

Makine Öğrenmesi Kullanarak Şampiyonlar Ligi Müsabakalarının Ev Sahibi Takım Özelinde Kazanma Durumunun Tahmin Edilmesi

Muhammet Semih KELEŞ
İzmir Kâtip Çelebi Üniversitesi
İzmir, Türkiye
semihkeles1997@hotmail.com

I. GİRİŞ (INTRODUCTION)

Futbol tarihin en eski sporlarından biridir. İlk ne zaman oynandığı tam belli olmasa da duvar resimlerinin ve yazılı kaynakların incelenmesi sonucu M.Ö 5000-2500 tarihleri arasında Çin'de ve yine yakın tarihler arasında Mısır'da başladığı belirtilmiştir. [1] Futbol her ülkede farklı şekillerde ve farklı turnuvalarla oynanmıştır. Günümüzde de 1904 yılında kurulan FIFA organizasyonuna bağlı olarak ülkeler arası milli maçlar düzenlenmektedir. İnsanlar tarafından en çok izlenen ve oynanan sporlardan biri olan futbol tarihin bilinen en eski sporlarından biridir. Öyle ki 1950 yılında FIFA tarafından düzenlenen Dünya Kupası organizasyonunda Uruguay ve Brezilya arasında Brezilya Maracanã Stadyumu'nda oynanan final karşılaşmasını 173.850 biletli taraftar izlemiştir. Gazetecilerle birlikte bu sayının 200.000'in üzerinde olduğu belirtilmektedir. 15 Haziran 1954 yılında kurulan UEFA organizasyonu ile de ülkeler kendi sıralamalarına göre liglerinde yer alan takımlardan bir veya birkaç takımı Şampiyonlar Ligi (1955), Avrupa Ligi (1971) ve Konferans Ligi (2021) organizasyonlarına göndermektedir. Dünyanın kulüpler çerçevesinde en büyük organizasyonu olarak bilinen UEFA Şampiyonlar Ligi organizasyonu ise insanlar tarafından çok büyük ilgiyle izlenmektedir. Ülkelerin sıralamalarına göre her ülkeden bir veya birkaç takımın katılımıyla oluşturulan organizasyon ön eleme maçları ile başlayıp, grup aşamaları ile devam edip sonrasında Son 32, Son 16, Çeyrek Final ve Yarı Final karşılaşmaları ile devam edip Final karşılaşması ile son bulmaktadır. Takımların elde ettiği başarılarla ülke puanı belirlenmektedir. Dolayısıyla bu karşılaşmalar ülkeler çapında da büyük önem arz etmektedir. Bu organizasyonla beraber takımlar galibiyet veya beraberlikle sonuçlandığı müsabakalarda gelir de elde etmektedirler. Dolayısıyla futbol takımlarının hem ülke puanı açısından hem de elde edilebilecek gelirler açısından karşılaşmaları kazanmaları oldukça büyük bir önem arz etmektedir. Bu çalışmanın amacı da Şampiyonlar liginde 2016-2022 tarihleri arasında oynanan maçları inceleyerek ev sahibi takım özelinde karşılaşmanın galibiyet, beraberlik ya da mağlubiyet ile sonuçlanma durumunun tahmin edilmesidir. İlgili çalışma 744 örnekten ve "Ev Sahibi Takım Adı", "Deplasman Takım

Adı", "Stadyum", "Ev Sahibi Takım Skoru", "Deplasman Takım Skoru", "Seyirci Sayısı" olmak üzere 6 öznitelikten oluşmaktadır. Ancak çalışmada "Ev Sahibi Takım Skoru" ve "Deplasman Takım Skoru" öznitelikleri kullanılmayacaktır. Bu öznitelikler yerine her takımın ilgili sezonda futbolcularının piyasa değerleri toplamı olan "Ev Sahibi Takım Piyasa Değeri" ve "Deplasman Takım Piyasa Değeri" özniteliği yer almaktadır. İlgili veri seti ile ev sahibi özelinde "Galibiyet", "Beraberlik" ya da "Mağlubiyet" sonuçları tahmin edilecektir. Araştırmada Makine Öğrenmesinin Denetimli Öğrenme (Supervised Learning) modeli, tekniklerinden Sınıflandırma (Classification) tekniğinin, "Galibiyet/Beraberlik/Mağlubiyet" sonuçları olmak üzere 3 farklı etiket seçeneği olduğundan Çoklu Sınıf Sınıflandırma (Multi-class Classification) çeşidi seçilmiştir. Algoritma olarak ise; sınıflandırma işlemlerinde en çok kullanılan Supervised Learning yöntemlerinden biri olan KNN (K-En Yakın Komşu), Logistic Regression (Lojistik Regresyon), Navie Bayes ve Decision Tree (Karar Ağacı) algoritmaları kullanılmıştır.

