TBF495 Python Programlama ve Veri Bilimi Dersi

2nci Proje Raporu

(Makine Öğrenmesi)

Hazırlayanlar

* 21993774 Eylül Begüm Özdaş

Mayıs 2022

# İÇİNDEKİLER

[İÇINDEKİLER 2](#_Toc66019504)

[1. Projenin Tanımı ve Başlatılma Aşaması 3](#_Toc66019505)

[1.1. Proje Çerçevesi 3](#_Toc66019506)

[1.2. Projenin Amaçları 4](#_Toc66019507)

[1.3. Proje Başlatma Belgesi 4](#_Toc66019508)

[1.3.1 Proje Yönetim Modeli 5](#_Toc66019509)

[1.3.2 Projenin Başarı Ölçütleri 5](#_Toc66019510)

[1.3.3 Proje Paydaşları ve Ekibi 6](#_Toc66019511)

[2. Veri Bilimi Aşaması 6](#_Toc66019512)

[2.1. Veri Kaynağı ve Elde Edilmesi 6](#_Toc66019513)

[2.2. Verinin Hazırlanması 6](#_Toc66019514)

[2.2.1 Verinin Temizlenmesi 7](#_Toc66019515)

[2.2.2 Verinin Birleştirilmesi 7](#_Toc66019516)

[2.2.3 Verinin Dönüştürülmesi 7](#_Toc66019517)

[2.2.4 Keşfe Yönelik Veri Analizi (Exploratory Data Analysis) 7](#_Toc66019518)

[3. Makine Öğrenmesi Aşaması 7](#_Toc66019519)

[3.1. Problemin Çözümüne Yönelik Modeller (Algoritma + Veri) 7](#_Toc66019520)

[3.2. Modellerin Eğitimi ve Test Edilmesi 8](#_Toc66019521)

[3.3. Modellerin Değerlendirilmesi 8](#_Toc66019522)

[4. Proje Sonlandırma Aşaması 8](#_Toc66019523)

[4.1. Proje Sonlandırma İşlemleri 8](#_Toc66019524)

[4.2. Proje Görevlerinin Takım Üyelerine Dağılımı 8](#_Toc66019525)

# Projenin Tanımı ve Başlatılma Aşaması

* Proje sonuç raporunda aşağıdaki proje çerçevesinde belirtilen, Ürün-2’e ait veri bilimi ve makine öğrenmesin ait süreçlerle ilgili bilgiler bu raporda yer alacaktır.

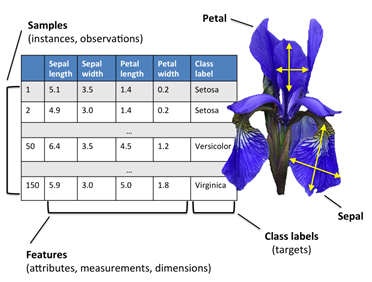
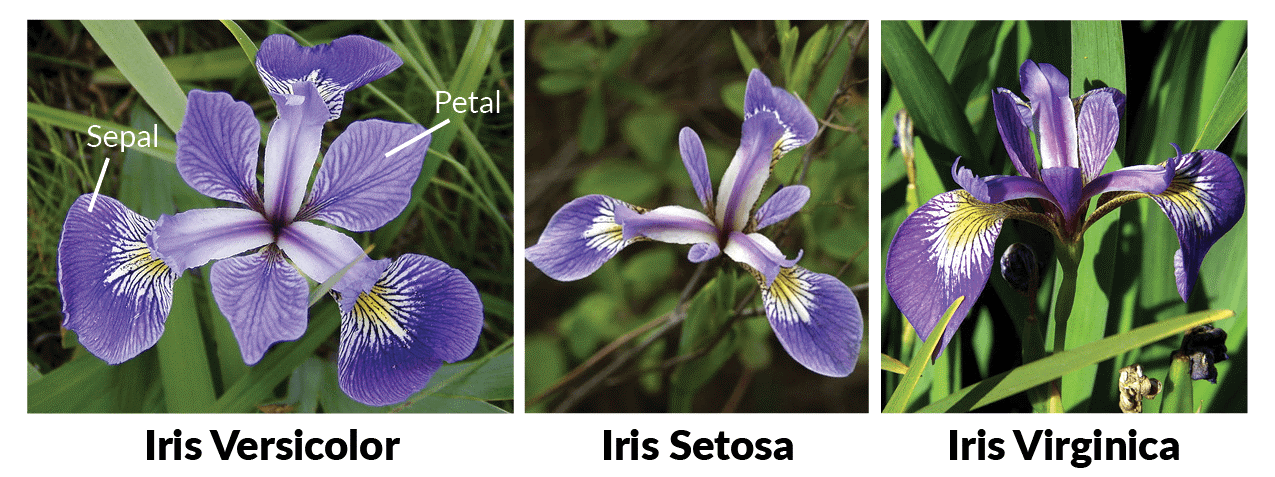
## Proje Çerçevesi

