

Makine Öğrenmesi Teknikleriyle Potansiyel Yetenek Avı

Merve Mirza Parıldar^{1*} ve Cemil Zalluhoğlu²

¹Veri ve Bilgi Mühendisliği / Bilişim Enstitü, Hacettepe Üniversitesi, Türkiye

²Bilgisayar Mühendisliği / Hacettepe Üniversitesi, Türkiye

*(mervemirzaparildar@gmail.com)

(Received: 03 November 2023, Accepted: 13 November 2023)

(2nd International Conference on Contemporary Academic Research ICCAR 2023, November 4-5, 2023)

ATIF/REFERENCE: Parıldar, M. M. & Zalluhoğlu, C. (2023). Makine Öğrenmesi Teknikleriyle Potansiyel Yetenek Avı. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(10), 240-245.

Özet – Bu çalışma, futbolculardaki potansiyel yeteneklerini tahmin etmek için makine öğrenimi tekniklerini araştırmaktadır. İki veri kümesi kullanılarak futbolcuların istatistiksel özellikleri analiz edilmiş ve çeşitli makine öğrenimi algoritmaları ile potansiyel etiketler tahmin edilmiştir. Seçilen birden fazla makine öğrenmesi modellerinden en başarılı sonuç Rastgele Orman yaklaşımından elde edilmiştir. Sonuçlar, makine öğrenimi tekniklerinin futbolculardaki potansiyeli tahmin etmede etkili olduğunu göstermektedir. Bu yaklaşım, futbol kulüplerine, veri odaklı oyuncu seçimi yapabilme, potansiyel yetenek gösteren oyuncuları daha doğru bir şekilde tanımlayarak gelecekteki başarılarını artırma fırsatı sunmaktadır.

Potansiyel Yetenek, Makine Öğrenmesi, Veri Analizi, Sınıflandırma Modelleri, Özellik Seçimi

I. GİRİŞ

Futbol, karmaşık bir spor dalıdır ve birçok faktörün birleşimiyle oyuncuların performansı ve potansiyeli belirlenir. Geleneksel yöntemlerle futbolculardaki potansiyeli tahmin etmek, belirsizlikler ve yanılgılarla dolu olabilir. Makine öğrenmesi algoritmaları, futbolculardaki performansı etkileyen birçok faktörü dikkate alarak, istatistiksel analizler ve örüntü tanıma yöntemleri kullanarak oyuncuları sınıflandırabilir. Örneğin, bir oyuncunun hızı, pas yüzdesi, şut becerisi, top sürme yeteneği gibi özellikleri, potansiyel yeteneklerini belirlemede önemli rol oynar. Makine öğrenmesi algoritmaları, bu tür özelliklerin karmaşıklığını ve ilişkilerini modelleyerek, futbolculardaki potansiyel yetenekleri tahmin etmek için bir öngörü yapabilir. Temelde çalışmada kullanılan algoritma, futbolculardaki performans verilerini kullanarak potansiyel yetenekleri tahmin etmedeki başarısını analiz etmeyi amaçlamaktadır.

II. MATERYAL VE YÖNTEM

A. Veri Seti

Geliştirilecek olan algoritma için kullanılan veri seti, maçlarda gözlemlenen futbolcuların özelliklerine göre scoutların değerlendirdikleri futbolcuların, maç içerisinde puanlanan özellikleri ve puanlarını içeren bilgileri tutan 2 ayrı veri dosyasından oluşmaktadır [1]. Bir scoutun bir maçta bir oyuncuyla ilgili özellik değerlendirmelerini içeren veri seti Tablo 1'de gösterilmiştir. Bu veri seti oyuncunun pozisyonu, değerlendirildiği özellik, her bir özelliği için aldığı puan gibi değişkenlerin bulunduğu 8 değişkenden ve 10.730 gözlemden oluşmaktadır. Bir scoutun bir maçta bir takımın kadrosundaki tüm oyunculara dair değerlendirmelerini içeren veri seti ise Tablo 2'de gösterilmiştir. Bu veri seti ise oyuncuyla ilgili nihai kararı belirten hedef değişkeninin (etiket) de bulunduğu 5 değişkenden ve 322 gözlemden oluşmaktadır.

