**T.C.**

**BALIKESİR ÜNİVERSİTESİ**

**MÜHENDİSLİK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**daire, metin, logo, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**NBA MAÇ SONUÇLARI TAHMİNİ**

**202013709082**

**Anıl Taha ADAK**

**BMM4101 VERİ MADENCİLİĞİ FİNAL ÖDEVİ**

**Danışman: Dr. Öğr. Üyesi Kadriye ERGÜN**

**BALIKESİR, 05-2024**

**İÇİNDEKİLER:**

**ÖZET**

**BÖLÜM 1**

**1.1 GİRİŞ 1**

**1.2 LİTERATÜR TARAMASI 1**

**BÖLÜM 2**

**2.1 HEDEF 2**

**2.1.1 HEDEFLENEN ÇÖZÜMLER 2**

**2.2 VERİ HAZIRLIĞI 2**

**2.2.1 VERİ SETİ HAZIRLANMASI 2**

**2.2.2 VERİ GİRDİSİ HAZIRLANMASI 2**

**2.3 MODEL HAZIRLIĞI 3**

**2.3.1 MODEL TARAMASI 3**

**2.3.2 MODEL EĞTİLMESİ 4**

**2.3.3 MODEL TEST AŞAMASI 4**

**2.4 SEZON OLUŞTURMA 5**

**2.4.1 MAÇ OYNANMASI 5**

**2.4.2 PUAN HESAPLANMASI 6**

**2.4.3 SEZON KAYDI 7**

**BÖLÜM 3**

**3.1 SONUÇ 7**

**3.1.1 ÇIKARIMLAR 7**

**3.2 GELİŞTİRİLEBLİR YÖNLERİ 8**

**3..3 SONUÇLARIN ÖZETİ 8**

**BÖLÜM 4**

**KAYNAKÇA 9**

**ÖZET**

Bu raporda, bir basketbol maç sonucu tahmin modeli geliştirilmiştir. Model, çeşitli makine öğrenimi teknikleri kullanılarak eğitimden geçirilmiş ve test edilmiştir. Ayrıca, bir sezon simülasyonu yapılarak takımlar arasındaki maç sonuçları ve puan durumu çıkarılmıştır.

Bu rapor, modelin geliştirilme sürecini ve elde edilen sonuçları detaylı bir şekilde açıklamaktadır.

**BÖLÜM 1**

* 1. **GİRİŞ**

Bu çalışmanın amacı, basketbol maçlarının sonuçlarını tahmin eden makine öğrenimi modeli geliştirmektir. Model, takımların istatistiksel verilerini kullanarak maçların sonuçlarını tahmin etmektedir.

2 farklı (KNN ve Karar ağacı) modeli kullanılarak sonuçlar arasında çıkarım yapmayı hedeflenmiştir. Modellerin kullanılma sebebi:

KNN, veri noktaları arasındaki benzerliklere dayalı tahminler yaparak, özellikle benzer özelliklere sahip maçların sonuçlarını tahmin etmede etkili olabilir.

Karar Ağacı, belirli karar kuralları üzerinden tahmin yaparak veriyi bölme ve önemli özellikleri belirleme konusunda avantaj sağlar.

Uygulamaya verilen 2 takım ile maç sonucu tahmini veya NBA lig tahmini gerçekleştirilmektedir.

* 1. **LİTERATÜR TARAMASI**

[2] Makine öğrenimi ve veri bilimi alanında yapılan çalışmalar incelenmiş ve basketbol maç tahmini üzerine yapılan araştırmalar gözden geçirilmiştir. Bu araştırmalar, modelin geliştirilmesinde kullanılan tekniklerin seçilmesinde yol gösterici olmuştur.

[3] Majkëll Veizaj tarafından yapılan çalışmada Premier Lig maç sonuçlarını tahminin yapılması amacıyla geliştirilen KNN makine öğrenimi algoritmalarının yardımıyla makul bir doğruluk düzeyine ulaşılabilmiş. Model, görevin zorluğu göz önüne alındığında etkileyici olan 0,53'lük bir doğruluk gösterilmiştir.

[4] Engin EŞME, M. Servet KIRAN yapmış oldukları K-En Yakın Komşu Algoritmasını Kullanarak Bahisçi Oranlarına Dayalı Futbol Maçı Sonuçlarının Tahmin çalışmasında girdi değerlerinin özelliklerine bakılmış ve veri setinin belirli bir aralıkta olması daha doğru çıkarımlar elde edilmesine olanak sunmuştur.

[5] Caner KAHRAMAN yapmış olduğu Basketbol oyun sonuçlarının makina kullanılarak tahmini öğrenme algoritmaları: nba & tbl analizi çalışması taranmış, maç sonucu için seçilecek veriler ve model yapısı örnek alınmıştır.