II. YÖNTEMLER (METHODS)

A. Kısaltmalar

TP	True Positive	Pozitif değeri doğru bir şekilde etiketlemek
TN	True Negative	Negatif değeri doğru bir şekilde etiketlemek
FN	False Negative	Negatif değeri hatalı şekilde etiketlemek
FP	False Positive	Pozitif değeri hatalı şekilde etiketlemek

B. Veri Seti (Data Set)

Bu çalışmada kullanılan veri seti 2016-2022 tarihleri arasında gerçekleşen Şampiyonlar Ligi müsabakalarını

çermektedir. İlgili veri setinde "Ev Sahibi Takım Adı", "Deplasman Takım Adı", "Stadyum", "Ev Sahibi Takımı Kadro Değeri", "Deplasman Takımı Kado Değeri", "Seyirci Sayısı" olmak üzere 6 öznitelik ve 744 örnek bulunmaktadır. İlgili öznitelikler ve açıklamaları *Tablo 1* 'de gösterilmektedir.

Öznitelik	Açıklama
Home_Team	Ev Sahibi Takım
Away_Team	Deplasman Takımı
Stadium	Stadyum Adı
Home_Team_Market_Value	Ev Sahibi Takım Piyasa Değeri
Away_Team_Market_Value	Deplasman Takım Piyasa Değeri
Attendance	Katılım Sayısı

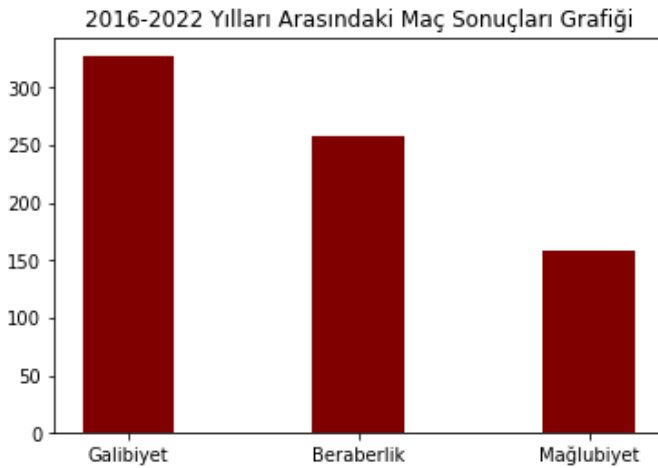
Tablo 1. Veri seti öznitelikleri ve açıklamaları

Müsabaka sonuçlarının "Galibet / Beraberlik / Mağlubiyet" durumu ise *Tablo 2* 'de gösterilmektedir.

Durum	Etiket
Galibiyet	1
Beraberlik	0
Mağlubiyet	2

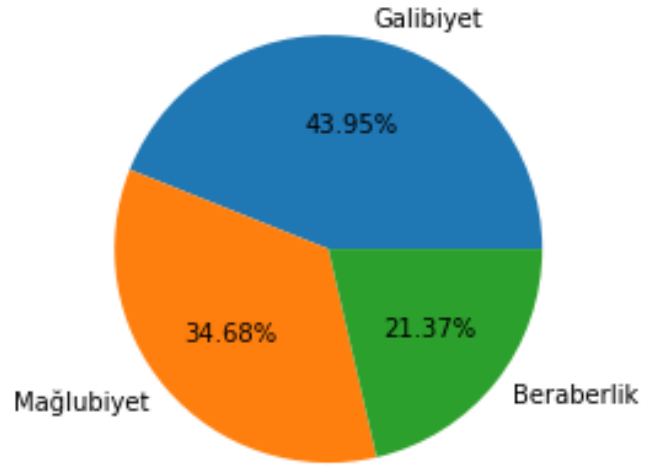
Tablo 2. Müsabaka sonuçları durumu

İlgili veri setinde yer alan 2016-2017, 2017-2018, 2018-2019, 2019-2020, 2020-2021 ve 2021-2022 sezonları olmak üzere toplam 6 sezonda 744 karşılaşma oynanmış ve bu karşılaşmalarda 327 tanesini ev sahibi takım kazanmış, 258 karşılaşma beraberlikle sonuçlanmış ve 159 karşılaşmayı da deplasman takımı kazanmıştır. (*Grafik v1*)



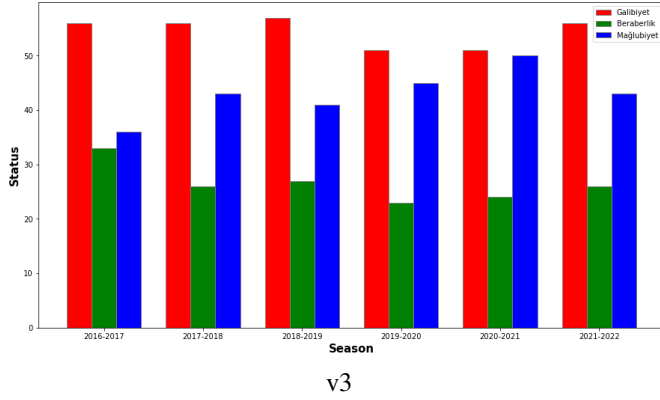
v1

Dolayısıyla verileri pasta grafiğinde yüzdelik dilimde gösterecek olursak ev sahibi takımın galibiyet durumu için %43.95, beraberlikle sonuçlanan karşılaşmalar için %21.37 ve deplasman takımının kazandığı karşılaşmalar için de %34.68 değerlerini görebiliriz. (*Grafik v2*)



