



Proje

Milli Teknoloji Akademisi programı kapsamında İleri Seviye Python derslerinde elde ettiğiniz bilgileri kullanarak bu projeyi gerçekleştirmenizi beklemekteyiz. Sizlerle paylaşılan veri setini kullanarak aşağıdaki adımları gerçekleştirmeniz gerekmektedir.

Veri Seti : [Breast Cancer Wisconsin \(Diagnostic\) Data Set](#)

Breast Cancer Wisconsin (Diagnostic)		
Donated on 10/31/1995		
Diagnostic Wisconsin Breast Cancer Database.		
Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Health and Medicine	Classification
Feature Type	# Instances	# Features
Real	569	30

Veri seti ile ilgili detaylı açıklama yukarıdaki linkte mevcuttur.

Streamlit Nedir?

Streamlit, veri bilimcileri ve geliştiriciler için hızlı ve etkileşimli web uygulamaları oluşturmayı kolaylaştıran bir Python kütüphanesidir. Streamlit, veri analizi, görselleştirme ve model sonuçlarını paylaşmak isteyen kişilere, kullanıcı dostu bir arayüz üzerinden etkileşimli deneyimler sunma imkanı sağlar. Streamlit kullanmanın bazı avantajları şunlardır:

Kolay Kullanım: Streamlit, kullanımı basit ve anlaşılır bir API'ye sahiptir. Python bilen bir kişi, hızlıca bir web uygulaması oluşturabilir.

Hızlı Prototip Oluşturma: Streamlit, hızlı prototip oluşturmayı destekler. Veri bilimcileri veya geliştiriciler, fikirlerini hızla test etmek ve görselleştirmek için Streamlit'i kullanabilirler.



Proje

Etkileşimli Görselleştirmeler: Veri bilimcileri, analistler veya geliştiriciler, veri analizi sonuçlarını interaktif grafikler, tablolar veya diğer görselleştirmelerle paylaşabilirler.

Doğrudan Python Kullanımı: Streamlit ile uygulama geliştirmek için başka bir dil veya araç öğrenmeye gerek yoktur. Sadece Python bilgisi yeterlidir.

Hızlı Dağıtım: Oluşturulan uygulamalar, basit bir şekilde paylaşılabilir ve dağıtılabilir. Bu, işbirliği ve son kullanıcılarla paylaşım açısından avantaj sağlar.

Örneğin, veri analiz sonuçlarını veya makine öğrenmesi modellerini hızlı bir şekilde paylaşmak isteyen bir veri bilimcisi, Streamlit kullanarak bu sonuçları interaktif bir web uygulamasına dönüştürebilir.

Daha detaylı inceleme için [tıklayınız](#).

Örnek Python uygulaması için [tıklayınız](#).

Sizden istenen ilgili veri setini kullanarak aşağıdaki proje adımlarını gerçekleştirmenizdir.

Görevler	Görev Adımı	Puanlama
Görev 1	Data Analizi	10p
Görev 2	Data Ön İşlemleri	15p
Görev 3	Model Gerçeklemesi	15p
Görev 4	Model Analizi	20p