Veri setinde oyuncuların değerlendirildiği her bir özelliğin id'si için bu özelliklerin ne tip özellikler olduğu bilgisi bulunmamaktadır. Fakat bu konu ile ilgili yapılan literatür araştırmasına göre bu özellik id'leri için muhtemel anlamların hız, dayanıklılık, güç, pas, şut, top sürme, savunma, beceri, top kontrolü, pas hızı, çalım, oyun zekâsı, ikili mücadele, hava topu, kafa vuruşu, taktik, baskı yapma, serbest vuruş, penaltı vuruşu gibi özellikler olduğu yorumu çıkarılabilir.

Tablo 1. scoutium_attributes veri dosyası

task_response_id	Bir scoutun bir maçta bir takımın kadrosundaki tüm oyunculara dair değerlendirmelerinin kümesi
match_id	İlgili maçın id'si
evaluator_id	Değerlendiricinin (scout'un) id'si
player_id	İlgili oyuncunun id'si
position_id	İlgili oyuncunun o maçta oynadığı pozisyonun id'si 1: Kaleci, 2: Stoper, 3: Sağ bek, 4: Sol bek, 5: Defansif orta saha, 6: Merkez orta saha, 7: Sağ kanat, 8: Sol kanat, 9: Ofansif orta saha, 10: Forvet
analysis_id	Bir scoutun bir maçta bir oyuncuya dair özellik değerlendirmelerini içeren küme
attribute_id	Oyuncuların değerlendirildiği her bir özelliğin id'si
attribute_value	Bir scoutun bir oyuncunun bir özelliğine verdiği değer(puan)

Tablo 2. scoutium_potential_labels veri dosyası

task_response_id	Bir scoutun bir maçta bir takımın kadrosundaki tüm oyunculara dair değerlendirmelerinin kümesi
match_id	İlgili maçın id'si
evaluator_id	Değerlendiricinin(scout'un) id'si
player_id	İlgili oyuncunun id'si
potential_label	Bir scoutun bir maçta bir oyuncuyla ilgili nihai kararını belirten etiket. (hedef değişken)

B. Veri Ön İşleme Çalışmaları

İlk olarak veri setinin tamamlayıcı bilgilerini içeren verileri bir araya getirmek amacıyla iki ayrı veri dosyası birleştirilmiştir. Birleştirme işlemi sonucunda elde edilen veri seti, her bir futbolcunun özellik değerlendirmelerini ve potansiyel değerlendirmelerini içeren geniş kapsamlı bir veri tabanını temsil etmektedir.

Değişkenler arasındaki ilişkilerin incelenmesi amacıyla korelasyon matrisi oluşturulmuş olup, bir takımın oyuncularının bir maçtaki ortalama puanının, o takımın oyuncularının puanlarının ortalamasına yakın olduğu görülmüştür. Takım oyuncularının bir maçtaki performanslarının genellikle takımın genel performansını yansıtmaması ise beklenen bir durumdur.

Veri analizi veya modelleme üzerinde olası yanılsamaları önlemek için pozisyonu kaleci olan oyuncular veri setinden çıkarılmıştır. Kalecilerin farklı özellikleri, genel eğilimleri veya ilişkileri yanıltıcı bir şekilde etkileyebilir. Bu çalışmanın amaçladığı futbolcuların performansını tahmin etmeye yönelik modelleme için kalecilerin çıkarılması, diğer futbolcuların performans tahminlerini daha doğru hale getirecektir. Bu ön işleme adımıyla 700 kalecinin verisi veri setinden çıkarılmıştır.

Veri setindeki sınıf etiketlerinin dengeli bir şekilde dağılıp dağılmadığı incelenmiş ve "potansiyeli düşük" sınıfının dağılımı çok düşük olduğu için bu sınıfa ait veriler veri setinden çıkarılmıştır. Geriye "potansiyeli ortalama" ve "potansiyeli yüksek" olmak üzere 2 sınıf kalmıştır. Daha sonra futbolcuların pozisyonlarına ve potansiyel etiketlerine göre özelliklerin dağılımını daha iyi gösterebilmek ve veri setinin daha anlaşılır olmasını sağlamak amacıyla bir pivot tablosu oluşturulmuştur.

