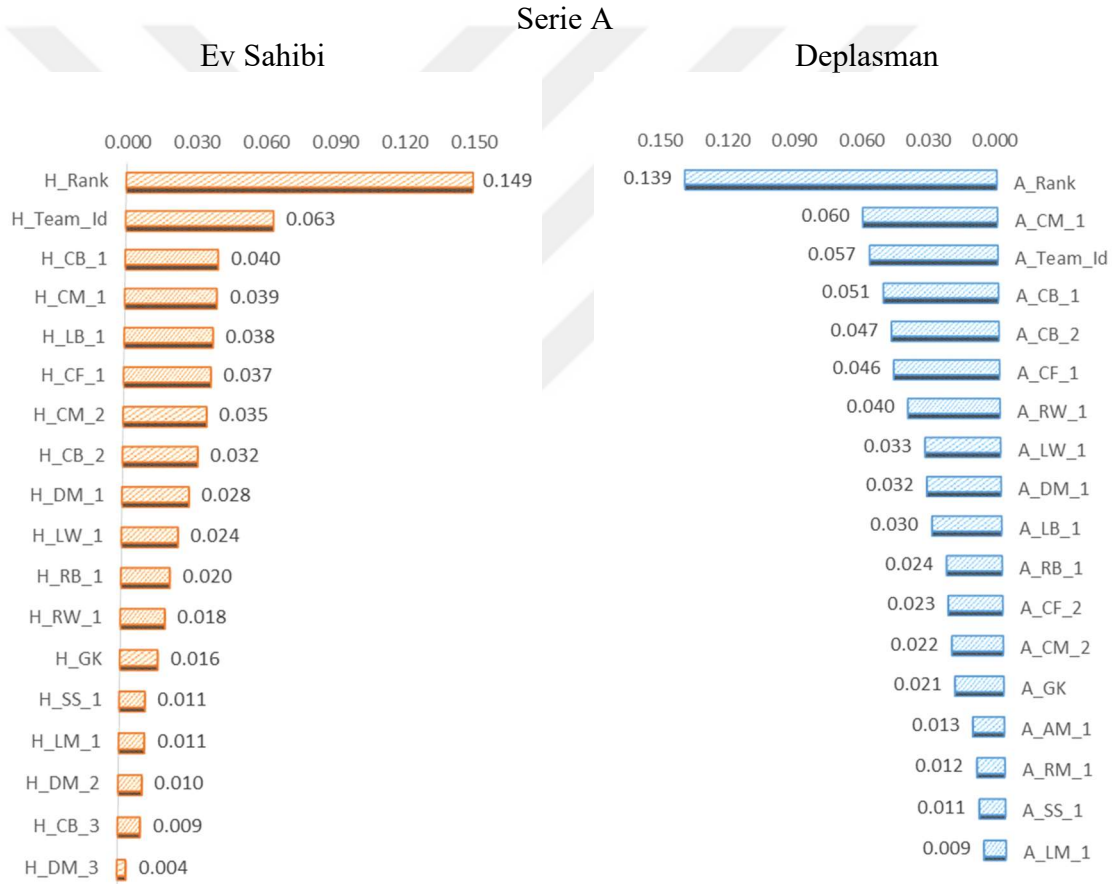
Bir veri madenciliği çalışmasında her özelliğin etkisi eşit değildir. Bir futbol maçında taktiğe dahil olan her oyuncunun etkisinin aynı olması beklenemez. Bu çalışmada, özelliklerin etkisini araştırmak için bilgi kazanımı kullanılmıştır. Ross Quinlan (Quinlan J. , 1986) tarafından önerilen bilgi kazancı (IG), özelliklerin değerinin analiz edilmesiyle elde edilen bilgi miktarıdır. Yani IG, sınıfla ilgili bir özelliğin değerini ölçmektedir.

$$IG(c, f) = H(c) - H(c|f) \quad (4.16)$$

Burada c sınıftır (Galibiyet, beraberlik, mağlubiyet) ve f özelliktir. IG, eğitim veri kümesini bölmeden önceki ve sonraki iki entropi (H) arasındaki farkı hesaplar. Bu nedenle IG, c'deki safsızlığı belirler. Şekil 3.1, Serie A için öğrenme modelinde çok önemli faktörler olan özellikleri ve bilgi kazanım değerlerini içermektedir.

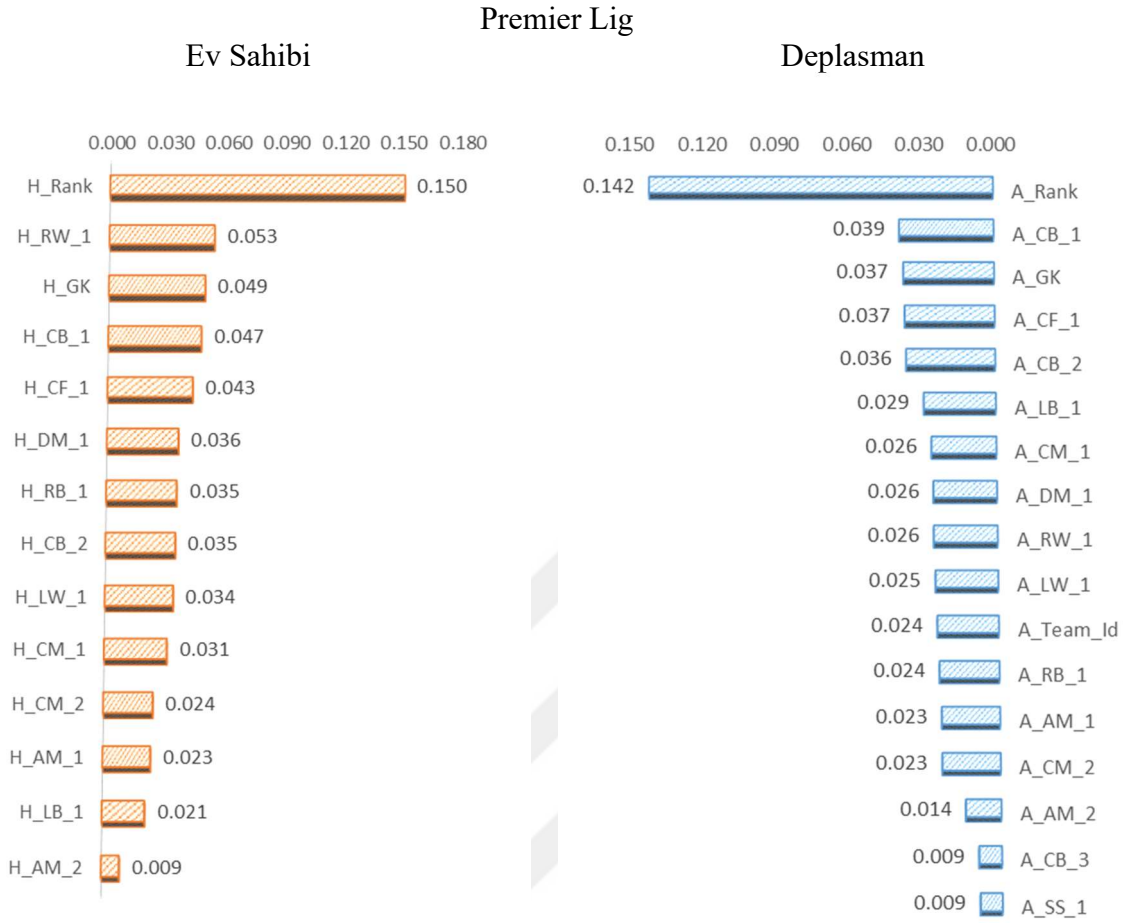
İtalyan Serie A liginde takım sıralaması ve takıma verilen benzersiz değer (Id) en üstün özellikleri arasındadır. İtalya Sere A'da deplasman takımı açısından sırasıyla orta saha merkezini (D_CM_1) ve defansın merkezini (D_CB_1 ve D_CB_2) içeren özellikler tahmin modelini olumlu etkilediği görülmektedir. Başka bir deyişle, Bölge 1 ve Bölge 3, İtalyan futbolunun savunma özelliklerini göstermektedir. D_CF_1, D_RW_1 ve D_LW_1 (deplasmanda kontra atak özellikleri) özellikleri Serie A'daki bir maçı etkileyen özellikler arasındadır. D_DM_1, D_LB_1 ve D_RB_1 özellikleri, İtalya'daki bir deplasman takımının

savunmaya önem vermesi gerektiğini göstermektedir. İkinci forvet (D_CF_2) ve merkez-orta saha (D_CM_2) özelliklerinin etkisi varken, kalecinin (D_GK) diğer özelliklerden daha az etkisinin olduğu görülmektedir. Şekil 3.1’de, ev sahibi takımda, E_CB_1, E_CM_1, E_LB_1, E_CM_2, E_CB_2, E_DM_1, E_RB_1 ve E_GK dahil olmak üzere ilginç bir şekilde savunma özellikleri baskın görünmektedir. Serie A ligi için sadece puanı etkileyen özelliklerin sayısı ve değerleri incelendiğinde E_CF_1, E_LW_1, E_RW_1 ve E_SS_1 gibi özelliklerin de az etkili olduğu görülmüştür. Geriye kalan tüm özellikler defansif iken, İtalyan futbolunda E_CB_3 gibi üçlü defans oynamanın da tahmin modelini etkilediğini görülmektedir. Sonuç olarak Şekil 3.1 incelendiğinde İtalyan futbolunda savunmanın hem ev sahibi takım hem de deplasman takımı için önemli özellikler içerdiği görülmektedir.