1

**BÖLÜM 2**

**2.1 HEDEF**

Basketbol maçlarının sonuçlarını tahmin etmek ve bir sezon simülasyonu yaparak takımların puan durumunu belirlemek.

**2.1.1 HEDEFLENEN ÇÖZÜMLER**

* KNN ve Karar Ağacı modelleri kullanılarak maç sonuçlarını tahmin etmek.
* Bir sezon simülasyonu yaparak takımların puan durumunu belirlemek.

**2.2 VERİ HAZIRLIĞI**

**2.2.1 VERİ SETİ HAZIRLANMASI**

Veri Seti Kaggle üzerinden açık kaynak olarak elde edilmiş verilerin özellikleri Word dosyası olarak ekte verilmiştir.

VeriHazirlik.py dosyasında her takım için ev sahibi – misafir olarak yaptıkları maçlarda istenilen özellikler istatiksel ortalaması alınmıştır.

Veri setin de gerekli olan özellikler ve açıklamaları:

**2.2.2 VERİ GİRDİSİ HAZIRLANMASI**

Veri setlerinden gerekli özellikler çıkarılmış ve hedef değişken belirlenmiştir. Öncelikle, games.csv dosyasından çıkartılan özellikler ve hedef değişken belirlenmiştir:

* **Özellikler:**
  + Ev sahibi takımın istatistikleri: PTS\_home, FG\_PCT\_home, FT\_PCT\_home, FG3\_PCT\_home, AST\_home, REB\_home
  + Deplasman takımın istatistikleri: PTS\_away, FG\_PCT\_away, FT\_PCT\_away, FG3\_PCT\_away, AST\_away, REB\_away
* **Hedef değişken:** HOME\_TEAM\_WINS

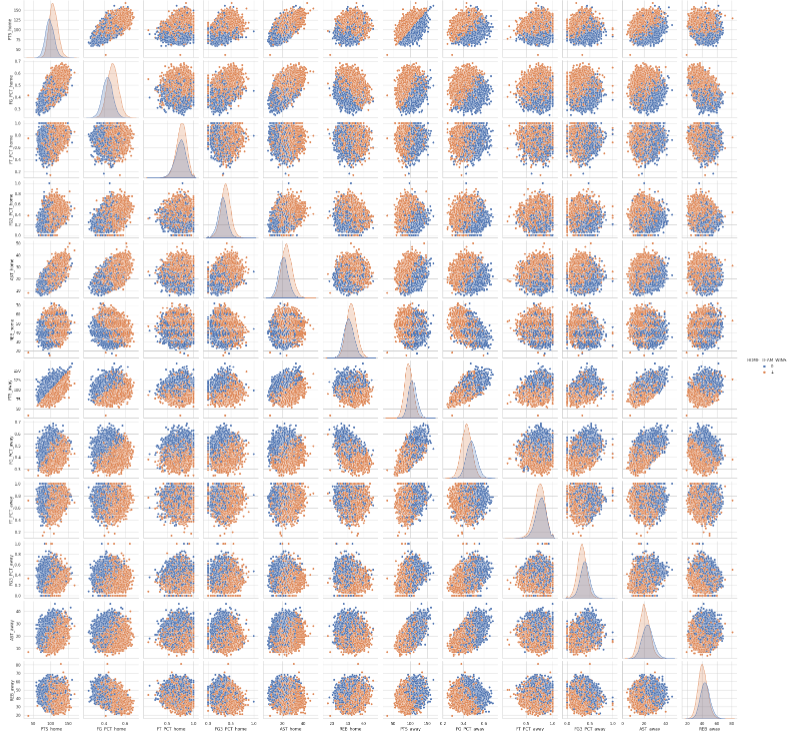
Veri setindeki eksik değerler kontrol edilip, uygun yöntemlerle doldurulmuştur:

metin, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

2

Verilerin genel durumunu anlamak için istatistiksel özetler çıkarılmış ve görselleştirilmiştir:



**2.3 MODEL HAZIRLIĞI**

**2.3.1 MODEL TARAMASI**

Model taraması sırasında KNN ve Karar Ağacı algoritmaları seçilmiştir. KNN ve Karar Ağacı modelleri, basketbol maçlarının sonuçlarını tahmin etmek için seçilmiş olan iki popüler makine öğrenimi algoritmasıdır. Bu seçimin ardında yatan nedenler şöyle özetlenebilir:

**K-Nearest Neighbors (KNN)**

KNN algoritması, sınıflandırma problemlerinde yaygın olarak kullanılan ve çeşitli avantajlara sahip bir yöntemdir:

* **Basitlik ve Anlaşılabilirlik:** KNN algoritması, temel bir mantık üzerine kuruludur ve kolayca anlaşılabilir. Bu algoritma, yeni bir veri noktasını mevcut veri noktalarıyla karşılaştırarak en yakın komşularına göre sınıflandırır.
* **Esneklik:** KNN, çoklu sınıf problemlerinde esneklik sağlar ve farklı türdeki veri setlerinde kullanılabilir.