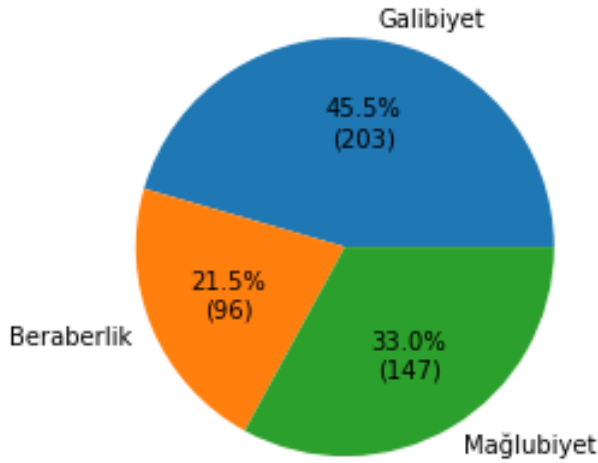
v2

Her sezon için ayrı ayrı incelediğimizde ise 2016-2017 sezonu için 56 karşılaşmayı ev sahibi takımın kazandığını, 33 karşılaşmanın beraberlik ile sonuçlandığını ve 36 karşılaşmanın ise deplasman takımı tarafından kazanıldığını; 2017-2018 sezonu için 56 karşılaşmayı ev sahibi takımın kazandığını, 26 karşılaşmanın beraberlikle sonuçlandığını ve 43 karşılaşmanın deplasman takımı tarafından kazanıldığını; 2018-2019 sezonu için 57 karşılaşmayı ev sahibi takımın kazandığını, 27 karşılaşmanın beraberlikle sonuçlandığını ve 41 karşılaşmanın deplasman takımı tarafından kazanıldığını; 2019-2020 sezonu için 51 karşılaşmayı ev sahibi takımın kazandığını, 23 karşılaşmanın beraberlikle sonuçlandığını ve 45 karşılaşmanın deplasman takımı tarafından kazanıldığını; 2020-2021 sezonu için 51 karşılaşmayı ev sahibi takımın kazandığını, 24 karşılaşmanın beraberlikle sonuçlandığını ve 50 karşılaşmanın deplasman takımı tarafından kazanıldığını; 2021-2022 sezonu için ise 56 karşılaşmanın ev sahibi takım tarafından kazanıldığını, 26 karşılaşmanın beraberlikle sonuçlandığını ve 43 karşılaşmanın deplasman takımı tarafından kazanıldığı görülmektedir. (*Grafik v3*)

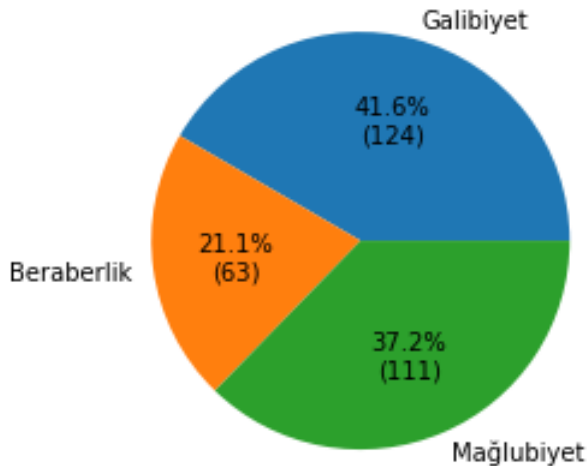


Eğitim ve Test veri seti olmak üzere 744 verinin yaklaşık %60'ına denk gelen 446 veri eğitim için 298 veri ise test için ayrılmıştır. Ayrılan veri, test için ayrılan kısımda ise 124 galibiyet, 63 beraberlik ve 111 mağlubiyet ile sonuçlanan örnek bulunmaktadır.

Eğitim Seti İstatistikler



Test Seti İstatistikler



C. Deneylerde Kullanılan Algoritmalar

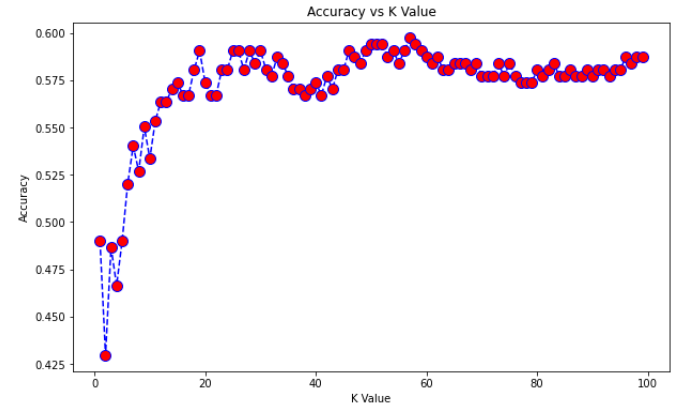
K-En Yakın Komşu Algoritması (K-Nearest Neighbour KNN Algorithm)

