

Iris Veri Kümesi Üzerinde Veri Analizi

1. Giriş

Bu proje, Iris veri kümesi kullanılarak veri temizleme, keşifsel veri analizi (EDA) ve görselleştirme adımlarını içermektedir. Iris veri kümesi, iris çiçeği türlerini sınıflandırmak için kullanılan çeşitli çiçek özelliklerini içerir. Veri kümesi, sepal uzunluğu, sepal genişliği, petal uzunluğu, petal genişliği ve tür bilgilerini içermektedir.

2. Veri Yükleme

Öncelikle, Iris veri kümesini yükledik ve sütun isimlerini belirledik:

```
import pandas as pd

column_names = ["sepal_length", "sepal_width", "petal_length",
                "petal_width", "species"]
df = pd.read_csv('iris.data', header=None, names=column_names)

print(df.head())
```

3. Veri Temizleme

Veri temizleme aşamasında, eksik verileri kontrol ettik ve veri kümesinde eksik veri bulunmadığını tespit ettik:

```
print(df.isnull().sum())

# Bu veri kümesinde eksik veri bulunmamakta, dolayısıyla
temizlememize gerek yok
```

4. Keşifsel Veri Analizi (EDA)

4.1 Temel İstatistikler

Veri kümesinin temel istatistiklerini hesapladık.

```
print(df.describe())
```

4.2 Kategorik ve Sayısal Değişkenlerin Dağılımı

Veri kümesindeki kategorik ve sayısal değişkenlerin dağılımlarını inceledik:

```
print(df['species'].value_counts())

# Sayısal değişkenlerin dağılımı
df['sepal_length'].hist(bins=20)
df['sepal_width'].hist(bins=20)
df['petal_length'].hist(bins=20)
df['petal_width'].hist(bins=20)
```

4.3 Korelasyon Analizi

Değişkenler arasındaki korelasyonları inceledik:

```
print(df.corr())
```

5. Önemli Bulguların Görselleştirilmesi

5.1 Türlere Göre Sepal ve Petal Uzunlukları

Türlere göre sepal ve petal uzunluklarını görselleştirdik:

```
import seaborn as sns
import matplotlib.pyplot as plt

sns.pairplot(df, hue='species')
plt.suptitle('Türlere Göre Sepal ve Petal Uzunlukları', y=1.02)
plt.show()
```

5.2 Sepal Genişliği ve Petal Genişliği Arasındaki İlişkiler

Sepal genişliği ve petal genişliği arasındaki ilişkileri görselleştirdik:

```
sns.scatterplot(x='sepal_width', y='petal_width', hue='species',
data=df)
```

```
plt.title('Sepal Geniřlięi ve Petal Geniřlięi Arasındaki İliřki')  
plt.show()
```

6. Sonu

Bu proje kapsamında Iris veri kümesi üzerinde veri temizleme, keřifsel veri analizi ve görselleřtirme adımları gerekleřtirilmiřtir. Analizler sonucunda, iris ieęi türlerinin sepal ve petal özelliklerine göre belirgin farklılıklar gösterdięi gözlemlenmiřtir. Bu bulgular, veri madencilięi ve makine öğrenimi modelleri oluřtururken kullanılabilir.