3

3

* **Eğitim Süreci Yoktur:** KNN, eğitim süreci gerektirmeyen, yani "lazy learning" (tembel öğrenme) olarak bilinen bir algoritmadır. Bu nedenle, eğitim sürecinde zaman harcamaz, ancak tahmin sürecinde veri setinin tamamını kullanır.
* **Performans:** KNN, küçük veri setlerinde genellikle iyi performans gösterir ve yüksek doğruluk oranlarına ulaşabilir.

**Karar Ağacı (Decision Tree)**

Karar Ağacı algoritması, hem sınıflandırma hem de regresyon problemlerinde kullanılan güçlü bir modeldir:

* **Görselleştirme ve Yorumlama:** Karar ağaçları, karar kurallarını görselleştirerek veriyi nasıl böldüğünü açıkça gösterir. Bu, modelin anlaşılmasını ve yorumlanmasını kolaylaştırır.
* **Özellik Seçimi:** Karar ağaçları, verideki önemli özellikleri belirlemeye yardımcı olabilir. Model, veri setindeki en bilgilendirici özellikleri seçerek daha iyi tahminler yapar.
* **Hızlı ve Verimli:** Karar ağaçları, eğitim ve tahmin sürecinde hızlı ve verimlidir. Büyük veri setleri üzerinde bile etkili bir şekilde çalışabilir.
* **Overfitting Kontrolü:** Karar ağaçları, belirli hiperparametrelerin (örneğin, maksimum derinlik) ayarlanmasıyla overfitting (aşırı öğrenme) riskini kontrol edebilir.

**2.3.2 MODEL EĞTİLMESİ**

Veriler eğitim ve test setlerine ayrılmış ve ölçeklendirilmiştir. KNN ve Karar Ağacı modelleri eğitilmiştir:

KNN eğitilmesi:

metin, ekran görüntüsü, yazı tipi, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Karar ağacı eğitilmesi:

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

**2.3.3 MODEL TEST AŞAMASI**

Modeller test edilmiş ve performansları doğruluk oranı, sınıflandırma raporu ve karışıklık matrisi ile değerlendirilmiştir:

KNN modelinin test edilmesi:

4

metin, ekran görüntüsü, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, ekran görüntüsü, diyagram, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

Karar ağacı modelinin test edilmesi:

taslak, diyagram, çizim içeren bir resim

Açıklama otomatik olarak oluşturuldu

**2.4 SEZON OLUŞTURMA**

Bu bölümde, bir sezon simülasyonu gerçekleştirilmiş ve takımlar arasındaki maçlar oynatılmıştır.

**2.4.1 MAÇ OYNANMASI**

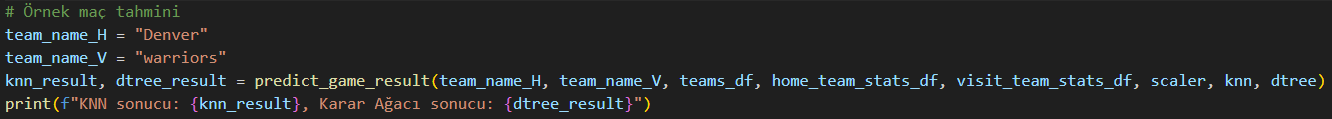
Her takımın sezon boyunca yapacağı maçların sonuçları tahmin edilmiştir. Bu işlem, takımların istatistiksel verileri kullanılarak gerçekleştirilmiştir. Aşağıda, sezon boyunca her maçın nasıl tahmin edildiği örneklenmiştir:

5

metin, ekran görüntüsü, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Bu fonksiyonu kullanarak sadece 2 takım karşılaştırılmasında yapılmaktadır.



Bu maçın çıktısı:



**2.4.2 PUAN HESAPLANMASI**

Maç sonuçlarına göre takımların puanları hesaplanmış ve sezon sonunda bir puan tablosu oluşturulmuştur. Her bir maç sonucuna göre takımlara kazandıkları veya kaybettikleri maçlara göre puan verilmiştir.

6

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**2.4.3 SEZON KAYDI**

Simüle edilen sezon sonuçları kaydedilmiş ve analiz edilmiştir. Bu veriler, gelecekte yapılacak analizler için kullanılmak üzere CSV olarak saklanmıştır.