Türkiye’nin çeşitli coğrafi bölgelerinde bulunan Süsen (İris) çiçekleriyle ilgili ülke çapında bir bilimsel araştırma yürütülmektedir. Söz konusu çalışmayı desteklemek amacıyla proje ekibinizden veri bilimi ve makine öğrenmesi destekli bütünleşik bir yazılım geliştirmeniz beklenmektedir. Dolayısıyla dönem projesinde, aşağıda belirtilen ihtiyaçları karşılamak üzere makine öğrenmesi yazılımı ürün olarak geliştirilecektir:

* **Ürün-2:** **İris çiçeklerinin türünü tahmin etmede kullanılacak, Python Programlama Diliyle geliştirilmiş bütünleşik makine öğrenmesi yazılımıdır.**

İris çiçek türleri ve özellikleri Şekil 1.1’de gösterilmiştir.:

**Şekil-1.1**.İris Çiçek Türleri ve Özellikleri (Feature)



Makine öğrenmesi projesi yazılımın ihtiyaçları veri bilimi ve makine öğrenmesi başlıkları altında gruplanarak aşağıdaki tabloda gösterilmiştir:

|  |  |
| --- | --- |
| **S. Nu** | **İhtiyaç Listesi** |
| **Veri Bilimi** | |
| 1 | Verinin Hazırlanması (temizlenmesi, birleştirilmesi, dönüştürülmesi ve görselleştirilmesi) Etkinlikleri |
| 2 | Keşfe Yönelik Veri Analizi Etkinlikleri |
| **Makine Öğrenmesi** | |
| 1 | Modellerinin (Algoritma + Veri) Belirlenmesi ve Uygun Modelin Seçimi |
| 2 | Modellerin Eğitimi ve Test Edilmesi |
| 3 | Modellerin Değerlendirilmesi ve İyileştirilmesi |
| **Sonuç Raporu** | |
| 4 | Proje Sonuçların Raporlanması ve Sunulması |

## Projenin Amaçları

Iris verileri analiz edilerek, türleri tahmin edilmesini sağlayan bir makine öğrenmesi çalışmaları yapılacak.

Algoritmalar test edilip en yüksek, doğru sonuç tespit edilecek.

## Proje Başlatma Belgesi

Projenin en az %90 başarılı olması beklenmektedir. Proje de çiçek türlerinin en az %90 şekilde doğruluk payı vermesi beklenmektedir. Proje de 5 algoritma kullanılacaktır.

Naive Bayes-sınıflandırıcı metottur. İstatistik bağımsızlık koşulunu esnetilerek kullanılır. İris kümesini sınıflandırmak için kullanacağız. Sonuç %90 beklenmektedir

Random Frost-karar ağacıdır. Problemin tipine göre sınıf veya sayı tahmini yapan yöntemdir. İris verisinde eğitip test edeceğiz. Sonuç %90 beklenmektedir.

Decison tree-ağaç tabanlı öğrenme modelleridir.