Sınıflandırma ve regresyon işlemlerinde kullanılabilen ve sıklıkla tercih edilen algoritmalarından biridir. Özniteliklerin birbirlerine olan uzaklıklarının Öklid işlemi ile hesaplanması ile isminde de geçen K parametresi ile en yakın K sayıdaki örneğe bakarak sınıflandırma işleminin yapıldığı algoritmadır. [2] [3] İşlemin sonunda örneğimiz, K parametresi kadar örneğin içerisinde çoğunlukla hangi sınıfa giriyor ise KNN algoritması çoğunluğun bulunduğu etiketi vermektedir. Örneğin K parametresi 3 verilmiş ve işlemin sonunda ulaşılmış örneklerin sınıf etiketleri "Başarılı", "Başarılı" ve "Başarısız" ise KNN algoritması etiketlemek istediğimiz örneği "Başarılı" etiketi ile etiketleyecektir. Özellikle sınıflandırma işlemlerinde genellikle K değeri tek sayı olarak seçilmektedir. Bunun sebebi işlemin sonunda elde edilecek etiketlerin yarı yarıya olma durumunu engellemektir. Örneğin K değerinin 4 seçilme durumunda ve işlemin sonunda "Başarılı", "Başarılı", "Başarısız" ve "Başarısız" etiketlerine sahip örneklerin getirilmesi sonucunda KNN algoritması sağlıklı bir etiketleme işlemi yapamayacaktır. Regresyon işleminde de aynı işlemler yapıp K kadar örneğin ortalaması hesaplanmaktadır. İlgili örnek için KNN algoritması hazırlanırken öncelikle uygun K değeri bulunmalıdır. Bunun için farklı K değerleri ile model eğitilmiş ve grafiğe dönüştürülmüştür. (Grafik KNN - 1) İlgili tabloda K değerleri ile hata oranları karşılaştırılmıştır.



KNN - 1

Grafikten de anlaşıldığı üzere hata oranı en düşük K değeri 55-60 aralığında görülmektedir. Yapılan çalışmadan en uygun değeri 56 olarak bulunmuştur Aynı şekilde farklı K değerleri üzerinden Accuracy değerleri de hesaplanmış olup en uygun K değeri yine 56 olarak bulunmuştur. (Grafik KNN - 2) K= 56 için Accuracy değeri 0.5906 olarak bulunmuştur.



KNN - 2

İlgili değerlerden yola çıkarak K'nın en uygun değeri 56 olarak bulunmuş olup KNN algoritmasında K değeri 56 olarak belirlenmiştir. Bu değerlerle model eğitildiğinde

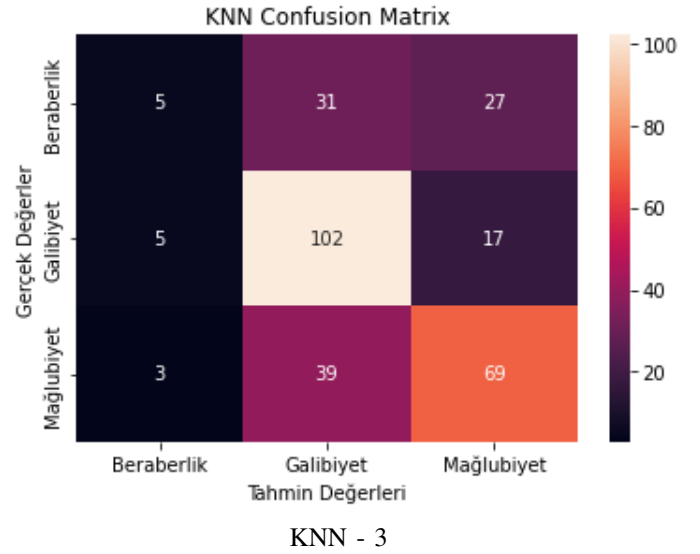
hata oranı 0.402 ve Accuracy değeri ise 0.590 olarak bulunmaktadır. Ancak Accuracy değeri her zaman tek başına yeterli olamayabilmektedir. Çünkü Accuracy değeri doğru olarak yapılan tahminlerin tüm tahminlere oranını almaktadır. Dolayısıyla elde edilen sonuçlar yanıltıcı olabilmektedir. Accuracy değeri ile birlikte Precision ve Recall değerlerinin de hesaplanması gerekir. Daha sonra tüm bu değerlerin harmonik ortalaması alınmalıdır. Precision (Kesinlik) değeri pozitif olarak tahmin edilen değerlerin gerçekten kaç adedinin pozitif olduğunu hesaplamak için kullanılmaktadır. Recall (Duyarlılık) değeri ise pozitif tahin edilmesi gereken değerlerin ne kadarının pozitif olarak tahmin edildiğini göstermektedir. Daha sonra elde edilen Precision ve Recall değerlerinin harmonik ortalaması alınmalıdır. Bunun sebebi veri setindeki uç durumların göz ardı edilmemesi gerektiğindendir. Harmonik ortalama ise F1 Score ile hesaplanabilmektedir.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