Şekil 3.1. Serie A'daki en iyi özelliklerin bilgi kazanımı

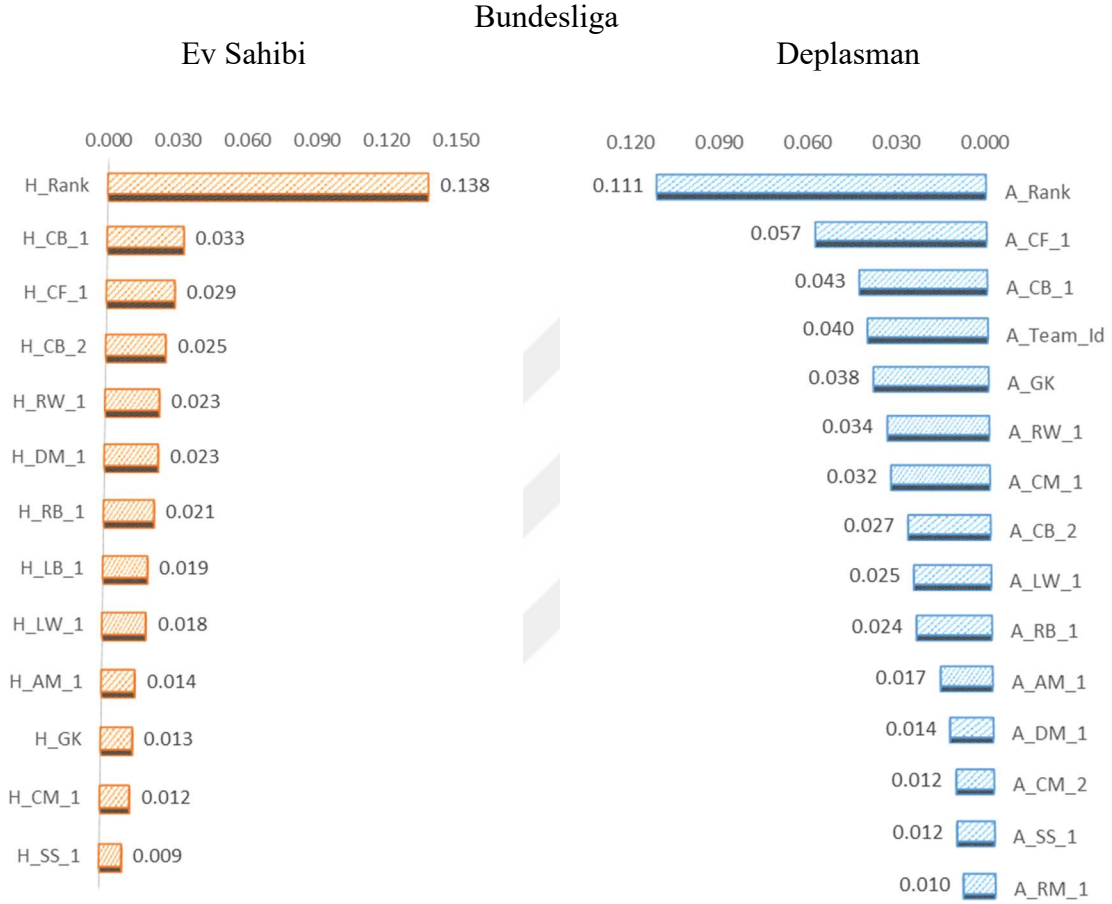
Aşağıdaki Şekil 3.2’de Premier Lig'deki bilgi kazanım değerleri incelenmiştir.



Şekil 3.2. Premier Lig'deki en iyi özelliklerin bilgi kazanımı

Şekil 3.2 incelendiğinde Premier Lig için de Serie A'daki gibi takım sıralamasının önemli bir faktörken olduğu görülmekte olup takıma verilen benzersiz değerin (Team_Id) diğer liglere göre daha az önemli olduğunu tespit edilmiştir. Buna göre takımların adının Premier Lig için Seri A'dakinden daha az önemli bir özellik olduğunu görülmektedir. Premier Lig için E_RW_1 özelliğinin yani forveti destekleyen sağ kanat özelliğinin ev sahibi takım için en değerli özellik olduğu görülmektedir. Deplasman takımları için ise beklendiği gibi D_CB_1 yani stoper özelliğini etkili bir özelliktir. Premier Lig'de hem ev sahibi hem de deplasman takımı için kaleciler baskın bir şekilde ön plana çıkmaktadır. Bu durum Premier Lig için kalecilere yaptıkları yüksek yatırım ile paralellik göstermektedir. Ayrıca bu yatırım kalecilerin piyasa değerlerine de yansımaktadır. Ev sahibi takımlar için E_CB_1, E_DM_1, E_RB_1 ve E_CB_2 gibi savunma pozisyonundaki özellikler en önemli özellikler olarak kendini göstermektedir. Bu lig için oyunun sonucunu etkileyen en önemli özelliklerden biri de E_CF_1

özelliği yani forvet oyuncudur. E_RW_1 özelliği yani sağ kanat ev sahibi takım için en önemli özelliklerden biri iken, deplasman takım için D_LB_1 yani sol bek özelliği en önemli özelliktir ki pozisyon itibari ile bu oyuncun görevi skoru doğrudan etkileyen ev sahibi takım için. E_RW_1 özelliğine karşılık gelen sağ kanat oyuncusunu durdurmaktadır. Alman Bundesliga için bir maçın sonuçlarına ışık tutan özellikler aşağıdaki Şekil 3.3'te verilmiştir.



Şekil 3.3. Bundesliga'daki en iyi özelliklerin bilgi kazanımı

Şekil 3.3'e göre Bundesliga için takımın sıralaması (Rank) ve benzersiz değeri (Team_Id) Premier Lig ile benzer özelliklere sahip olduğu görülmüştür. Bundesliga'da, deplasman takımları için forvet (A_CF_1) ve ev sahibi takımlar için merkez defans oyuncusu-stoper (H_CB_1) gibi özellikler, tahmin modeli üzerinde olumlu bir etkiye sahip olduğu görülmüştür. Kaleci faktörü ev sahibi takıma nazaran deplasman takımı adına çok daha yüksek önem arz etmektedir. Şekil 3.3 incelendiğinde deplasman takıma ait bilgi kazancı değerlerinin, ev sahibi takımın değerlerinden daha fazla oyunun sonucuna etki ettiği görülmektedir.

Serie A için toplam 30 özelliğten 15'i hem ev sahibi takım hem de deplasman takım taktiklerinde yer alarak tahmin modelinde aktif rol oynamıştır. Premier Lig için bu özellik sayısı ev sahibi ve deplasman takımları için sırasıyla 12 ve 14'e düşmektedir. Bundesliga'da ise bu değerler sırasıyla 11 ve 12'ye düşmektedir. Serie A'da oyunun sonucuna etki eden özellik sayısının Premier Lig ve Bundesliga'dan fazla olması çok farklı taktiklerin ve bu taktikler içindeki farklı pozisyonların oyunun sonuçlarını etkilediği işaret etmektedir. Diğer taraftan Bundesliga için temel pozisyonların maç sonuçlarını etkilediği tespit edilmiştir. Diğer liglere nazaran Premier Lig için kalecilerin çok daha büyük öneme sahip olduğu ortaya çıkmıştır. Ancak her ligdeki en önemli özelliklerden birinin stoper pozisyonu olduğu tespit edilmiştir. Oyunun sonucunu etkileyen en önemli özellik olarak ortaya çıkan stoper pozisyonun diğer bölge oyuncu piyasa değerleri ile karşılaştırıldığında en düşük değere sahip olması ilginç bir durumdur. İtalya Serie A için orta saha merkezindeki ilk 5 özelliği en önemli özellik olarak kendini gösterirken diğer iki ligde bu özelliklerin öneminin azaldığı görülmektedir. Ancak Serie A için ortalama piyasa değeri açısından ise en düşük değere orta saha oyuncuları sahiptir. Premier Lig ve Bundesliga'daki en iyi 5 özellik arasında, santrafor ve dramatik olarak sağ kanat pozisyonlarının çok aktif bir rol oynadığı tespit edilmiştir. Sonuç olarak Serie A'da taktik anlayışının savunma ve orta saha bölgelerinde yoğunlaştığı, diğer iki ligde ise savunma ve hücum taraflarının tahmin modeline etki ettiği tespit edilmiştir