Proje

Görev 5

Streamlit Entegrasyonu

40p

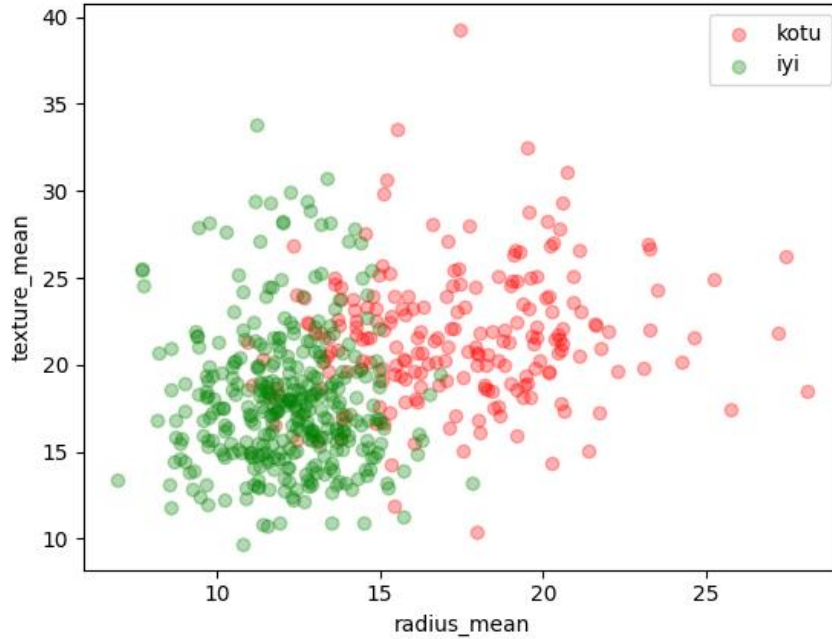
Görev 1

- Datayı lokalden seçip yükle. (Streamlit arayüzde sidebar ile Veri Seti ismi seçilecek, ilgili veri seçildiğinde Python datayı yükleyecek.)
- Streamlit ana ekrana datanın ilk 10 satırını içeren dataframe'i ve datanın sütunlarını göster.

Görev 2

Bu adımda data temizle ve data ön işleme adımları gerçekleştirilecektir.

- Gereksiz columnları temizle.
- Datanın son 10 row değerini göster.
- 'diagnosis' sütunundaki M değerini 1, B değerini 0 olacak şekilde değiştir. Bu sütunu Y etiket verisi olarak kullan, geri kalan sütunları ise X öznitelik verisi olarak kullan.
- Seaborn kütüphanesi ile korelasyon matrisini çizdir. Malignant ve benign olacak şekilde datayı ayır ve x:radiusmean y:texturemean olacak şekilde aşağıdaki gibi çizdir.



- Veriyi X_train, Y_train, X_test ve Y_test olarak yüzde 80-20 oranında ayır.



Proje

! Bu görevi streamlitten bağımsız olarak yap. Yani buton ya da herhangi bir şeyle yapılmayacak. Data yüklenince script kendisi yapacak.

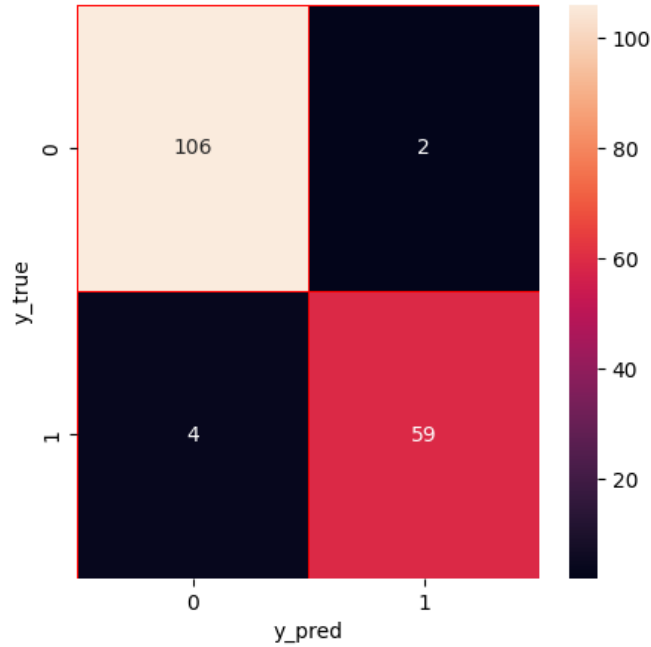
Görev 3

- Model implementasyonunu gerçekleştir.
- Streamlit sidebar üzerinden KNN, SVM, Naïve Bayes yönteminden birini seçince bu modele göre işlemler devam etsin.
- Gridsearch ile en iyi parametrelerini bul.
- Optimum parametrelere göre modeli X_train ve Y_train verisi ile eğit.

Görev 4

Model sonuçlarını X_test ve Y_test verisi için streamlit arayüzünde göster.

> Accuracy, precision, recall, f1-score ve confusion matrix (aşağıdaki gibi).



Görev 5: İlk 4 görev içerisinde istenmektedir.