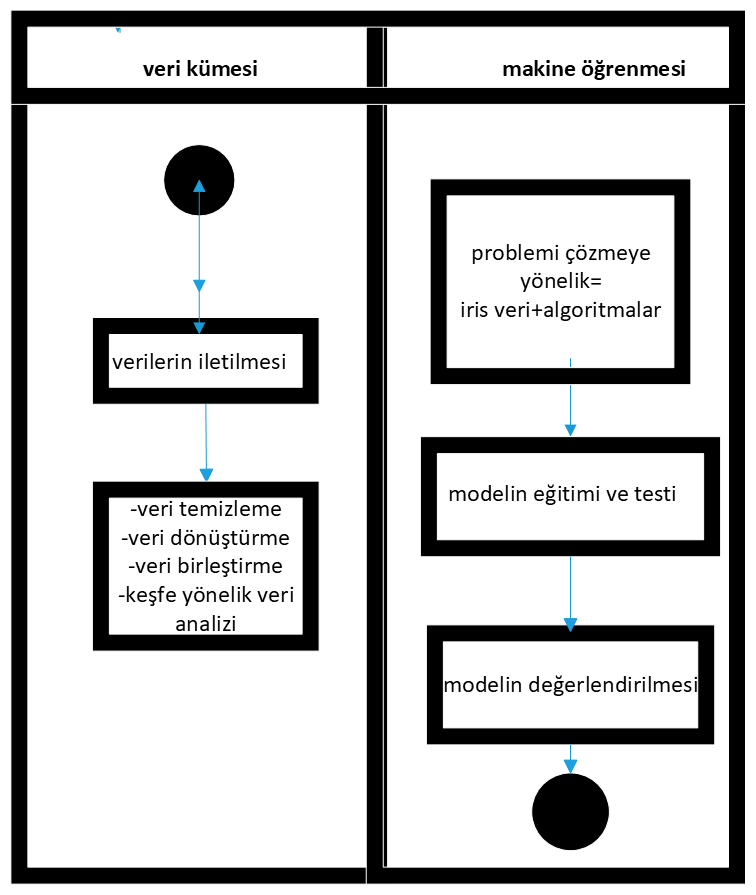
KNN-tahmin edilecek değerin bağımsız değişkenlerinin oluşturduğu en yakın komşuyu,hangi sınıfta yoğun olduğu bulmak için kullanılır. İris veri kümesinde en yakın kümeleri bulacak.

KMeans-küme içi benzerliğin maksimum ve kümeler arası benzerliğin ise minumum olmasını sağlamalıyız. İris veri kümesinde maks, min değerleri inceleyeceğiz

Logistic Resgrssion-beklenen değerlerin olasılık olarak elde edildiği bir regrasyon yöntemidir. İris veri kümesinde beklenen sonuç%90dır

### Proje Yönetim Modeli

Projede kullanacağınız proje yönetim modeli bu proje sonuç raporunun içendekiler tablosundaki ilgili başlıklar (veri bilimi ve makine öğrenmesi) seçilerek belirlenecektir. Bu adımlar UML etkinlik çizeneğiyle gösterimi yapılarak aşağıda gösterilecektir.

****

**Şekil-1.3.1.** Proje Yönetim Etkinlik Çizeneği (Activity Diagram)

### Projenin Başarı Ölçütleri

* Geliştirilecek makine öğrenmesi yazılımının çıktılarının, ihtiyaç makamı olan kurum veya organizasyonun hangi ihtiyaçlarını karşılayacağı ile bunun nasıl ölçüleceğine ilişkin ölçütleri belirtiniz.
* Geliştireceğiniz kestirim, sınıflama veya kümelemeye yönelik makine öğrenmesi modelinin yüzdelik cinsinden başarı ölçütlerini belirtiniz.

### Proje Paydaşları ve Ekibi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S. Nu.** | **Projedeki Görevi** | **Görev Kodu** | **Öğrenci Numarası** | **Adı ve Soyadı** |
| 1 | Takım Lideri | T-1 | 21993774 | Eylül Begüm Özdaş |

# Veri Bilimi Aşaması

## Veri Kaynağı ve Elde Edilmesi

* Sklearn kütüphanesinden Iris veri kümesini kullanacağız. Veriler de Iris çiçeğinin özellikleri inlecenecek.