Modelin dağılımı fazla olan sınıfı daha iyi öğrenmesi ve yanlış sınıflandırması durumunun önüne geçmek amacıyla SMOTE yöntemi kullanılarak sınıf dengesizliği giderilmiştir.

Veri setindeki farklı ölçeklerdeki değişkenlerin etkisini eşitleyerek daha tutarlı bir analiz yapılabilmesi için sayısal değerleri standartlaştırma adımı gerçekleştirilmiştir.

Son olarak bağımlı ve bağımsız değişkenler belirlenmiş, veri setinin %80'lik kısmı makine öğrenmesi algoritmalarını eğitme, kalan %20'lik kısmı ise bu modellerin başarımını ölçmede kullanılmak üzere ayrılmıştır.

C. Modelleme

1. Değerlendirme Metrikleri

Potansiyel yetenek avı tahminlemesi için gözetimli öğrenme içerisinde yer alan 6 sınıflandırma modeli tercih edilmiştir. Algoritmanın hedef bir nitelik ile ilgili tahminleri ile gerçek değerlerin karşılaştırıldığı karmaşıklık matrisi; sınıflandırma modelinin doğruluk, hassasiyet, duyarlılık metriklerini $p = \frac{e^{\alpha+bx}}{1 + e^{\alpha+bx}}$ gibi performans hesaplamak için temel

oluşturmaktadır. Bu matris, modelin sınıflandırma sonuçlarını gerçek sınıflarla karşılaştırarak farklı sonuçları kategorize eder.

Çalışmada kullanılan modellerin değerlendirilmesi aşağıdaki kriterlere göre yapılmıştır.

Doğruluk (Accuracy): $TP+TN / TP + TN + FP + FN$ eşitliği ile hesaplanır. Modelin genel performansını ortaya koyar. Bu sınıflandırma metriği ile doğru tahminlerin toplam tahminler içindeki oranı hesaplanmış olur [1].

Kesinlik (Precision): $TP / TP+FP$ eşitliği ile hesaplanır. Modelin, pozitif olarak tahminlediği örneklerin ne kadarının gerçekten doğru olduğunu gösterir [2].

Geri çağırma (Recall): $TP / TP+FN$ eşitliği ile hesaplanır. Gerçek pozitif örneklerin oranını verir.

F1 skoru: $2TP / 2TP+FP+FN$ eşitliği ile hesaplanır. Dengesiz sınıflar için yararlı hibrit bir metrik olup, kesinlik (precision) ve geri çağırma (recall) skorlarının ağırlıklı ortalamasını ifade etmektedir [3].

ROC-AUC (Area Under the Receiver Operating Characteristic Curve): Modelin çıktıları ve gerçek sınıf etiketleri kullanılarak hesaplanan TPR (True Positive Rate- Gerçek Pozitif Oran) ve FPR (False Positive Rate- Yanlış Pozitif Oran) değerleri üzerinden hesaplanır. ROC-AUC, TPR ve FPR değerlerinin birbirleriyle olan ilişkisini temsil eden ROC eğrisinin altında kalan alanı ifade eder [2]. ROC-AUC değeri, 0 ile 1 arasında bir değer alır. 1'e ne kadar yakınsa, modelin sınıflandırma performansı o kadar iyidir. Öte yandan, ROC-AUC değeri 0.5'e yakınsa, modelin performansı rastgele tahmin yapmaktan farksızdır.

2. Kullanılan Makine Öğrenmesi Algoritmaları

a) Logistic Regression (Lojistik Regresyon) Modeli

Lojistik regresyon, ikili (binary) değere sahip olabilen bir sonuc olasılığını tahmin edici bir

analiz yöntemidir. Bu yöntem, veri setindeki bir veya birkaç sayısal veya kategorik bağımsız değişkenin kullanımına dayanarak bir olayın gerçekleşme olasılığını tahmin etmek için kullanılır. Lojistik regresyon, 0 ile 1 arasındaki değerlerle sınırlı olan lojistik eğrisi üretir [4].

Bir lojistik regresyon modeli, bir $ax+b$ denklemi aracılığıyla olayın gerçekleşme olasılığını tahmin etmek için katsayıları üretir. Bir $ax+b$ denklemi için bir olayın gerçekleşme olasılığı:

Bu tahmin işlemi, logit dönüşümü kullanılarak gerçekleştirilir. Logit fonksiyonu (p) olayın gerçekleşme olasılığını 0 ile 1 arasında sınırlayan bir dönüşüm sağlar [5]. Logit fonksiyonu, olasılığın doğal logaritması olan $(p/(1-p))$ olarak ifade edilir [6].