Bu çalışmadaki bilgi kazanım değerleri kullanılarak Bölge 1 ile Bölge 5 arasındaki pozisyonları içeren taktikler oluşturulabilir. Yani kaleci hariç ilk 10 pozisyon ile bölgedeki oyuncu sayısı tespit edilerek bir taktik oluşturulabilir. Bu çalışmadaki bilgi kazanım değerlerine göre ev sahibi takımlara Serie A için 41221 taktiği önerilebilir. Bu taktiğin, eğitim veri setinde en çok kullanılan taktik olduğu tespit edilmiştir. Ancak bu ligde 41122 taktiği sadece deplasman takımları tarafından kullanıldığı görülmüştür. Şu ana kadar tercih edilen taktiklerle arasında bu taktiği tercih etmediği tespit edilmiştir. Bu sonuç, deplasman takımının Serie A'da kazanmak için daha ofansif oynaması gerektiğini göstermektedir. Benzer bir analiz Premier Lig için uygulandığında ev sahibi takım için 31231 ve deplasman için 41131 gibi iki zorlayıcı taktikle karşılaşılmaktadır. Bu durum, Premier Lig'de deplasman takımlarının daha kontrollü futbol oynadığının kanıtlamaktadır. Bilgi kazanım değerleri açısından, Bundesliga'da ev sahibi takım için 41131 ve deplasman takımı için 31231 taktiği tavsiye edilebilir. Bu durum Serie A'ya paralellik göstermektedir. Diğer bir ifade ile ev sahibi takımlar bir maçı kazanmak için savunma yapılarını düzenlemektedirler.

4. SONUÇ VE ÖNERİLER

Bu çalışma ile bir futbol maçındaki olası üç sonuç olan galibiyet, beraberlik ve mağlubiyet sonuçlarını bulmak için genişletilebilir tahmin sistemi geliştirilmiştir. Bu tahmin sisteminde takım sıralaması, takım adı, maç haftası gibi özelliklerin yanı sıra literatürde ilk kez önerilen yeni özelliklerden menajer taktiği ve oyuncunun transfer piyasa değeri gibi özellikleri web sayfalarından otomatik olarak toplanmaktadır. Çalışmanın deneysel bölümünde, özelliklerin etkisini görmek amacıyla üç büyük lige ait 2014-2020 sezonları içeren veri seti toplanmıştır. Geliştirilen tahmin sisteminin değerlendirilmesinde 2014-2019 sezonlarına ait veriler eğitim verisi olarak kullanılırken 2019-2020 sezonuna ait veriler test verisi olarak kullanılmıştır. İlgili tahmin sisteminde üç popüler veri madenciliği yönteminin performans sonuçları karşılaştırılmıştır. Sistemimiz ile Serie A, Premier Lig ve Bundesliga için sırasıyla 0.600, 0584 ve 0.554 doğruluk değerleri elde edilmektedir. Elde edilen sonuçların analizinden sonra takımın ek organizasyonlarda yer alması, kış transfer sezonu, ligin son haftaları gibi bazı faktörlerin sonuçlara olumsuz etkisi olduğu ortaya çıkmıştır. Sonuçlara olumsuz etkisi olan bu haftalar çıkarılarak oluşturulan alt test veri setinde bu ligler için sırasıyla Serie A için 0,619, Premier Lig için 0,600 ve Bundesliga için 0,582 doğruluk değerleri elde edilmiştir.

İnternette ulaşım ve kullanım oranının gün ve gün artması verilerin hızla artmasına ve büyük bir veri kaynağının ortaya çıkmasına sebep olmuştur. Veri madenciliği çalışmaları bu verilerden anlamlı bilgiler ortaya çıkarmak için ortaya çıkmıştır. Ancak internet ortamındaki ham haldeki bu veri kaynağını veri madenciliği için uygun hale getirmek bile zahmetli bir iştir. Bu çalışmada, web sayfalarından futbolcuların pozisyonları ve piyasa değerlerine ilişkin veriler ayrıştırılmıştır. Bir futbol maçı öncesi takımlara ait yorgunluk, moral ve sakatlık durumu ile futbolcuyla ilgili magazin haberleri farklı web sayfalarından toplanabilmektedir. Diğer taraftan bir maçın sonuna etki eden örneğin bir maçın taraftarları, hakemi veya yöneticisindeki değişiklik gibi daha birçok faktör üretilebilir. Bu çalışmada geliştirilen sistem için maç öncesi verileri kullanılmıştır. Ancak utulmamalıdır ki maç esnasında meydana gelen sakatlık, kırmızı kart, sarı kart, gol ve değişen taktikler gibi maçın sonucunu değiştiren birçok faktör vardır. Maç boyunca bu veriler internet üzerinden anlık olarak toplanabilir. Anlık verileri toplamak için daha karmaşık web tarayıcıları geliştirilebilir. Bu sistemi maç esnasındaki verileri de kullanarak daha da iyi gale getirmek için verilen web tarama sürecini hızlandıracak web tarayıcısı üzerinde çalışılabilir. Bir diğer etken de internetteki futbolseverlerin yorumlarıdır. Bahis sitelerinde dahi çok fazla yorum mevcuttur. Bu yorumları otomatik olarak keşfederek (Uçar, Uzun, & Tüfekci,

2017) aralarındaki spam yorumları bulmak ve tahmin modelini geliřtirmek mmkndr. Mevcut sistemin geliřtirilmesi iin bu konuda alıřma yapılabilir.

Bu alıřmada, sistemi analiz etmek iin  veri madencilięi yntemi ve  lig zerinde alıřılmıřtır. Bir bařka deyiřle kurmuř olduęumuz sistem iin 9 farklı tahmin modeli oluřturulmuřtur. İncelemeye konu olan her ligin kendine has zelliklerinin olduęu tespit edilmiřtir. rnek olarak Serie A iin bilgi kazanımı deęerlerini incelenirken stoper, merkez orta saha gibi oyuncuların ma sonularında etkin rol oynadıkları grlmřtur. Bundesliga'da ise merkez orta saha ve forvet oyuncularının tahmin modelini pozitif ynde etkiledięi grlmřtur. Premier Lig'de saę kanat ve kaleci pozisyonlarının oyunun sonucu zerinde nemli bir etkisinin olduęu grlmřtur. İngiliz futbolunun kanatlardan yapılan orta ile sonuca gittięi gereęini kurmuř olduęumuz sistem kanıtlamıřtır. Kurmuř olduęumuz sistem ile arařtırma konusu olmayan dięer liglerin karakteristik zellikleri arařtırılabilir. Bu makale konusu alıřmamız ile ileriye ynelik olarak farklı ligler ve farklı veri madencilięi yntemleri kullanarak yukarıda saydıęımız zellikler de dahil edilerek sistemimiz daha da iyileřtirilebilir.

KAYNAKLAR

- Dixon, M., & Coles, S. (1997). *Modelling Association Football Scores and Inefficiencies in the Football Betting Market*. *Applied Statistics*, 46(2), 265-280. doi:<https://doi.org/10.1111/1467-9876.00065>
- Joseph, A., Marco, B., Nelson, B., Sears, R., & Tygar, J. (2006). *Can machine learning be secure? Proceedings of the 2006 ACM Symposium on Information, computer and communications security (s. 16-25)*. New York, NY, USA: Association for Computing Machinery. doi:10.1145/1128817.1128824
- Karlis, D., & Ntzoufras, I. (2009). *Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference*. *IMA Journal of Management Mathematics*, 20(2), 133–145. doi:10.1093/imaman/dpn026
- Lessmann, S., Sung, M.-C., & Johnson, J. (2010). *Alternative methods of predicting competitive events: An application in horserace betting markets*. *International Journal of Forecasting*, 26(3), 1-38. doi:DOI:10.1016/j.ijforecast.2009.12.013
- Panigrahi, R., & Borah, S. (2018). *Rank Allocation to J48 Group of Decision Tree Classifiers using Binary and Multiclass Intrusion Detection Datasets*. *International Conference on Computational Intelligence and Data Science (ICCIDS 2018)*. 132, s. 323-332. ScienceDirect. doi:<https://doi.org/10.1016/j.procs.2018.05.186>
- Rue, H., & Salvesen, Q. (2000). *Prediction and retrospective analysis of soccer matches in a league*. *The Statistician*, 49(3), 399-418. doi: <https://doi.org/10.1111/1467-9884.00243>
- Rue, H., & Salvesen, O. (1997). *Prediction and retrospective analysis of soccer matches in a league*. *Royal Statistical Society*, 265-280.
- Sucky, R. (2020, 9). *Multiclass Classification Using Logistic Regression from Scratch in Python: Step by Step Guide*. <https://towardsdatascience.com/multiclass-classification-algorithm-from-scratch-with-a-project-in-python-step-by-step-guide-485a83c79992> adresinden alındı
- Asif, M., & McHale, I. (2016). *In-play forecasting of win probability in One-Day International cricket: A dynamic logistic regression model*. *International Journal of Forecasting-Elsevier*, 32(1), 34-43. doi:<https://doi.org/10.1016/j.ijforecast.2015.02.005>
- Baboota, R., & Kaur, H. (2019). *Predictive analysis and modelling football results using machine learning approach for English Premier League*. *International Journal of Forecasting*, 35(2), 741-755. doi:<https://doi.org/10.1016/j.ijforecast.2018.01.003>
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. (2011). *Data mining for credit card fraud: A comparative study*. *Elsevier*, 50(3), 602-613. doi:<https://doi.org/10.1016/j.dss.2010.08.008>

- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: springer.
- Boshnakov, G., Kharra, T., & McHale, I. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2), 458-466. doi:<https://doi.org/10.1016/j.ijforecast.2016.11.006>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123-140. doi:<https://doi.org/10.1007/BF00058655>
- Constantinou, A., Fenton, N., & Neil, M. (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, 36, 332-339. doi:DOI:10.1016/j.knosys.2012.07.008
- corporatefinanceinstitute. (2021, 10 12). What is Bagging (Bootstrap Aggregation)? corporatefinanceinstitute: <https://corporatefinanceinstitute.com/resources/knowledge/other/bagging-bootstrap-aggregation/> adresinden alındı
- Crowder, M., Dixon, M., Ledford, A., & Robinson, M. (2002). Dynamic modelling and prediction of English Football League matches for betting. *Journal of the Royal Statistical Society*, 51(2), 157-168. doi:DOI:10.1111/1467-9884.00308
- Fenton, N., & Constantinou, A. (2013). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries. *Journal of Quantitative Analysis in Sports*, 9(1), 37-50. doi:10.1515/jqas-2012-0036
- Forrest, D., & Simmons, R. (2000). Forecasting sport: the behaviour and performance of football. *International Journal of Forecasting*, 16(3), 317-331. doi:[https://doi.org/10.1016/S0169-2070\(00\)00050-9](https://doi.org/10.1016/S0169-2070(00)00050-9)
- Gibson, A., Nicholson, C., Patterson, J., Warrick, M., Black, A., Kokorin, V., . . . Eraly, S. (2016, Mayıs). Deeplearning4j: Distributed, opensource deep learning for java and scala on hadoop and spark. <https://figshare.com/>: https://figshare.com/articles/code/deeplearning4j-deeplearning4j-parent-0_4-rc3_8_zip/3362644 adresinden alındı
- Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21(2), 331-340. doi:<https://doi.org/10.1016/j.ijforecast.2004.08.002>
- Hand, D. (2012). *Principles of Data Mining*. Drug-Safety. 30, s. 621-622. Springer. doi:10.2165/00002018-200730070-00010
- <https://en.wikipedia.org/wiki/Bundesliga>. (2021, 10 4). Bundesliga. Wikipedia: <https://en.wikipedia.org/wiki/Bundesliga> adresinden alındı

- https://en.wikipedia.org/wiki/Premier_League. (2021, 10 4). [wikipedia.org](https://en.wikipedia.org/wiki/Premier_League).
https://en.wikipedia.org/wiki/Premier_League:
https://en.wikipedia.org/wiki/Premier_League adresinden alındı
- https://en.wikipedia.org/wiki/Serie_A. (2021, 10 4). *Serie A*. *Wikipedia*:
https://en.wikipedia.org/wiki/Serie_A adresinden alındı
- <https://www.mathworks.com/>. (2021, 10 12). *What Is Deep Learning? 3 things you need to know. 3 things you need to know*: <https://www.mathworks.com/discovery/deep-learning.html> adresinden alındı
- Joseph, A., Fenton, N., & Neil, M. (2006). *Predicting football results using Bayesian nets and other machine learning techniques*. *Knowledge-Based Systems*, 19(7), 544-553. doi:<https://doi.org/10.1016/j.knosys.2006.04.011>
- Kaur, H., Lechman, E., & Marszk, A. (2017). *Catalyzing Development through ICT Adoption*. Springer International Publishing.
- Koopman, S., & Lit, R. (2012). *A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League*. *Journal of the Royal Statistical Society*, 178(1), 167-186. doi:[DOI:10.2139/ssrn.2154792](https://doi.org/10.2139/ssrn.2154792)
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). *Imagenet classification with deep convolutional neural networks*. *Communications of the ACM*, 60(6), 84-90. doi:<https://doi.org/10.1145/3065386>
- Lang, S., Bravo-Marquez, F., Beckham, C., Hall, M., & Frank, E. (2019). *Weka deep learning 4j: A deep learning package for weka based on deep learning 4j*. *Knowledge-Based Systems*, 178, 48-50. doi:[DOI:10.1016/j.knosys.2019.04.013](https://doi.org/10.1016/j.knosys.2019.04.013)
- Litjens, G., Kooi, T., Bejnordi, B., Setio, A., Ciompi, F., Ghafoorian, M., . . . Sánchez, C. (2017). *A survey on deep learning in medical image analysis*. *Medical Image Analysis*, 42, 60-88. doi:[doi: 10.1016/j.media.2017.07.005](https://doi.org/10.1016/j.media.2017.07.005).
- Maher, M. (1982). *Modelling association football scores*. *Statistica Neerlandica*, 36(3), 109-118. doi: <https://doi.org/10.1111/j.1467-9574.1982.tb00782.x>
- Manikandan, G., Vasudev, A., & Balasubramanian, A. (2018). *A Survey to Identify an Efficient Classification Algorithm for Heart Disease Prediction*. *International Journal of Pure and Applied Mathematics*, 119(12), 13337-13345. <https://core.ac.uk/download/pdf/160114503.pdf> adresinden alındı
- Meshram, S., Safari, M., Khosravi, K., & Meshram, C. (2021). *Iterative classifier optimizer-based pace regression and random*. Springer, 28(1), 11637–11649. doi:[DOI:10.1007/s11356-020-11335-5](https://doi.org/10.1007/s11356-020-11335-5)

- Odachowski, K., & Grekow, J. (2013). *Using Bookmaker Odds to Predict the Final Result of Football Matches. Proceedings of the 16th international conference on Knowledge Engineering, Machine Learning and Lattice Computing with Applications*, (s. 196-205).
- Owramipur, F., Eskandarian, P., & Mozneb, F. (2013). *Football Result Prediction with Bayesian Network in Spanish League-Barcelona Team. International Journal of Computer Theory and Engineering*, 5(5), 812-815. doi:DOI:10.7763/IJCTE.2013.V5.802
- Peeters, T. (2018). *Testing the Wisdom of Crowds in the field: Transfermarkt. International Journal of Forecasting*, 34(1), 17-29. doi:DOI:10.1016/j.ijforecast.2017.08.002
- Powers, D. (2011). *Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. Journal of Machine*, 2(1), 37-63. https://www.researchgate.net/publication/228529307_Evaluation_From_Precision_Recall_and_F-Factor_to_ROC_Informedness_Markedness_Correlation_adresinden_alindi
- Quinlan, J. (1993). *C4.5: Programs for Machine Learning. San Mateo, California: Morgan Kaufmann Publishers Inc.(eds).*
- Quinlan, J. (1986). *Induction of decision trees. Machine Learning*, 1, 81-106. doi:http://dx.doi.org/10.1007/BF00116251
- Rahul Baboota, H. K. (2019). *Predictive analysis and modelling football results using machine learning approach for English Premier League. International Journal of Forecasting*, 35(2), 741-755. doi:https://doi.org/10.1016/j.ijforecast.2018.01.003
- Rish, I. (2001). *An Empirical Study of the Naïve Bayes Classifie. ResearchGate*, 41-46. doi:doi=10.1.1.330.2788
- Rudrapal, D., Boro, s., Srivastava, J., & Shyamu, S. (2019). *A Deep Learning Approach to Predict Football Match Result. Spinger*, 93-99. doi: https://doi.org/10.1007/978-981-13-8676-3_9
- Schmidhuber, J., Meier, U., & Ciresan, D. (2012). *Multi-column deep neural networks for image classification. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (s. 3642-3649). Los Alamitos, CA, USA: IEEE Computer Society. doi:DOI:10.1109/CVPR.2012.6248110
- Simmons, R., Bryson, A., & Frick, B. (2013). *The Returns to Scarce Talent: Footedness and Player Remuneration in European Soccer. Journal of Sports Economics*, 14(6), 607-628. doi:DOI:10.1177/1527002511435118
- Sperandei, S. (2014). *Understanding logistic regression analysis. Biochemia Medica*, 12-18.
- Stehman, S. (1997). *Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment*, 62(1), 77-89. doi:https://doi.org/10.1016/S0034-4257(97)00083-7