## Verinin Hazırlanması

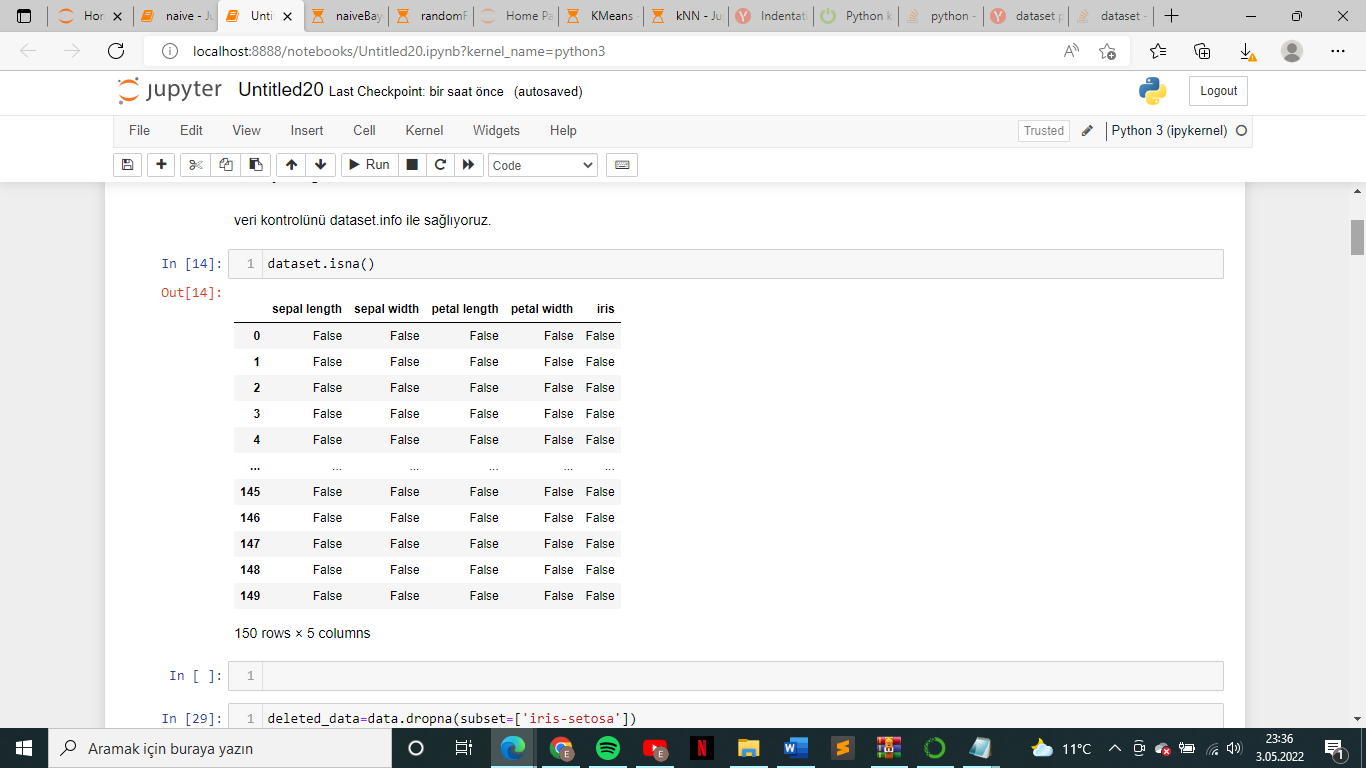
İris veri kümesi bilgilerini çağırıyoruz.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

### Verinin Temizlenmesi

Temizlenmiş ilk 150 veriyle çalıştığı için eksik değer veri kümesi yoktur

.