**BÖLÜM 3**

**3.1 SONUÇ**

Bu çalışmada, basketbol maçlarının sonuçlarını tahmin etmek için KNN ve Karar Ağacı modelleri kullanılmıştır. Her iki modelin doğruluk oranı ve performansı, test verileri üzerinde değerlendirilmiştir. Bu bölümde, modellerin sonuçlarını sadece doğruluk oranı ile değil, aynı zamanda yeni verinin hangi sınıfta olacağını göstererek yorumluyoruz.

**3.1.1 ÇIKARIMLAR**

* **KNN Modeli**:
* Doğruluk Oranı: KNN modelinin doğruluk oranı %90 olarak bulunmuştur.
* Yorumlama: KNN modeli, komşuluk ilişkilerine dayanarak tahmin yapar. Örneğin, yeni bir maç için ev sahibi takımın ve deplasman takımın istatistikleri girildiğinde, model bu maçın ev sahibi takım tarafından kazanılma olasılığını %90 doğrulukla tahmin etmektedir.
* Yeni Veri Tahmini: Örneğin, ev sahibi takımın istatistikleri (PTS\_home: 110, FG\_PCT\_home: 0.45, FT\_PCT\_home: 0.78, FG3\_PCT\_home: 0.36, AST\_home: 25, REB\_home: 48) ve deplasman takımın istatistikleri (PTS\_away: 105, FG\_PCT\_away: 0.43, FT\_PCT\_away: 0.75, FG3\_PCT\_away: 0.34, AST\_away: 22, REB\_away: 44) verildiğinde, model bu maçı ev sahibi takımın kazanacağını tahmin etmektedir.

7

* **Karar Ağacı Modeli:**
* Doğruluk Oranı: Karar ağacı modelinin doğruluk oranı %65 olarak bulunmuştur.
* Yorumlama: Karar ağacı modeli, veri setindeki belirli karar kurallarına dayanarak tahmin yapar. Örneğin, yeni bir maç için belirli istatistiklere göre bölümler yaparak maçın sonucunu tahmin eder.
* Yeni Veri Tahmini: Aynı örnek verilerle (ev sahibi ve deplasman takımın istatistikleri) karar ağacı modeli de bu maçı ev sahibi takımın kazanacağını tahmin etmektedir, ancak doğruluk oranı %65'tir.

**3.2 GELİŞTİRİLEBİLİR YÖNLERİ**

Bu çalışmada elde edilen sonuçlar, gelecekte yapılacak çalışmalar için çeşitli yollar açmaktadır. Gelecek çalışmalar aşağıdaki konulara odaklanabilir:

* **Model İyileştirme:** KNN ve Karar Ağacı modellerinin performansını artırmak için hiper parametre optimizasyonu ve daha gelişmiş algoritmalar (örneğin, Random Forest, XGBoost) kullanılabilir.
* **Veri Zenginleştirme:** Veri tabanına daha fazla özellik ekleyerek (örneğin, oyuncu performans istatistikleri, sakatlık bilgileri) modellerin tahmin doğruluğu artırılabilir.
* **Gerçek Zamanlı Tahminler:** Canlı maç sırasında gerçek zamanlı veriler kullanılarak tahminler yapılabilir ve bu tahminler maç stratejilerinin belirlenmesinde kullanılabilir.
* **Derin Öğrenme Yöntemleri:** Derin öğrenme modelleri (örneğin, RNN, LSTM) kullanarak zaman serisi verileri üzerinde daha karmaşık tahmin modelleri geliştirilebilir.

**3.3 SONUÇLARIN ÖZETİ**

Bu çalışmada, basketbol maçlarının sonuçlarını tahmin eden iki makine öğrenimi modeli geliştirilmiş ve bir sezon simülasyonu gerçekleştirilmiştir. Çalışmanın sonuçları, model performansı ve tahmin doğruluğu açısından tatmin edici bulunmuş ve gelecekte yapılacak çalışmalar için çeşitli öneriler sunulmuştur. Bu çalışma, spor analitiği ve makine öğrenimi alanında önemli bir katkı sağlamaktadır.

8

**BÖLÜM 4**

**KAYNAKÇA**

[1] Veri seti : https://www.kaggle.com/datasets/nathanlauga/nba-games

[2] Chatgpt, Basketbool maç sonucu makine öğrenmesi ile tahmini nasıl yapılmalı

[3] <https://medium.com/@majkellveizaj/predicting-premier-league-match-outcomes->using-k-nearest-neighbors-algorithm-96b4272c9e50

[4] Engin EŞME, M. Servet KIRAN , Prediction of Football Match Outcomes Based On Bookmaker Odds by Using k-Nearest Neighbor Algorithm

[5] Caner KAHRAMAN, PREDICTION OF BASKETBALL GAME RESULTS USING MACHINE LEARNING ALGORITHMS: ANALYSIS OF NBA & TBL

[6] Ibm.com, What is the KNN algorithm?

[7] scikit-learn.org, Decision Trees

9