**Rapor: Basit Bir Tahmin Modeli Oluşturma**

**1. Giriş**

Bu rapor, Iris veri kümesini kullanarak gerçekleştirilen veri temizleme, keşifsel veri analizi (EDA) ve görselleştirme adımlarını ve bu süreçten elde edilen bulguları özetlemektedir. Ayrıca modelleme süreci, değerlendirme sonuçları ve potansiyel iyileştirmeler hakkında bilgiler sunulmaktadır.

**2. Veri Kümesi Hakkında**

Iris veri kümesi, iris çiçeklerinin sepal uzunluğu, sepal genişliği, petal uzunluğu ve petal genişliği gibi özelliklerini içerir. Ayrıca, her bir gözlem için çiçeğin türü ('setosa', 'versicolor', 'virginica') belirtilmiştir. Bu veri kümesi, veri analizi ve makine öğrenimi algoritmalarını test etmek için yaygın olarak kullanılır.

**3. Veri Temizleme**

Veri temizleme aşamasında, eksik veri olup olmadığı kontrol edildi ve tüm değerlerin mevcut olduğu görüldü. Ayrıca veri türleri kontrol edildi ve tüm sütunların doğru veri türlerine sahip olduğu doğrulandı.

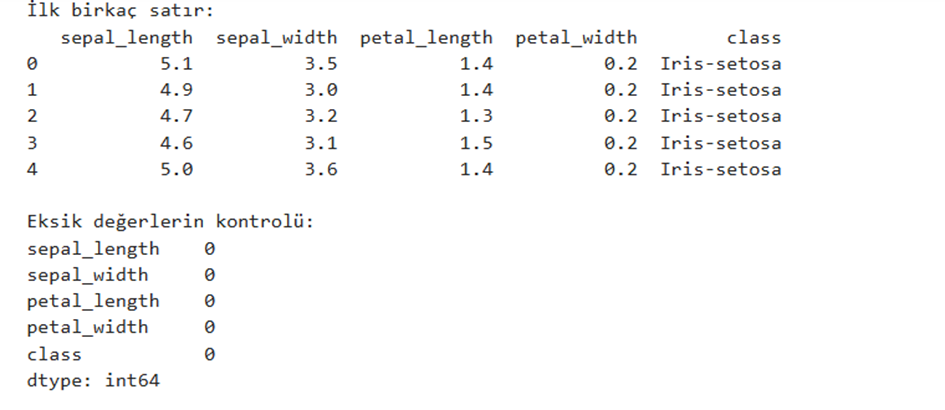
**# Eksik değerlerin kontrolü**

**print(iris.isnull().sum())**

**# Veri türlerinin kontrolü**

**print(iris.dtypes)**

Sonuç:

* Veri kümesinde eksik değer bulunmamaktadır.
* Tüm sütunlar uygun veri türlerine sahiptir.
* 

**4. Keşifsel Veri Analizi (EDA)**

Keşifsel veri analizi sırasında temel istatistikler hesaplandı ve sınıf dağılımı incelendi.

**Temel İstatistikler:**

**print(iris.describe())**

* Sepal ve petal uzunluk ve genişlikleri arasında belirgin farklar bulunmaktadır.
* Her bir özelliğin ortalama ve standart sapma değerleri incelenmiştir.

**Sınıf Dağılımı:**

**print(iris['class'].value\_counts())**

* Üç sınıf da eşit sayıda gözlem içermektedir (her biri 50 gözlem).
* metin, ekran görüntüsü, yazı tipi, menü içeren bir resim

  Açıklama otomatik olarak oluşturuldu

**5. Veri Görselleştirme**

Çiftler arası ilişkileri göstermek için bir pairplot oluşturuldu ve özellikler arasındaki korelasyonlar bir ısı haritası ile gösterildi.

**Pairplot:**

**sns.pairplot(iris, hue='class')**

**plt.show()**

* Farklı sınıfların özellikler arasındaki dağılımlar ve ilişkiler görselleştirildi.metin, diyagram, harita içeren bir resim

  Açıklama otomatik olarak oluşturuldui

**Korelasyon Matrisinin Isı Haritası:**

**numeric\_columns = iris.drop(columns=['class'])**

**corr\_matrix = numeric\_columns.corr()**

**sns.heatmap(corr\_matrix, annot=True, cmap='coolwarm')**

**plt.show()**

* Sepal ve petal uzunluk ve genişlikleri arasındaki korelasyonlar incelendi.
* Özellikle petal uzunluğu ve genişliği arasında yüksek bir korelasyon gözlemlendi.

ekran görüntüsü, kare, dikdörtgen, metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

**6. Modelleme Süreci**

Modelleme sürecinde, sınıflandırma problemini çözmek için bir karar ağacı sınıflandırıcısı kullanıldı. Model, eğitim ve test setlerine bölünmüş veri üzerinde eğitildi ve değerlendirildi.

**Modelin Eğitimi ve Değerlendirilmesi:**

**from sklearn.model\_selection import train\_test\_split**

**from sklearn.tree import DecisionTreeClassifier**

**from sklearn.metrics import accuracy\_score, classification\_report**

**# Veri ve etiketlerin ayrılması**

**X = iris.drop(columns=['class'])**

**y = iris['class']**

**# Eğitim ve test setlerine bölünmesi**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)**

**# Modelin oluşturulması ve eğitilmesi**

**model = DecisionTreeClassifier()**

**model.fit(X\_train, y\_train)**

**# Modelin tahmin yapması ve değerlendirilmesi**

**y\_pred = model.predict(X\_test)**

**print("Accuracy:", accuracy\_score(y\_test, y\_pred))**

**print("Classification Report:\n", classification\_report(y\_test, y\_pred))**

Sonuçlar:

* Model, %97 doğruluk oranı ile sınıflandırma yapmıştır.
* Setosa sınıfı tamamen doğru sınıflandırılmıştır, ancak versicolor ve virginica sınıfları arasında bazı karışıklıklar gözlemlenmiştir.

**7. Potansiyel İyileştirmeler**

* Daha karmaşık modellerin (örneğin, rastgele ormanlar veya destek vektör makineleri) kullanılması performansı artırabilir.
* Özellik mühendisliği yaparak veri kümesindeki özellikleri zenginleştirmek model performansını iyileştirebilir.
* Modelin hiperparametre ayarları (örneğin, karar ağacı derinliği) optimize edilerek daha iyi sonuçlar elde edilebilir.
* Çapraz doğrulama kullanarak modelin genelleme yeteneği artırılabilir.

**8. Sonuç**

Bu çalışma, Iris veri kümesi üzerinde gerçekleştirilen veri temizleme, keşifsel veri analizi ve görselleştirme adımlarını özetlemekte ve modelleme süreci ve değerlendirme sonuçlarını sunmaktadır. Elde edilen bulgular, veri kümesinin genel özelliklerini ve sınıflandırma probleminin çözümünde kullanılan modelin performansını göstermektedir. Potansiyel iyileştirmeler, modelin daha da geliştirilmesi için öneriler sunmaktadır.

Bu rapor, Iris veri kümesi üzerinde temel veri analizi ve modelleme süreçlerini anlamak için faydalı bir kaynak sağlamaktadır.