K'nın en iyi değeri olan 56 için hesaplanan Precision skoru 0.555, Recall skoru 0.590 ve F_1 skoru ise 0.544 olarak bulunmatadır. İlgili sonuçlar Confusion Matrix ile grafikleştirilmiş ve ortaya KNN - 3 tablosu çıkmıştır. Bu sonuçlardan görüldüğü üzere üretilen model; beraberlik ile sonuçlanan 5 karşılaşmayı (TP), ev sahibi takımın mağlup olduğu 'Mağlubiyet' etiketine sahip 3 karşılaşmayı (FP) ve ev sahibi takımın galip geldiği 'Galibiyet' etiketine sahip 5 karşılaşmayı (FP) da 'Beraberlik' etiketi ile etiketlemiştir. Aynı şekilde ev sahibi takımın galip geldiği 102 karşılaşmayı (TP), 'Beraberlik' ile sonuçlanan 31 karşılaşmayı (FP) ve 'Mağlubiyet' ile etiketlenmiş olan 29 karşılaşmayı (FN) 'Galibiyet' olarak etiketlemiştir. Yine aynı şekilde ev sahibi takımın mağlubiyeti ile sonuçlanan 69 karşılaşmayı (TP), 'Beraberlik' ile sonuçlanan 27 karşılaşmayı (FP) ve 'Galibiyet' ile sonuçlanan 27 karşılaşmayı (FP) 'Mağlubiyet' olarak etiketlemiştir. İlgili sonuçlardan da anlaşılacağı üzere model 'Galibiyet' ile sonuçlanmış karşılaşmaların 'Galibiyet' ile etiketlenmesi konusunda oldukça başarılı, 'Mağlubiyet' etiketine sahip karşılaşmaların 'Mağlubiyet' olarak etiketlenmesi konusunda başarılı ancak 'Beraberlik' ile sonuçlanmış karşılaşmaların 'Beraberlik' olarak etiketlenmesi konusunda oldukça başarısız olmuştur.



Naive Bayes Algoritması

Genellikle metin sınıflandırma projelerinde kullanılan Naive Bayes algoritması supervised (denetimli) öğrenme modeli ile gerçekleşen bir sınıflandırma algoritmasıdır. Bağımsız değişkenin, bağımlı değişken üzerindeki etkilerine göre istatistik ve olasılık temellerine göre yeni bir durumu sınıflandırma işlemidir.

Durumların olasılık değerlerinin hesaplandığı formül:

$$X = [x_1, x_2, x_3, \dots, x_n]$$

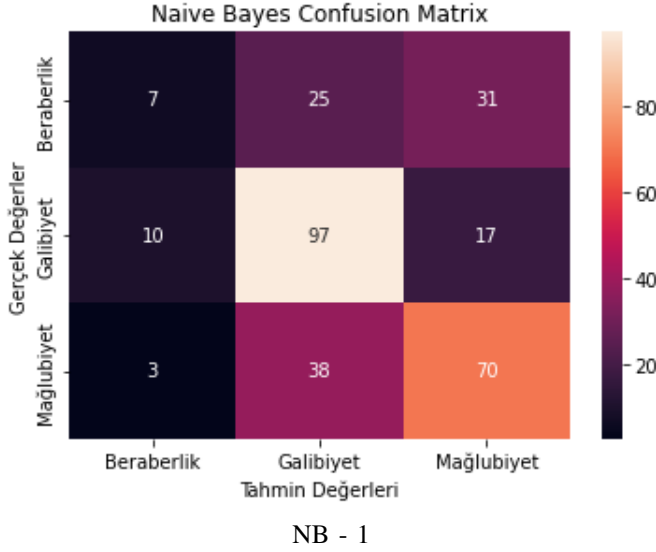
$$C = [c_1, c_2, c_3, \dots, c_n]$$

$$P(C_i/x) = \frac{p(x/C_i)}{P(x)}$$

Yeni durum ise $\text{argmax}_{c_i} P(x/c_i)P(c_i)$ formülü kullanılarak hesaplanan en yüksek olasılık değerinin sınıfına atanır [4] [3]

İlgili örnek için Naive Bayes algoritması kullanıldığında elde edilen Accuracy değeri 0.583, Precision değeri 0.547, Recall değeri 0.583 ve F_1 değeri ise 0.547 olarak bulunmaktadır. Elde edilen sonuçlarla oluşturulan Confusion Matrix'te ise (NB - 1) modelin beraberlik ile sonuçlanan 7 karşılaşmayı (TP), 'Galibiyet' ile sonuçlanan 10 karşılaşmayı (FP) ve 'Mağlubiyet' ile sonuçlanan 3 karşılaşmayı (FP) 'Beraberlik' etiketi ile etikeldiği görülmektedir. Aynı şekilde ev sahibi takımın galibiyeti ile sonuçlanan 97 karşılaşmayı (TP), 'Beraberlik' ile sonuçlanan 25 karşılaşmayı (FP) ve 'Mağlubiyet' ile sonuçlanan 38 karşılaşmayı (FP) 'Galibiyet' olarak etiketlediği görülmektedir. Yine aynı şekilde ev sahibi takımın mağlubiyeti ile sonuçlanan 70 karşılaşmayı (TP), 'Beraberlik' etiketi ile sonuçlanan 31 karşılaşmayı (FP) ve 'Galibiyet' etiketi ile sonuçlanan 17 karşılaşmayı 'Mağlubiyet' olarak etiketlediği görülmektedir. İlgili sonuçlara bakıldığında ilgili algoritmanın 'Galibiyet' ile etiketlenmiş karşılaşmaları 'Galibiyet' ile etiketleme konusunda ve 'Mağlubiyet' ile etiketlenmiş karşılaşmaları 'Mağlubiyet' ile etiketleme konusunda başarılı sayılabileceği

