**Rapor: Iris Veri Kümesi Analizi ve Modelleme**

**1. Giriş**

Bu rapor, Iris veri kümesi üzerinde gerçekleştirilen veri temizleme, keşifsel veri analizi (EDA), görselleştirme ve modelleme sürecini kapsamaktadır. Amaç, iris çiçeklerinin çeşitli özelliklerini kullanarak türlerini sınıflandırmak ve modelin performansını değerlendirmektir. Ayrıca, potansiyel iyileştirmeler ve gelecek çalışmalar hakkında öneriler sunulmuştur.

**2. Veri Kümesi Hakkında**

Iris veri kümesi, iris çiçeklerinin sepal uzunluğu, sepal genişliği, petal uzunluğu ve petal genişliği gibi özelliklerini içermektedir. Her bir gözlem için çiçeğin türü ('setosa', 'versicolor', 'virginica') belirtilmiştir. Veri kümesi, toplamda 150 gözlem ve 5 sütun içermektedir.

**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)

Sonuçlar:

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

**4. Keşifsel Veri Analizi (EDA)**

EDA sırasında temel istatistikler hesaplandı ve sınıf dağılımı incelendi.

Temel İstatistikler:

print(iris.describe())

Sınıf Dağılımı:

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

Bulgular:

* Sepal ve petal uzunluk ve genişlikleri arasında belirgin farklar bulunmaktadır.
* Her bir özelliğin ortalama ve standart sapma değerleri incelenmiştir.
* Üç sınıf da eşit sayıda gözlem içermektedir (her biri 50 gözlem).

**5. Veri Görselleştirme**

Veri görselleştirme aşamasında, 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()

Isı Haritası:

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

corr\_matrix = numeric\_columns.corr()

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

plt.show()

Bulgular:

* Farklı sınıfların özellikleri arasındaki dağılımlar ve ilişkiler görselleştirildi.
* Sepal ve petal uzunluk ve genişlikleri arasındaki korelasyonlar incelendi.
* Özellikle petal uzunluğu ve genişliği arasında yüksek bir korelasyon gözlemlendi.

**6. Potansiyel İyileştirmeler**

* **Özellik Mühendisliği:** Veri kümesindeki özellikleri zenginleştirerek model performansı iyileştirilebilir.
* **Hiperparametre Optimizasyonu:** Modelin hiperparametre ayarları optimize edilerek daha iyi sonuçlar elde edilebilir.
* **Çapraz Doğrulama:** Çapraz doğrulama kullanılarak modelin genelleme yeteneği artırılabilir.

**8. Sonuç**

Bu çalışma, Iris veri kümesi üzerinde gerçekleştirilen veri temizleme, keşifsel veri analizi ve görselleştirme adımlarını özetlemektedir.

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