Logit fonksiyonu:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

Bu çalışmada model, futbolcuların özelliklerini temsil eden bağımsız değişkenleri kullanarak oyuncuların hangi sınıfa (average - ortalama veya highlighted - öne çıkan) ait olduğunu tahmin etmeyi hedeflemektedir. Bu tahminleme sürecinde, veri setindeki "potential_label" sütunu hedef değişken olarak kullanılmış ve bu sütunun tahmin edilmesiyle futbolcuların potansiyel değerlendirmelerinin yapılabilmesi amaçlanmıştır.

b) K-Neighbours (K-En Yakın Komşu) Modeli

K-NN algoritmasının temel prensibi, tahmin edilecek değer için bağımsız değişkenlerinin oluşturduğu vektörün en yakın komşularının hangi sınıfta yoğun olduğu bilgisini kullanarak sınıfını tahmin etmektir. Bu bilgiye dayanarak, yeni bir veri noktasının sınıf etiketini tahminlemek için en yakın komşularına bakılır ve kendisine en yakın komşuların olduğu sınıfa dahil edilir.

Bu çalışmada K-NN modeli, futbolcuların özelliklerine dayalı olarak benzerlik ölçüsüne göre en yakın komşuları belirler ve bu komşuların sınıfları üzerinden çoğunluk kararını alarak tahminleme yapar.

c) Decision Tree (Karar Ağacı Sınıflandırıcısı) Modeli

Bu algoritma, özelliklerin değerlerine dayanarak bir karar ağacı oluşturur ve sınıflandırma işlemini gerçekleştirir. Karar Ağacı Sınıflandırıcısı, veri setindeki özelliklerin değerlerine göre bir kök düğüm oluşturur ve bu düğümü alt düğümlere bölerek ağacı inşa eder. Her düğüm, bir özellik değeriyle ilişkilendirilir. Bölünme işlemi, homojenlik veya belirli bir sınıfa hedefleme gibi belirli bir kriteri sağlamak amacıyla gerçekleştirilir. Dallar ve yapraklar karar ağacı yapısının diğer elemanlarını oluşturur. Ağaçta başlangıç kısmı kök en son kısım yaprak olarak adlandırılır. Bu iki yapı arasında kalan kısımlar ise dal olarak ifade edilir [7]. Bu şekilde ağaç, sınıflandırma kararları için özellikler arasındaki ilişkileri kullanır. Tahminleme, oluşturulan karar ağacı üzerinden gerçekleştirilir. Bir veri noktası ağaç yapısında ilerler ve son düğüme ulaştığında, bu düğümün sınıf etiketiyle sınıflandırılır. Sınıflandırma işlemi, veri noktalarının karar ağacı üzerindeki yolculuğu ve sınıf etiketiyle eşleştirilmesiyle gerçekleştirilir.

d) RandomForest (Rastgele Orman Sınıflandırıcısı) Modeli

Rastgele Orman Sınıflandırıcısı, birçok karar ağacının bir araya gelerek oluşturduğu bir denetimli ensemble (topluluk) sınıflandırma modelidir. Temel sınıflandırıcı olarak karar ağacını kullanır [5]. Leo Breiman tarafından keşfedilen bu yöntem, temeli birden çok karar ağacının ürettiği tahminlerin bir araya getirilerek değerlendirilmesine dayanmaktadır [8].

Random Forest algoritması, her bir ağaç için rastsal gözlem ve değişken seçimini kullanır. Gözlem seçimi, bootstrap rastgele örnekleme yöntemiyle gerçekleştirilir. Bu yöntem, veri setinden rastgele gözlemler seçerek ağaçların oluşturulmasında tekrarlamaya izin verir. Böylece, her ağaç farklı bir alt küme üzerinde eğitilir ve farklı bir perspektif sunar.