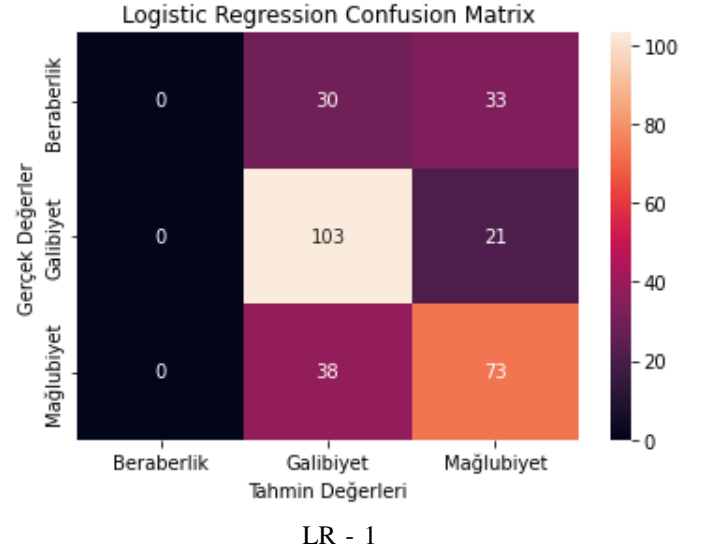
görüldürken 'Beraberlik' etiketi ile etiketlenmiş karşılaşmaları 'Beraberlik' ile etiketleme konusunda oldukça başarısız olduğu görülmektedir.



Lojistik Regresyon (Logistic Regression)

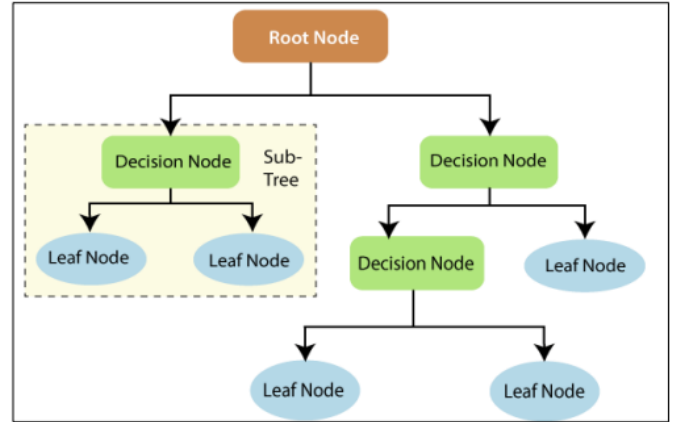
Lojistik Regresyonda bağımlı ve bağımsız değişken arasında en az değişken ile en iyi uyum sağlamak amaçlanmaktadır. Lojistik Regresyon, bir sınıfı tahmin edebilmek için bağımsız çoklu değişkenlerin etkisini hesaplamaktadır. [5]

İlgili örnek için Lojistik Regresyon algoritması kullanıldığında elde edilen Accuracy değeri 0.590, Predision değeri 0.464, Recall değeri 0.590 ve F1 değeri ise 0.519 olarak gözlemlenmiştir. Elde edilen sonuçlarla oluşturulan Confusion Matrix'te (LR - 1) ise beraberlik ile sonuçlanan hiçbir karşılaşmayı sınıflayamadığını, 'Beraberlik' etiketi ile etiketlenen 30 karşılaşmayı (FP), 'Galibiyet' etiketi ile etiketlenen 103 karşılaşmayı (FP) ve 'Mağlubiye' etiketi ile etiketlenen 38 karşılaşmayı (FP) 'Galibiyet' olarak etiketlediği; beraberlikle sonuçlanan 33 karşılaşmayı (FP), 'Galibiyet' ile etiketlenen 21 karşılaşmayı (FP) ve 'Mağlubiye' ile etiketlenen 73 karşılaşmayı 'Mağlubiye' etiketi ile etiketlediği görülmektedir. Bu sonuçlar ışığında Lojistik Regresyon'un 'Galibiyet' ile etiketlenmiş örnekleri 'Galibiyet' ile etiketleme konusunda oldukça başarılı, 'Mağlubiye' ile etiketlenen örnekleri 'Mağlubiye' etiketi ile etiketleme konusunda başarılı ancak 'Beraberlik' ile etiketlenen örnekleri 'Beraberlik' ile etiketleme konusunda oldukça başarısız olduğu görülmektedir.