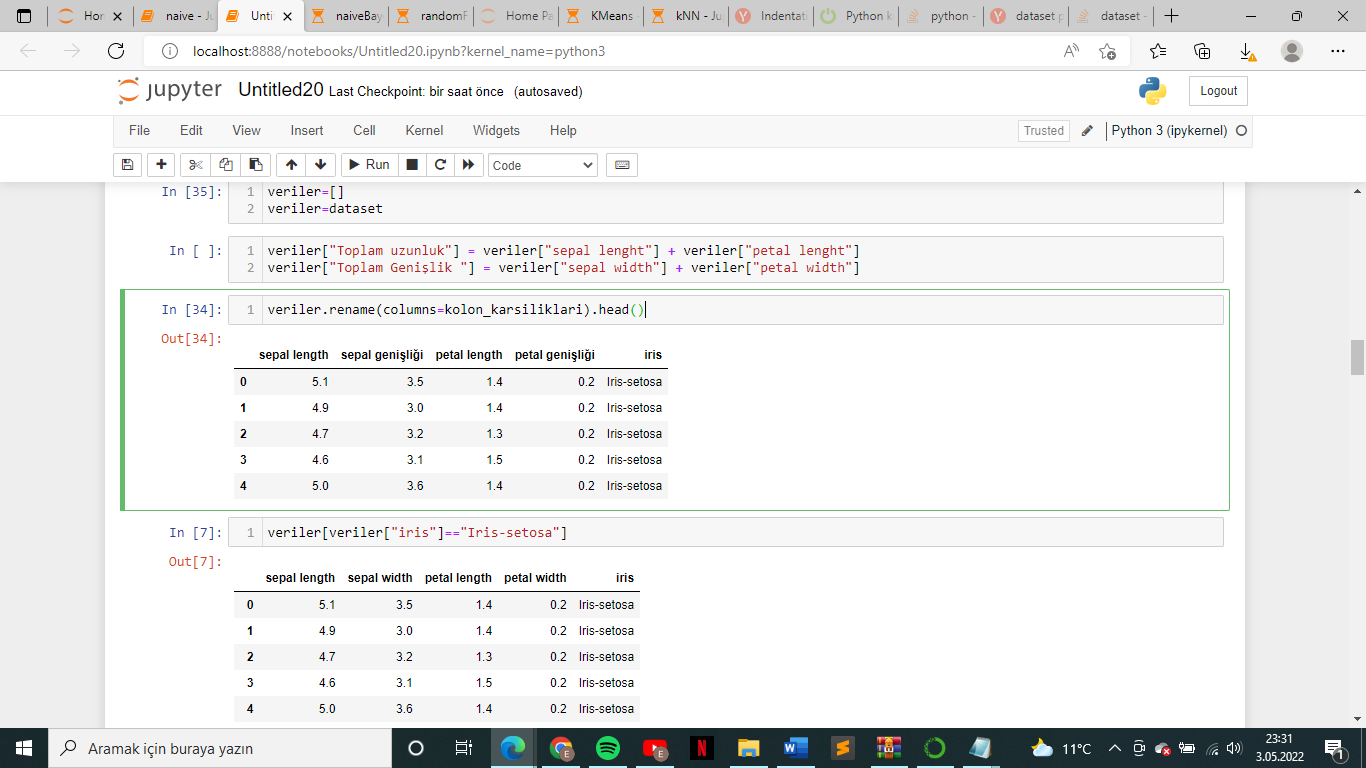
Isna().sum() komutu eksik ya da boş deperlerin miktarını gösterir. Görüldüğü gibi bu küme boş.

metin, ekran görüntüsü, iç mekan içeren bir resim

Açıklama otomatik olarak oluşturuldu

### Verinin Birleştirilmesi

Iris çiçeğinin yaprakları ve genişlikleri alıyoruz.



### Verinin Dönüştürülmesi

Çiçek türleirni etiketleyip değerlerini atadık.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

### Keşfe Yönelik Veri Analizi (Exploratory Data Analysis)

Bu bölümde verinin temsili, gösterimi ve görselleştirilmesiyle ilgi yapılan etkinlikler, ilgili *grafik ve tablolar*, Python program kodları ver ekran görüntüleri yer alacaktır.