Değişken seçimi ise random subspace yöntemiyle gerçekleştirilir. Bu yöntem, tüm değişkenlerin bulunduğu bir kümeden rastgele bir alt küme seçer ve sadece bu alt kümedeki değişkenler kullanılarak ağaçlar oluşturulur. Bu, her ağacın farklı özelliklerle eğitilmesini sağlar ve modelin daha genelleyici olmasına yardımcı olur. Sonuçta tüm bireysel ağaçları birleştirerek rastgele orman modeli oluşturulur [9].

Her bir ağaç, seçilen gözlem ve değişkenler üzerinde dallanma işlemleri yaparak oluşturulur.

Sonuç olarak, Random Forest birden çok karar ağacının tahminlerini bir araya getirerek ensemble tahminini üretir. Random Forest, gözlem ve değişken seçiminde rastsallığı koruyarak farklı örüntüleri modellemeyi mümkün kılar. Bu, veri setinin yapısında gizli olan kompleks ilişkileri daha iyi yakalayabilmesini sağlar, aşırı uyumu (overfitting) önler, yanlılık-bağımsızlık (bias-variance) dengesini iyileştirir ve böylece tahmin başarısını artırır.

e) GradientBoosting (Gradyan Artırma Sınıflandırıcısı) Modeli

Gradyan Artırma Sınıflandırıcısı, birçok zayıf öğrencinin (weak learner) bir araya gelerek güçlü bir sınıflandırma modeli oluşturduğu bir ensemble yöntemidir. Hataları azaltarak ve zayıf öğrencilerin hatalarını düzelterek adım adım güçlü bir tahmin modeli oluşturur.

Gradyan artırma algoritması, bir hedef fonksiyonu minimize etmek için gradyan inişini kullanır. İlk olarak, bir zayıf öğrenci, veri setini sınıflandırmaya çalışır ve hataları hesaplanır. Ardından, hataların gradyanı, bir sonraki zayıf öğrencinin eğitimi için kullanılır. Bu, hataları azaltmaya yönelik bir düzeltme faktörü olarak kullanılır ve yeni öğrenci, önceki öğrencinin zayıflıklarını düzeltmeye çalışır. Yani sonraki öğrenciler önceki öğrencilere bağımlıdır.

Gradyan artırma, bir modelin performansını ölçmek için bir kayıp (loss) fonksiyonunu kullanır. Bu kayıp fonksiyonu, modelin tahminlerinin gerçek değerlerden ne kadar sapma gösterdiğini ölçer. Gradyan yükseltme algoritması, zayıf öğrenciler olarak adlandırılan basit modellerden oluşur. Her bir zayıf öğrenci tek başına iyi tahminler yapamaz, ancak bir araya geldiklerinde güçlü bir tahmin modeli oluştururlar. Gradyan yükseltme, her bir zayıf öğrenciye katkıda bulunacak bir model ekleyerek kayıp fonksiyonunu en aza indirmeyi hedefler. Bu katkı modeli, kayıp fonksiyonunu en aza indirecek şekilde öğrenilen gradyan bilgisini kullanır [10].

f) Support Vector (Destek Vektör Sınıflandırıcısı) Modeli

Destek Vektör Sınıflandırıcısı'nın temel fikri, verileri bir ayrıştırma hiper düzlemiyle en iyi şekilde ayırmaktır. Bu hiper düzlem, sınıflar arasındaki maksimum marjinal mesafeyi maksimize edecek şekilde optimize edilir. Marjinal mesafe, iki

sınıfa ait en yakın veri noktaları arasındaki uzaklığı ifade eder. Sınır düzlemlerinin tam ortasından geçer ve her iki düzleme de eşit uzaklıktadır [11].

Destek vektörler, bu marjinal mesafeyi belirleyen veri noktalarıdır. Bu yöntem, verileri yüksek boyutlu bir uzaya dönüştürerek sınıfları lineer

Tablo 3- Modeller ve Sonuçları

Model	Doğruluk	Kesinlik	Geri Çağırma	F1-Skor	ROC-AUC
Lojistik Regresyon	0.81	0.86	0.78	0.82	0.84
K-NN	0.84	0.9	0.81	0.85	0.90
Karar Ağaçları	0.87	0.88	0.86	0.87	0.87
Rastgele Orman	0.94	0.93	0.95	0.94	0.98
Gradyan Artırma	0.91	0.93	0.90	0.91	0.96
Destek Vektör	0.86	0.83	0.87	0.85	0.93

olarak ayrılabilir hale getirir. Çekirdek fonksiyonlarıyla bu dönüşüm gerçekleştirilir, böylece karmaşık ve non-lineer veri kümelerini sınıflandırmak mümkün hale gelir.