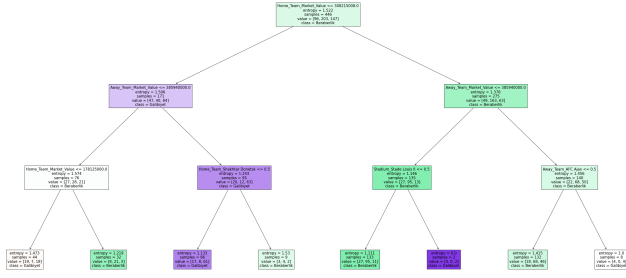
Karar Ağacı (Decision Tree)

Karar Ağacı isminden de anlaşılacağı üzere ağaç biçiminde bir yapı oluşturularak karar verme işleminin yapıldığı bir yapıdır. Karar ağaçlarında karar düğümleri (decision nodes) ve yaprak düğümleri (leaf nodes) bulunmaktadır. Karar düğümleri; karar verme, sınıflandırma yapma, tahmin etme gibi işlemleri gerçekleştirirken yaprak düğümleri ise bu kararların tutulduğu yerdir. Ağacın başı kök düğüm (root node) olarak isimlendirilir ve karara bağlanmak için kök düğümden yaprak düğümlere doğru belli bir yol izlenmektedir. [5]



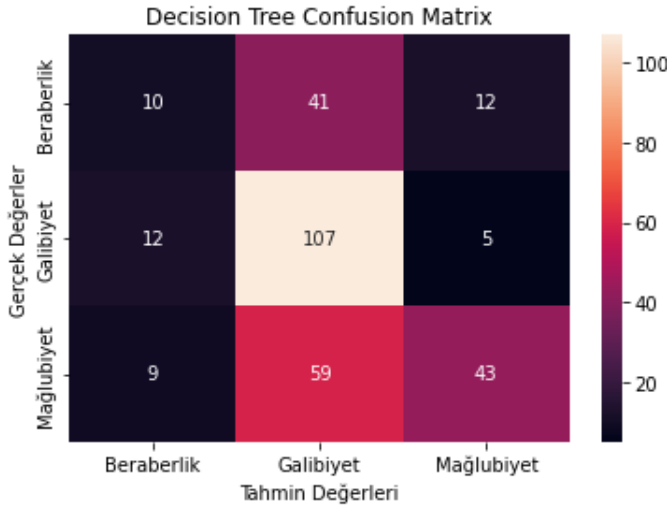
Karar Ağacı (Decision Tree) [6]

İlgili örnek için Karar Ağacı algoritmasında cross validation değeri 10 ve criterion değeri de 'entropy' olarak belirlenmiştir. Aynı zamanda algoritmada max_depth değeri 3-20 aralığında döngü kurularak en yüksek cross validation değeri için otomatik olarak ayarlanmaktadır. Dolayısıyla veri seti max depth değerini içeriğe göre otomatik olarak ayarlamaktadır. İlgili örnek için cross validation değerinin en yüksek değeri max_depth= 3 için belirlenmiştir. Max_depth= 3 için elde edilen Accuracy değeri 0.536, Precision değeri 0.550, Recall değeri 0.536 ve F1 değeri de 0.501 olarak bulunmuştur. Elde edilen karar ağacı modeli DT - 1'de gösterilmiştir.



DT - 1

Elde edilen sonuçlarla oluşturulan Confusion Matrix'te (DT - 2) modelin beraberlik ile sonuçlanan 10 karşılaşmayı (TP), ev sahibi takımın galibiyeti ile sonuçlanan 12 karşılaşmayı (FP) ve ev sahibi takımın mağlubiyeti ile sonuçlanan 9 karşılaşmayı (FP) 'Beraberlik' etiketi ile görülmektedir. Aynı şekilde ev sahibi takımının galibiyeti ile sonuçlanan 40 karşılaşmayı (FP), ev sahibi takımın galibiyeti ile sonuçlanan 106 karşılaşmayı (TP) ve ev sahibi takımın mağlubiyeti ile sonuçlanan 58 karşılaşmayı (FP) 'Galibiyet' etiketi ile etiketlemiştir. Yine aynı şekilde ev sahibi takımın galibiyeti ile sonuçlanan 13 karşılaşmayı (FP), ev sahibi takımın galibiyeti ile sonuçlanan 6 karşılaşmayı (FP) ve ev sahibi takımın mağlubiyeti ile sonuçlandığı 44 karşılaşmayı (TP) 'Mağlubiyet' olarak etiketlemiştir. Bu sonuçlar ışığında ilgili modelin ev sahibi takım özelinde galibiyet ile sonuçlanan karşılaşmaları 'Galibiyet' etiketi ile etiketlemesi konusunda oldukça başarılı, ev sahibi takımın mağlubiyeti ile sonuçlanan karşılaşmaların 'Mağlubiyet' ile etiketlenmesi konusunda başarılı ancak beraberlik ile sonuçlanan karşılaşmaların 'Beraberlik' ile etiketlenmesi konusunda başarısız olduğu gözlemlenmektedir.