- Štrumbelj, E., & Vračar, P. (2012). *Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. International Journal of Forecasting*, 532-542.
- The Universty of Waikato. (2021, 30 09). <https://www.waikato.ac.nz/>
<https://www.waikato.ac.nz/>: <https://deeplearning.cms.waikato.ac.nz/> adresinden alındı
- TIGHE, S. (2012, Ekim 18). *bleacherreport. bleacherreport:*
<https://bleacherreport.com/articles/1375589-15-tactical-formations-and-what-theyre-good-for> adresinden alındı
- transfermarkt. (2021, 10 11). *transfermarkt. transfermarkt:* <https://www.transfermarkt.com.tr/>
adresinden alındı
- Transfermarkt. (2021, 10 12). *transfermarkt. Transfermarkt:*
<https://www.transfermarkt.com.tr/> adresinden alındı
- Uçar, E., Uzun, E., & Tüfekci, P. (2017). *A novel algorithm for extracting the user reviews from web pages. Journal of Information Science*, 45(3), 696-712.
doi:<https://doi.org/10.1177/0165551516666446>
- University of Waikato. (2021, 9 90). <https://www.cs.waikato.ac.nz/>
<https://www.cs.waikato.ac.nz/>: <https://www.cs.waikato.ac.nz/ml/weka/> adresinden alındı
- UZUN, E. (2020). *A novel web scraping approach using the additional information obtained from web pages. IEEE Access*, 8, 61726–61740. doi:10.1109/ACCESS.2020.2984503
- UZUN, E. (2020). *A regular expression generator based on CSS selectors for efficient extraction from HTML pages. Turkish Journal of Electrical Engineering & Computer Sciences*, 28, 3389–3401. doi:doi:10.3906/elk-2004-67
- Uzun, E. (2021, 11 30). 04.04. *Naive Bayes Classifier.* <https://erdincuzun.com/>:
https://erdincuzun.com/makine_ogrenmesi/naive-bayes-classifier/ adresinden alındı
- Uzun, E., & Özhan, E. (2018). *Examining the impact of feature selection on classification of user reviews in web pages. 2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, (s. 1-8).
- UZUN, E., KIRAT, o., & YERLİKAYA, T. (2018). *Comparision Of Pyhton Libraries Used For Web Data Extraction. International Scientific Conference on Engineering, Technologies and Systems TECHSYS 2018*. 24, s. 87-92. Sofia, Plovdiv branch, Bulgaria: Technical University – Sofia.
- Vračar, P., Štrumbelj, E., & Kononenko, I. (2016). *Modeling basketball play-by-play data. Expert Systems with Applications-Elsevier*, 58-66.
- YERLİKAYA, T., KIRAT, O., & UZUN, E. (2018). *Comparision of Phyton Libraries Used For Web Data Extraction. International Scientific Conference on Engineering, Technologies and*

Systems TECHSYS 2018. 24, s. 87-92. Sofia, Plovdiv brach, Bulgaria: Technical Universty-Sofia.

Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. Decision Support Systems, 50(3), 570-575. doi:10.1016/j.dss.2010.08.007



EK-1. Serie A – Ev Sahibi Taktik ve Yüzdeleri

Taktik	Toplam	Kazanma	Berberlik	Kaybetme	Kazanma%	Berberlik%	Yüzde%
41221	593	259	146	188	43.68%	24.62%	31.70%
31402	434	173	103	158	39.86%	23.73%	36.41%
41212	325	137	84	104	42.15%	25.85%	32.00%
42031	104	50	23	31	48.08%	22.12%	29.81%
41131	92	45	28	19	48.91%	30.43%	20.65%
30412	91	45	18	28	49.45%	19.78%	30.77%
31312	68	29	20	19	42.65%	29.41%	27.94%
32221	62	33	12	17	53.23%	19.35%	27.42%
40402	56	24	14	18	42.86%	25.00%	32.14%
41122	46	27	8	11	58.70%	17.39%	23.91%
30421	40	17	11	12	42.50%	27.50%	30.00%
31321	39	17	11	11	43.59%	28.21%	28.21%
30502	36	19	8	9	52.78%	22.22%	25.00%
40231	36	20	10	6	55.56%	27.78%	16.67%
40222	29	16	2	11	55.17%	6.90%	37.93%
32212	24	14	5	5	58.33%	20.83%	20.83%
41302	23	9	4	10	39.13%	17.39%	43.48%
40312	22	11	2	9	50.00%	9.09%	40.91%
41311	17	2	4	11	11.76%	23.53%	64.71%
31222	16	2	6	8	12.50%	37.50%	50.00%
40321	11	7	2	2	63.64%	18.18%	18.18%
40411	11	4	1	6	36.36%	9.09%	54.55%
42022	10	4	3	3	40.00%	30.00%	30.00%
41401	10	1	3	6	10.00%	30.00%	60.00%
52102	10	3	2	5	30.00%	20.00%	50.00%
52012	9	4	1	4	44.44%	11.11%	44.44%
42121	8	3	2	3	37.50%	25.00%	37.50%
32302	7	4	1	2	57.14%	14.29%	28.57%
42202	6	2	1	3	33.33%	16.67%	50.00%
30403	5	3	1	1	60.00%	20.00%	20.00%
42112	4	3	1	0	75.00%	25.00%	0.00%
43012	4	0	1	3	0.00%	25.00%	75.00%
41041	3	2	1	0	66.67%	33.33%	0.00%
31231	3	2	1	0	66.67%	33.33%	0.00%
30322	3	2	1	0	66.67%	33.33%	0.00%
32203	3	2	1	0	66.67%	33.33%	0.00%
31303	3	1	0	2	33.33%	0.00%	66.67%
41203	3	0	2	1	0.00%	66.67%	33.33%
43021	2	1	0	1	50.00%	0.00%	50.00%
30241	1	1	0	0	100.00%	0.00%	0.00%
42301	1	0	0	1	0.00%	0.00%	100.00%
32131	1	1	0	0	100.00%	0.00%	0.00%
40303	1	0	0	1	0.00%	0.00%	100.00%

50401	1	0	1	0	0.00%	100.00%	0.00%
33202	1	0	0	1	0.00%	0.00%	100.00%
30232	1	1	0	0	100.00%	0.00%	0.00%
32032	1	0	1	0	0.00%	100.00%	0.00%
51121	1	0	1	0	0.00%	100.00%	0.00%
31132	1	1	0	0	100.00%	0.00%	0.00%
51112	1	1	0	0	100.00%	0.00%	0.00%
42211	1	1	0	0	100.00%	0.00%	0.00%
2280	1003	547	730	43.99%	23.99%	32.02%	

EK-2. Serie A – Deplasman Taktik ve Yüzdeleri

Taktik	Toplam	Kazanma	Beraberlik	Kaybetme	Kazanma%	Beraberlik%	Yüzde%
41221	552	188	135	229	34.06%	24.46%	41.49%
31402	459	141	101	217	30.72%	22.00%	47.28%
41212	306	74	67	165	24.18%	21.90%	53.92%
42031	113	53	25	35	46.90%	22.12%	30.97%
41131	92	38	25	29	41.30%	27.17%	31.52%
30412	76	26	27	23	34.21%	35.53%	30.26%
32221	74	24	19	31	32.43%	25.68%	41.89%
31312	64	24	18	22	37.50%	28.13%	34.38%
31321	53	15	18	20	28.30%	33.96%	37.74%
40402	50	13	12	25	26.00%	24.00%	50.00%
40231	42	16	11	15	38.10%	26.19%	35.71%
30421	37	7	11	19	18.92%	29.73%	51.35%
30502	33	9	7	17	27.27%	21.21%	51.52%
40222	31	11	7	13	35.48%	22.58%	41.94%
41122	27	16	4	7	59.26%	14.81%	25.93%
40312	27	4	7	16	14.81%	25.93%	59.26%
52102	24	0	6	18	0.00%	25.00%	75.00%
41311	21	5	7	9	23.81%	33.33%	42.86%
41401	19	5	4	10	26.32%	21.05%	52.63%
32212	19	6	4	9	31.58%	21.05%	47.37%
40411	17	10	1	6	58.82%	5.88%	35.29%
32302	16	4	3	9	25.00%	18.75%	56.25%
21402	15	4	5	6	26.67%	33.33%	40.00%
41302	14	4	3	7	28.57%	21.43%	50.00%
31222	12	4	3	5	33.33%	25.00%	41.67%
42022	11	6	1	4	54.55%	9.09%	36.36%
40321	8	5	2	1	62.50%	25.00%	12.50%
41203	6	3	1	2	50.00%	16.67%	33.33%
52012	5	1	2	2	20.00%	40.00%	40.00%
42112	5	2	1	2	40.00%	20.00%	40.00%
30322	5	1	3	1	20.00%	60.00%	20.00%
31303	5	0	0	5	0.00%	0.00%	100.00%
31231	4	1	1	2	25.00%	25.00%	50.00%