Destek Vektör Sınıflandırıcısı, ayrıştırma hiper düzlemi üzerindeki destek vektörlerini kullanarak yeni veri noktalarını sınıflandırır. Sınıflandırma işlemi, yeni veri noktasının hangi tarafta (hangi sınıfta) yer aldığını belirlemek için destek vektörlerden yararlanır. Bu sayede, veri noktaları sınıflara doğru şekilde ayrılabilir. Bulgular açık ve öz olmalıdır. Bulguların en önemli özellikleri ve eğilimleri açıklanmalı, ancak ayrıntılı olarak yorumlanmamalıdır.

3. Özellik Değerlendirmesi

Future importance yani özellik değerlendirme, makine öğrenimi ve veri madenciliği alanlarında yaygın olarak kullanılan bir yöntemdir. Bu yöntem, bir veri kümesinde yer alan özniteliklerin veya değişkenlerin hedef değişken üzerindeki etkisini ölçmeye ve önem düzeylerini belirlemeye odaklanır. Özellik değerlendirme, özellikle öznitelik seçimi, boyut azaltma veya veri ön işleme gibi aşamalarda önemli bir rol oynamaktadır. Modelin performansını artırmak, anlaşılabilirliği sağlamak ve gereksiz bilgi girişlerini azaltmak amacıyla önemli bir adımdır.

Bu çalışma, öznitelikler rakam olarak tutulmuş olsa da en iyi sonucu veren Rastgele Orman ve Gradyan Artırma modellerindeki özniteliklerin önemini değerlendirmek amacıyla özellik değerlendirme yöntemlerini de kullanmıştır. Her bir özneliğin hedef değişken üzerindeki etkisi ölçülmüştür.

Özellik değerlendirme sonuçlarına göre, her iki modelde de "4325" olarak kodlanan öznitelik en

yüksek öneme sahip olarak belirlenmiştir. Ancak, veri setinde özniteliklerin hangi özellikleri temsil ettiği belirtilmediğinden bu özneliğin anlamı

hakkında kesin bir yorum yapılamamaktadır. Bu özneliğin şut yeteneği veya top sürme gibi ofansif beceriler ya da savunma veya hava topu gibi defansif beceriler olabileceği düşünülebilir, ancak kesin bir yorum yapabilmek için özneliğin neyi temsil ettiği bilgisine ihtiyaç vardır.

Öte yandan, ilginç bir bulgu olarak, pozisyon özneliği (position_id) beklenildiği gibi yüksek bir öneme sahip çıkmamıştır. Bu sonuç, futbolcuların pozisyonlarının tahmin performansı üzerindeki etkisinin sınırlı olduğunu göstermektedir. Bu durum, futbolculardaki potansiyel yetenekleri tahmin etmek için pozisyon bilgisinin tek başına yeterli olmayabileceğini düşündürülebilir.

III. SONUÇLAR

Bu çalışmanın temel amacı, scoutlar tarafından izlenen futbolcuların özelliklerine dayanarak oyuncuların hangi sınıfa (ortalama, öne çıkan) ait olduğunu tahmin etmek ve futbolculardaki potansiyel yetenekleri tahmin etmek için makine öğrenmesi tekniklerini araştırmaktır.

Çalışmada, farklı makine öğrenimi algoritmaları kullanılarak yapılan deneylerin sonuçları Tablo 3'te gösterilmiş olup, elde edilen metriklere dayanarak bir değerlendirme yapılabilir.

Random Forest sınıflandırıcı modeli, futbolcuların potansiyel yeteneklerini tahmin etmede en yüksek başarıyı sergilemiştir. Recall değeri 0.95 olarak hesaplanmıştır, bu da modelin öne çıkan sınıfa ait olan futbolcuların %95'ini doğru bir şekilde sınıflandırabildiğini gösterir. Precision (kesinlik) değeri ise 0.93'tür, yani modelin öne çıkan sınıfa ait olarak tahmin ettiği futbolcuların %93'ünün

gerçekten öne çıkan sınıfa ait olduğunu gösterir. F1 skoru modelin öne çıkan sınıfa ait futbolcuları %94 oranında doğru bir şekilde sınıflandırdığını ve tahminlerinin tutarlı olduğunu göstermektedir. ROC-AUC değeri ise 0.98'dir, bu da modelin yüksek bir ayırım gücüne sahip olduğunu ve iyi bir performans sergilediğini göstermektedir.