DT - 2

D. Öznitelik Seçimi (Feature Selection)

Bir müsabakanın galibiyet, beraberlik ya da mağlubiyet ile sonuçlanmasını etkileyen birçok özellik (feature) vardır ve bu özelliklerin doğru seçilmesi oldukça önemlidir. İlgili çalışmada "Ev Sahibi Takım Adı", "Deplasman Takım Adı", "Stadyum", "Ev Sahibi Takım Piyasa Değeri", "Deplasman Takım Piyasa Değeri", "Seyirci Sayısı" olmak üzere 6 öznitelik seçilmiştir. (Tablo 1) Bu öznitelikler neticesinde karşılaşmanın ev sahibi takım özelinde "Galibiyet", "Beraberlik" ya da "Mağlubiyet" ile sonuçlanma durumu (Tablo 2) tahmin edilmiştir.

III. SONUÇLAR VE TARTIŞMA

Bir futbol karşılaşması her ne kadar galibiyet, beraberlik ve mağlubiyet olmak üzere 3 farklı sonuçla bitiyor olsa da bu sonuçları etkileyen birçok etken bulunmaktadır. Bu çalışmada da ev sahibi takımın adı, deplasman takımının adı, maçın oynandığı stadyum, maçı izleyen seyirci sayısı, ev sahibi takımın piyasa değeri ve deplasman takımının piyasa değeri olmak üzere 6 öznitelikten oluşan 744 örnek incelenmiştir. Bu örneklerin %60'ı (446) eğitim veri seti, %40'ı (298) ise test veri seti olarak ayrılmıştır. Eğitim seti için KNN, Naive Bayes, Lojistik Regresyon ve Karar Ağacı algoritmaları ile model oluşturulmuş ve her model için eğitim veri seti eğitilmiştir. Buna bağlı olarak da her model için Accuracy, Precision, Recall ve F1 skorları hesaplanmış ve Confusion Matrix'leri çıkarılmıştır.

A. Sonuçların İncelenmesi

İlgili veri seti ile geliştirilen KNN, Naive Bayes, Logistic Regression ve Decision Tree algoritmaları ile ilgili elde edilen TP, FP, Accuracy, Precision, Recall ve F1 skorları Tablo S1 ve Tablo S2'de gösterilmektedir.

	TP	FP	F1 Score
KNN	176	122	0.5440
Naive Bayes	174	124	0.5476
Logistic Regression	176	122	0.5190
Decision Tree	160	138	0.5023

Tablo S1

	Accuracy	Precision	Recall
KNN	0.5906	0.5555	0.5906
Naive Bayes	0.5838	0.5472	0.5838
Lojistik Regresyon	0.5906	0.4647	0.5906
Karar Ağacı	0.5369	0.5445	0.5369

Tablo S2

Tüm bu sonuçlar ışığında F1 skorlarının Precision ve Recall değerlerinin harmonik ortalamasını verdiği de düşünüldüğünde en iyi algoritma seçiminde F1 skor değerinin karar vericiliğinin en yüksek olduğu düşünülebilir. Elde edilen sonuçlar incelendiğinde F1 skorları birbirine çok yakın değerler olsa da en iyi skoru 0.5476 değeriyle Naive Bayes algoritmasına ait olduğu görülmektedir. Dolayısıyla ilgili çalışmada denenen algoritmalar arasında en iyi sonucu veren algoritmanın Naive Bayes algoritması olduğu çıkarılabilir.

IV. TARTIŞMA

İlgili çalışmada her ne kadar en iyi algoritmanın Naive Bayes algoritması olduğu sonucu çıkarılsa da 3 etiketli bir problem için 0.5476 olan F1 skoru çok yeterli görülmektedir. Dolayısıyla ilgili veri setinde öznitelikler yeniden gözden geçirilebilir ya da yeni öznitelikler eklenebilir. Aynı şekilde ilgili veri seti mevcut haliyle ve/veya düzenlenmiş özniteliklerle birlikte farklı algoritmalar ile denenebilir.

KAYNAKLAR

- [1] Ü. TAZEGÜL, "Farklı toplumlarda futbolun tarihi," *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi*, no. 64, pp. 178–187, 2017.
- [2] Y. Özkan and V. M. Yöntemleri, "Papatya yayıncılık eğitim," *İstanbul, Mayıs-2008*, 2013.
- [3] M. Aydoğan and A. Karcı, "Meslek yüksekokulu öğrencilerinin başarı performanslarının makine öğrenmesi yöntemleri ile analizi," in *International Symposium on Multidisciplinary Studies and Innovative Technologies*, 2018.
- [4] I. H. M. Paris, L. S. Affendey, and N. Mustapha, "Improving academic performance prediction using voting technique in data mining," *International Journal of Computer and Information Engineering*, vol. 4, no. 2, pp. 306–309, 2010.
- [5] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [6] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, 2021.