42202	4	2	0	2	50.00%	0.00%	50.00%
52021	3	0	0	3	0.00%	0.00%	100.00%
43012	3	0	1	2	0.00%	33.33%	66.67%
42121	3	1	0	2	33.33%	0.00%	66.67%
21312	3	0	2	1	0.00%	66.67%	33.33%
42211	3	2	0	1	66.67%	0.00%	33.33%
41041	2	0	1	1	0.00%	50.00%	50.00%
20421	2	0	1	1	0.00%	50.00%	50.00%
50311	1	0	0	1	0.00%	0.00%	100.00%
42103	1	1	0	0	100.00%	0.00%	0.00%
31042	1	1	0	0	100.00%	0.00%	0.00%
32203	1	1	0	0	100.00%	0.00%	0.00%
30331	1	0	0	1	0.00%	0.00%	100.00%
31132	1	1	0	0	100.00%	0.00%	0.00%
22212	1	0	0	1	0.00%	0.00%	100.00%
32122	1	0	0	1	0.00%	0.00%	100.00%
20412	1	0	0	1	0.00%	0.00%	100.00%
50401	1	0	0	1	0.00%	0.00%	100.00%
40501	1	0	0	1	0.00%	0.00%	100.00%
52201	1	0	1	0	0.00%	100.00%	0.00%
51112	1	0	0	1	0.00%	0.00%	100.00%
50221	1	0	0	1	0.00%	0.00%	100.00%
31411	1	1	0	0	100.00%	0.00%	0.00%
2280	730	547	1003	32.02%	23.99%		

EK-3. Premier Lig – Ev Sahibi Taktik ve Yüzdeleri

Taktik	Toplam	Kazanma	Berberlik	Kaybetme	Kazanma%	Berberlik%	Yüzde%
41221	508	240	120	148	47.24%	23.62%	29.13%
41131	359	176	83	100	49.03%	23.12%	27.86%
42031	159	74	30	55	46.54%	18.87%	34.59%
41212	145	63	41	41	43.45%	28.28%	28.28%
32221	104	55	22	27	52.88%	21.15%	25.96%
41122	94	39	23	32	41.49%	24.47%	34.04%
31402	90	27	19	44	30.00%	21.11%	48.89%
40402	84	39	15	30	46.43%	17.86%	35.71%
42022	83	40	24	19	48.19%	28.92%	22.89%
41311	75	29	16	30	38.67%	21.33%	40.00%
40312	57	17	14	26	29.82%	24.56%	45.61%
42121	56	20	21	15	35.71%	37.50%	26.79%
42112	48	15	16	17	31.25%	33.33%	35.42%
52021	40	18	8	14	45.00%	20.00%	35.00%
42202	36	16	8	12	44.44%	22.22%	33.33%
32212	32	16	7	9	50.00%	21.88%	28.13%
31312	32	16	9	7	50.00%	28.13%	21.88%
41302	22	9	7	6	40.91%	31.82%	27.27%

52012	22	9	3	10	40.91%	13.64%	45.45%
42211	22	10	5	7	45.45%	22.73%	31.82%
40321	20	10	3	7	50.00%	15.00%	35.00%
31321	20	7	4	9	35.00%	20.00%	45.00%
30412	17	6	3	8	35.29%	17.65%	47.06%
51121	17	6	2	9	35.29%	11.76%	52.94%
50221	15	8	1	6	53.33%	6.67%	40.00%
40231	13	7	2	4	53.85%	15.38%	30.77%
30421	13	4	5	4	30.77%	38.46%	30.77%
40222	12	1	7	4	8.33%	58.33%	33.33%
41041	11	8	1	2	72.73%	9.09%	18.18%
40411	11	2	4	5	18.18%	36.36%	45.45%
41401	9	6	0	3	66.67%	0.00%	33.33%
50311	7	2	0	5	28.57%	0.00%	71.43%
32302	5	1	1	3	20.00%	20.00%	60.00%
52201	5	2	1	2	40.00%	20.00%	40.00%
32203	4	3	1	0	75.00%	25.00%	0.00%
52111	4	3	0	1	75.00%	0.00%	25.00%
30502	4	3	1	0	75.00%	25.00%	0.00%
31231	4	3	1	0	75.00%	25.00%	0.00%
52102	3	0	1	2	0.00%	33.33%	66.67%
51211	3	0	2	1	0.00%	66.67%	33.33%
31222	3	1	0	2	33.33%	0.00%	66.67%
50302	2	0	1	1	0.00%	50.00%	50.00%
50401	2	1	1	0	50.00%	50.00%	0.00%
51202	2	0	2	0	0.00%	100.00%	0.00%
40501	1	0	1	0	0.00%	100.00%	0.00%
43021	1	0	0	1	0.00%	0.00%	100.00%
51031	1	1	0	0	100.00%	0.00%	0.00%
30322	1	0	0	1	0.00%	0.00%	100.00%
2278	1013	536	729	44.47%	23.53%	32.00%	

EK-4. Premier Lig – Deplasman Taktik ve Yüzdeleri

Taktik	Toplam	Kazanma	Berberlik	Kaybetme	Kazanma%	Berberlik%	Yüzde%
41221	514	153	126	235	29.77%	24.51%	45.72%
41131	305	126	65	114	41.31%	21.31%	37.38%
42031	152	46	44	62	30.26%	28.95%	40.79%
41212	140	46	28	66	32.86%	20.00%	47.14%
32221	124	61	21	42	49.19%	16.94%	33.87%
41311	95	23	26	46	24.21%	27.37%	48.42%
31402	90	20	26	44	22.22%	28.89%	48.89%
41122	89	30	20	39	33.71%	22.47%	43.82%
40402	80	17	23	40	21.25%	28.75%	50.00%

42022	76	25	11	40	32.89%	14.47%	52.63%
42121	61	20	22	19	32.79%	36.07%	31.15%
40312	52	22	13	17	42.31%	25.00%	32.69%
42112	52	15	13	24	28.85%	25.00%	46.15%
52021	45	11	7	27	24.44%	15.56%	60.00%
42202	34	9	10	15	26.47%	29.41%	44.12%
50221	33	6	10	17	18.18%	30.30%	51.52%
32212	33	10	12	11	30.30%	36.36%	33.33%
52012	33	6	10	17	18.18%	30.30%	51.52%
41302	25	10	1	14	40.00%	4.00%	56.00%
31321	23	10	4	9	43.48%	17.39%	39.13%
31312	22	9	3	10	40.91%	13.64%	45.45%
40321	19	4	7	8	21.05%	36.84%	42.11%
42211	19	7	4	8	36.84%	21.05%	42.11%
51121	18	2	5	11	11.11%	27.78%	61.11%
41041	13	11	1	1	84.62%	7.69%	7.69%
40411	13	3	1	9	23.08%	7.69%	69.23%
30421	13	4	5	4	30.77%	38.46%	30.77%
30502	13	3	1	9	23.08%	7.69%	69.23%
40222	11	4	1	6	36.36%	9.09%	54.55%
41401	10	2	0	8	20.00%	0.00%	80.00%
40231	10	3	4	3	30.00%	40.00%	30.00%
30412	10	1	2	7	10.00%	20.00%	70.00%
31222	8	3	2	3	37.50%	25.00%	37.50%
52111	7	3	1	3	42.86%	14.29%	42.86%
51211	6	1	0	5	16.67%	0.00%	83.33%
50302	6	1	2	3	16.67%	33.33%	50.00%
52102	4	0	1	3	0.00%	25.00%	75.00%
50401	4	1	1	2	25.00%	25.00%	50.00%
32203	3	0	0	3	0.00%	0.00%	100.00%
50311	2	0	0	2	0.00%	0.00%	100.00%
52201	2	0	0	2	0.00%	0.00%	100.00%
51202	2	0	2	0	0.00%	100.00%	0.00%
32302	2	0	0	2	0.00%	0.00%	100.00%
40501	1	0	0	1	0.00%	0.00%	100.00%
41032	1	0	1	0	0.00%	100.00%	0.00%
33202	1	0	0	1	0.00%	0.00%	100.00%
31231	1	1	0	0	100.00%	0.00%	0.00%
31303	1	0	0	1	0.00%	0.00%	100.00%
	2278	729	536	1013	32.00%	23.53%	44.47%