Accuracy (doğruluk) değeri ise modelin, futbolcuların potansiyel yeteneklerini %94 doğruluk oranında tahmin edebildiğini gösterir. Yüksek doğruluk oranı, modelin genel performansının tatmin edici olduğunu göstermektedir.

Bu sonuçlar, futbol kulüplerinin potansiyel yetenek gösteren oyuncuları daha etkili bir şekilde belirlemeleri ve oyuncu seçiminde daha bilinçli kararlar almaları konusunda önemli bir adım sağlamaktadır. Makine öğrenmesi tekniklerinin kullanılması, veri analitiği ve sınıflandırma modellerinin futbolculardaki performansı değerlendirmede kullanılmasına yönelik bir yol haritası sunmaktadır.

Ancak, bu çalışma hala önemli zorluklarla karşılaşmaktadır. Veri setindeki öznitelikler birer rakamla tutulmaktadır. Ancak, bu rakamların hangi özellikleri temsil ettiği bilinmemektedir. Özniteliklerin açıklaması veya hangi özellikleri temsil ettiği bilindiğinde, daha anlamlı yorumlar yapmak mümkün olacaktır. Örneğin, hız, şut gücü, pas yeteneği gibi özelliklerin model performansına nasıl etki ettiği değerlendirilebilir. Buna ek olarak, daha geniş bir veri setinin kullanılması ve farklı özelliklerin incelenmesi gibi ileri çalışmalar, bu alanda yapılabilecek geliştirmeleri temsil etmektedir.

KAYNAKLAR

- [1] Enes Arslan, *Scoutium Player Potential Prediction*, <https://www.kaggle.com/code/enesarслан8/scoutium-player-potential-prediction-90-acc>, (Erişim Tarihi: 27 Ekim, 2023).
- [2] Sun, Y., Wong, A. K., & Kamel, M. S. (2009). *Classification of Imbalanced Data: A Review*. International Journal of Pattern Recognition and Artificial Intelligence, 23(04), 687–719.
- [3] [AKGÜN, K., & BULUT ÖZEK, M. (2020). *Eğitsel Veri Madenciliği Yöntemi ile ilgili Yapılmış çalışmaların incelenmesi: İçerik analizi*. Uluslararası Eğitim Bilim ve Teknoloji Dergisi, 6(3), 197–213.
- [4] D. G. Kleinbaum, ve M. Klein, *Logistic Regression: A Self-Learning Text* (Statistics for Biology and Health)", Third Edition. New York: Springer 2010.
- [5] He, S., Li, X., & Chen, J. (2021). *Application of data mining in predicting college graduates' employment*. 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD).
- [6] [5] Z. Drus ve H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review", Procedia Computer Science, c. 161, ss. 707–714, 2019, doi: 10.1016/j.procs.2019.11.174.
- [7] Quinlan J.R., 1993, "C4.5: Programs for Machine Learning", Morgan Kaufmann, San Mateo, CA, 302 s.
- [8] <http://www.jcomputers.us/vol7/jcp0712-09.pdf>. (An Improved Random Forest Classifier for Text Categorization
- [9] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001.
- [10] Anonim, *Topluluk Öğrenme Algoritmalarından Gradyan Yükseltme Algoritması İle Göğüs Kanserinin Tahmini Üzerine Bir Vaka Çalışması* <https://tevfikbulut.com/2020/06/27/topluluk-ogrenme-algoritmalarından-gradyan-yukseltme-algoritmasi-ile-gogus-kanserinin-tahmini-uzerine-bir-vaka-calismasi-a-case-study-on-the-prediction-of-breast-cancer-using-gradient-boosting-algori/> (Erişim Tarihi: 24 Haziran 2023).
- [11] Abdalla S., Erdoğan Ş., (2014). *Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi*. Eskişehir osmangazi üniversitesi iktisadi idari bilimler fakültesi dergisi. 9. 175-198.