EK-5. Bundesliga – Ev Sahibi Taktik ve Yüzdeleri

Taktik	Toplam	Kazanma	Beraberlik	Kaybetme	Kazanma%	Beraberlik%	Yüzde%
41131	274	138	62	74	50.36%	22.63%	27.01%
41221	239	108	54	77	45.19%	22.59%	32.22%
41122	182	101	38	43	55.49%	20.88%	23.63%
41212	154	64	36	54	41.56%	23.38%	35.06%
42022	106	39	28	39	36.79%	26.42%	36.79%
42031	98	49	19	30	50.00%	19.39%	30.61%
31402	74	31	20	23	41.89%	27.03%	31.08%
31312	69	27	24	18	39.13%	34.78%	26.09%
32221	59	24	18	17	40.68%	30.51%	28.81%
31222	50	27	10	13	54.00%	20.00%	26.00%
40231	45	21	10	14	46.67%	22.22%	31.11%
31321	44	21	9	14	47.73%	20.45%	31.82%
30412	44	19	14	11	43.18%	31.82%	25.00%
32212	35	11	8	16	31.43%	22.86%	45.71%
42112	31	10	9	12	32.26%	29.03%	38.71%
40222	29	14	6	9	48.28%	20.69%	31.03%
52012	28	11	4	13	39.29%	14.29%	46.43%
30421	26	10	11	5	38.46%	42.31%	19.23%
42121	25	6	5	14	24.00%	20.00%	56.00%
41311	23	9	5	9	39.13%	21.74%	39.13%
42202	22	9	7	6	40.91%	31.82%	27.27%
52021	21	11	8	2	52.38%	38.10%	9.52%
51121	16	5	5	6	31.25%	31.25%	37.50%
41041	16	10	3	3	62.50%	18.75%	18.75%
50221	15	4	3	8	26.67%	20.00%	53.33%
41302	14	4	3	7	28.57%	21.43%	50.00%
40312	10	2	2	6	20.00%	20.00%	60.00%
41401	10	4	4	2	40.00%	40.00%	20.00%
40402	8	2	4	2	25.00%	50.00%	25.00%
42211	7	2	2	3	28.57%	28.57%	42.86%
32122	5	3	0	2	60.00%	0.00%	40.00%
40321	4	2	1	1	50.00%	25.00%	25.00%
31231	4	2	2	0	50.00%	50.00%	0.00%
32203	4	2	2	0	50.00%	50.00%	0.00%
30502	3	2	1	0	66.67%	33.33%	0.00%
51211	3	2	0	1	66.67%	0.00%	33.33%
31132	3	2	0	1	66.67%	0.00%	33.33%
52111	3	0	2	1	0.00%	66.67%	33.33%
51112	3	0	1	2	0.00%	33.33%	66.67%
32302	3	0	0	3	0.00%	0.00%	100.00%
50401	2	1	0	1	50.00%	0.00%	50.00%
50302	2	1	1	0	50.00%	50.00%	0.00%
52102	2	1	1	0	50.00%	50.00%	0.00%

32032	2	0	1	1	0.00%	50.00%	50.00%
30222	2	0	2	0	0.00%	100.00%	0.00%
43012	1	0	0	1	0.00%	0.00%	100.00%
43021	1	0	1	0	0.00%	100.00%	0.00%
42301	1	0	0	1	0.00%	0.00%	100.00%
31303	1	1	0	0	100.00%	0.00%	0.00%
32131	1	1	0	0	100.00%	0.00%	0.00%
30132	1	0	1	0	0.00%	100.00%	0.00%
42021	1	0	0	1	0.00%	0.00%	100.00%
31141	1	0	0	1	0.00%	0.00%	100.00%
30241	1	0	1	0	0.00%	100.00%	0.00%
30331	1	0	1	0	0.00%	100.00%	0.00%
31213	1	0	0	1	0.00%	0.00%	100.00%
30232	1	0	0	1	0.00%	0.00%	100.00%
30403	1	0	0	1	0.00%	0.00%	100.00%
40411	1	1	0	0	100.00%	0.00%	0.00%
52201	1	0	0	1	0.00%	0.00%	100.00%
30322	1	0	0	1	0.00%	0.00%	100.00%
41032	1	0	1	0	0.00%	100.00%	0.00%
	1836	814	450	572	44.34%	24.51%	31.15%

EK-6. Bundesliga – Deplasman Taktik ve Yüzdeleri

Taktik	Toplam	Kazanma	Beraberlik	Kaybetme	Kazanma%	Beraberlik%	Yüzde%
41131	256	99	63	94	38.67%	24.61%	36.72%
41221	224	61	59	104	27.23%	26.34%	46.43%
41122	163	68	35	60	41.72%	21.47%	36.81%
41212	137	46	31	60	33.58%	22.63%	43.80%
42022	107	27	31	49	25.23%	28.97%	45.79%
42031	103	26	27	50	25.24%	26.21%	48.54%
31402	80	25	24	31	31.25%	30.00%	38.75%
31312	73	21	17	35	28.77%	23.29%	47.95%
32221	63	26	10	27	41.27%	15.87%	42.86%
31321	50	22	7	21	44.00%	14.00%	42.00%
31222	39	7	15	17	17.95%	38.46%	43.59%
30421	37	15	8	14	40.54%	21.62%	37.84%
41311	37	7	11	19	18.92%	29.73%	51.35%
40222	35	15	6	14	42.86%	17.14%	40.00%
40231	35	16	8	11	45.71%	22.86%	31.43%
30412	35	8	12	15	22.86%	34.29%	42.86%
42112	35	14	5	16	40.00%	14.29%	45.71%
32212	35	7	9	19	20.00%	25.71%	54.29%
52021	30	5	5	20	16.67%	16.67%	66.67%
52012	27	6	5	16	22.22%	18.52%	59.26%
42121	26	6	7	13	23.08%	26.92%	50.00%

42202	22	6	6	10	27.27%	27.27%	45.45%
51121	19	4	3	12	21.05%	15.79%	63.16%
41041	15	5	4	6	33.33%	26.67%	40.00%
41302	14	3	2	9	21.43%	14.29%	64.29%
40402	14	5	3	6	35.71%	21.43%	42.86%
50221	14	3	7	4	21.43%	50.00%	28.57%
41401	12	2	3	7	16.67%	25.00%	58.33%
42211	11	4	2	5	36.36%	18.18%	45.45%
40312	9	2	1	6	22.22%	11.11%	66.67%
51211	7	1	0	6	14.29%	0.00%	85.71%
32203	7	0	1	6	0.00%	14.29%	85.71%
31231	7	2	2	3	28.57%	28.57%	42.86%
32122	6	0	4	2	0.00%	66.67%	33.33%
50311	5	1	0	4	20.00%	0.00%	80.00%
40411	5	2	1	2	40.00%	20.00%	40.00%
32302	5	1	3	1	20.00%	60.00%	20.00%
40321	4	1	2	1	25.00%	50.00%	25.00%
52102	3	0	2	1	0.00%	66.67%	33.33%
52111	3	0	1	2	0.00%	33.33%	66.67%
30403	2	0	2	0	0.00%	100.00%	0.00%
43021	2	0	2	0	0.00%	100.00%	0.00%
51301	2	0	0	2	0.00%	0.00%	100.00%
51202	2	0	0	2	0.00%	0.00%	100.00%
50302	2	0	1	1	0.00%	50.00%	50.00%
52201	2	1	0	1	50.00%	0.00%	50.00%
30502	2	0	0	2	0.00%	0.00%	100.00%
30232	2	1	0	1	50.00%	0.00%	50.00%
31132	2	0	1	1	0.00%	50.00%	50.00%
51031	1	0	1	0	0.00%	100.00%	0.00%
40132	1	0	0	1	0.00%	0.00%	100.00%
31303	1	0	0	1	0.00%	0.00%	100.00%
50401	1	0	0	1	0.00%	0.00%	100.00%
30222	1	0	0	1	0.00%	0.00%	100.00%
32041	1	0	0	1	0.00%	0.00%	100.00%
32032	1	1	0	0	100.00%	0.00%	0.00%
33202	1	0	1	0	0.00%	100.00%	0.00%
30322	1	0	0	1	0.00%	0.00%	100.00%
	1836	572	450	814	31.15%	24.51%	44.34%

EK-7. Derin Öğrenme için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı

Serie A				
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	118	32	6
	Mağlubiyet	25	84	19
	Beraberlik	50	35	11
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.611	0.756	0.676
	Mağlubiyet	0.556	0.656	0.602
	Beraberlik	0.306	0.115	0.167
Accuracy(Doğruluk)		0.561		

Premier Lig				
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	124	17	3
	Mağlubiyet	52	88	13
	Beraberlik	47	29	7
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.556	0.861	0.676
	Mağlubiyet	0.657	0.575	0.613
	Beraberlik	0.304	0.084	0.132
Accuracy(Doğruluk)		0.576		

Bundesliga				
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	96	30	3
	Mağlubiyet	25	69	2
	Beraberlik	47	30	4
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.571	0.744	0.646
	Mağlubiyet	0.535	0.719	0.613
	Beraberlik	0.444	0.049	0.089
Accuracy(Doğruluk)		0.552		

EK-8. Yinelemeli Sınıflandırıcı İyileştirici için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı

		Serie A		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	126	22	8
	Mağlubiyet	30	90	8
	Beraberlik	57	27	12
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.592	0.808	0.683
	Mağlubiyet	0.647	0.703	0.674
	Beraberlik	0.429	0.125	0.194
Accuracy(Doğruluk)		0.60		

		Premier Lig		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	125	13	6
	Mağlubiyet	57	82	14
	Beraberlik	42	35	6
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.558	0.868	0.679
	Mağlubiyet	0.631	0.536	0.58
	Beraberlik	0.231	0.072	0.11
Accuracy(Doğruluk)		0.561		

		Bundesliga		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	104	17	8
	Mağlubiyet	34	57	5
	Beraberlik	51	21	9
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.55	0.806	0.654
	Mağlubiyet	0.6	0.594	0.597
	Beraberlik	0.409	0.111	0.175
Accuracy(Doğruluk)		0.556		

EK-9. Torbalama için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı

Serie A				
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	122	24	10
	Mağlubiyet	28	88	12
	Beraberlik	44	37	15
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.629	0.782	0.697
	Mağlubiyet	0.591	0.688	0.635
	Beraberlik	0.405	0.156	0.226
Accuracy(Doğruluk)		0.592		
Premier Lig				
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	115	15	14
	Mağlubiyet	36	94	23
	Beraberlik	37	33	13
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.612	0.799	0.693
	Mağlubiyet	0.662	0.614	0.637
	Beraberlik	0.26	0.157	0.195
Accuracy(Doğruluk)		0.584		
Bundesliga				
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	93	21	15
	Mağlubiyet	27	57	12
	Beraberlik	44	25	12
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.567	0.721	0.635
	Mağlubiyet	0.553	0.594	0.573
	Beraberlik	0.308	0.148	0.2
Accuracy(Doğruluk)		0.529		

EK-10. Navie Bayes için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı

		Serie A		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	103	23	30
	Mağlubiyet	25	81	22
	Beraberlik	37	31	28
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.624	0.66	0.642
	Mağlubiyet	0.6	0.633	0.616
	Beraberlik	0.35	0.292	0.318
Accuracy(Doğruluk)		0.558		

		Premier Lig		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	76	22	46
	Mağlubiyet	34	77	42
	Beraberlik	22	32	29
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.576	0.528	0.551
	Mağlubiyet	0.588	0.503	0.542
	Beraberlik	0.248	0.349	0.29
Accuracy(Doğruluk)		0.479		

		Bundesliga		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	65	24	40
	Mağlubiyet	15	49	32
	Beraberlik	25	26	30
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.619	0.504	0.556
	Mağlubiyet	0.495	0.51	0.503
	Beraberlik	0.294	0.37	0.328
Accuracy(Doğruluk)		0.471		

EK-11. J48 Sınıflandırıcı için Tüm Haftaların Dahil Olduğu Test Veri Setinde Liglerin Tahmin Performansı

		Serie A		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	105	22	29
	Mağlubiyet	30	62	36
	Beraberlik	38	33	25
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.607	0.673	0.638
	Mağlubiyet	0.53	0.484	0.506
	Beraberlik	0.278	0.26	0.269
Accuracy(Doğruluk)		0.505		

		Premier Lig		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	87	21	36
	Mağlubiyet	45	73	35
	Beraberlik	36	22	25
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.518	0.604	0.558
	Mağlubiyet	0.629	0.477	0.543
	Beraberlik	0.26	0.301	0.279
Accuracy(Doğruluk)		0.487		

		Bundesliga		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	75	24	30
	Mağlubiyet	36	45	15
	Beraberlik	33	22	26
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.521	0.581	0.549
	Mağlubiyet	0.495	0.469	0.481
	Beraberlik	0.366	0.366	0.342
Accuracy(Doğruluk)		0.477		

EK-12. Derin Öğrenme için Alt Test Veri Setinde Liglerin Tahmin Performansı

Serie A

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	69	12	2
	Mağlubiyet	15	52	11
	Beraberlik	29	15	5
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.611	0.831	0.704
	Mağlubiyet	0.658	0.667	0.662
	Beraberlik	0.278	0.102	0.149
Accuracy(Doğruluk)		0.60		

Premier Lig

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	73	6	3
	Mağlubiyet	27	56	12
	Beraberlik	23	17	3
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.593	0.89	0.712
	Mağlubiyet	0.709	0.589	0.644
	Beraberlik	0.167	0.07	0.098
Accuracy(Doğruluk)		0.60		

Bundesliga

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	56	13	1
	Mağlubiyet	11	41	1
	Beraberlik	24	21	3
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.615	0.8	0.696
	Mağlubiyet	0.547	0.744	0.641
	Beraberlik	0.6	0.063	0.113
Accuracy(Doğruluk)		0.585		

EK-13. Yinelemeli Sınıflandırıcı İyileştirici için Alt Test Veri Setinde Liglerin Tahmin Performansı

		Serie A		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	71	10	2
	Mağlubiyet	17	56	5
	Beraberlik	32	14	3
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.592	0.855	0.7
	Mağlubiyet	0.7	0.718	0.709
	Beraberlik	0.3	0.061	0.102
Accuracy(Doğruluk)		0.619		

		Premier Lig		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	74	7	1
	Mağlubiyet	32	55	8
	Beraberlik	21	18	4
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.583	0.902	0.708
	Mağlubiyet	0.688	0.579	0.629
	Beraberlik	0.308	0.093	0.143
Accuracy(Doğruluk)		0.605		

		Bundesliga		
		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	60	7	3
	Mağlubiyet	17	34	2
	Beraberlik	28	15	5
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.571	0.857	0.686
	Mağlubiyet	0.607	0.642	0.624
	Beraberlik	0.5	0.104	0.172
Accuracy(Doğruluk)		0.579		

EK-14. Torbalama için Alt Test Veri Setinde Liglerin Tahmin Performansı

Serie A

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	68	9	6
	Mağlubiyet	16	54	8
	Beraberlik	25	26	8
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.624	0.819	0.708
	Mağlubiyet	0.684	0.692	0.688
	Beraberlik	0.364	0.163	0.225
Accuracy(Doğruluk)		0.619		

Premier Lig

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	66	8	8
	Mağlubiyet	21	57	17
	Beraberlik	18	17	8
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.629	0.805	0.706
	Mağlubiyet	0.695	0.6	0.644
	Beraberlik	0.242	0.186	0.211
Accuracy(Doğruluk)		0.596		

Bundesliga

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	51	10	9
	Mağlubiyet	12	36	5
	Beraberlik	22	19	7
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.6	0.729	0.658
	Mağlubiyet	0.554	0.679	0.61
	Beraberlik	0.333	0.146	0.203
Accuracy(Doğruluk)		0.55		

EK-15. Navie Bayes için Alt Test Veri Setinde Liglerin Tahmin Performansı

Serie A

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	60	11	12
	Mağlubiyet	14	50	14
	Beraberlik	21	15	13
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.632	0.723	0.674
	Mağlubiyet	0.658	0.641	0.649
	Beraberlik	0.333	0.265	0.295
Accuracy(Doğruluk)		0.586		

Premier Lig

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	66	8	8
	Mağlubiyet	21	57	17
	Beraberlik	18	17	8
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.629	0.805	0.706
	Mağlubiyet	0.695	0.6	0.644
	Beraberlik	0.242	0.186	0.211
Accuracy(Doğruluk)		0.464		

Bundesliga

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	37	9	24
	Mağlubiyet	5	28	20
	Beraberlik	13	14	21
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.673	0.529	0.592
	Mağlubiyet	0.549	0.528	0.538
	Beraberlik	0.323	0.438	0.372
Accuracy(Doğruluk)		0.503		

EK-16. J48 Sınıflandırıcı için Alt Test Veri Setinde Liglerin Tahmin Performansı

Serie A

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	55	9	19
	Mağlubiyet	16	39	23
	Beraberlik	17	17	15
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.625	0.663	0.643
	Mağlubiyet	0.6	0.5	0.545
	Beraberlik	0.263	0.306	0.283
Accuracy(Doğruluk)		0.519		

Premier Lig

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	50	10	22
	Mağlubiyet	30	49	16
	Beraberlik	16	16	11
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.521	0.61	0.562
	Mağlubiyet	0.653	0.516	0.576
	Beraberlik	0.224	0.256	0.239
Accuracy(Doğruluk)		0.50		

Bundesliga

		Tahmin Edilen		
		Galibiyet	Mağlubiyet	Beraberlik
Gerçek	Galibiyet	44	10	16
	Mağlubiyet	22	27	4
	Beraberlik	18	16	14
		Precision	Recall	f -Meaure
Sınıf	Galibiyet	0.524	0.629	0.571
	Mağlubiyet	0.509	0.509	0.509
	Beraberlik	0.412	0.292	0.341
Accuracy(Doğruluk)		0.497		