Çiçek türüne göre sorgulaıyoruz

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

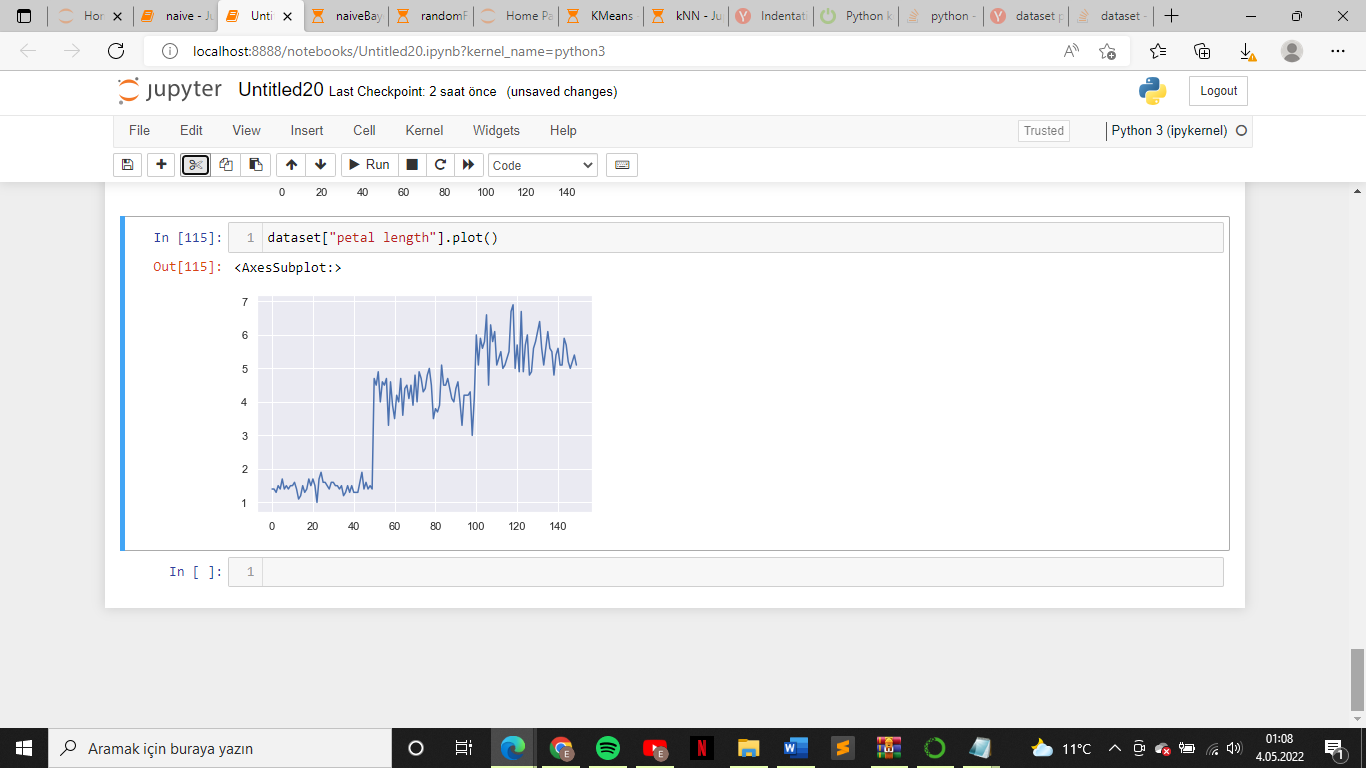
Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu



# Makine Öğrenmesi Aşaması

## Problemin Çözümüne Yönelik Modeller (Algoritma + Veri)

* Navie Bayes(sınıflama)
* Random Forest(sınıflama)
* Decions Tree(sınıflama)
* K-Means(kümeleme)
* K-NN(sınıflama)
* Logistic Regression.(sınıflama)

## Modellerin Eğitimi ve Test Edilmesi

* **Navie Bayes**

.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Random Forest**

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Decision Tree**

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **K-Means**

metin içeren bir resim

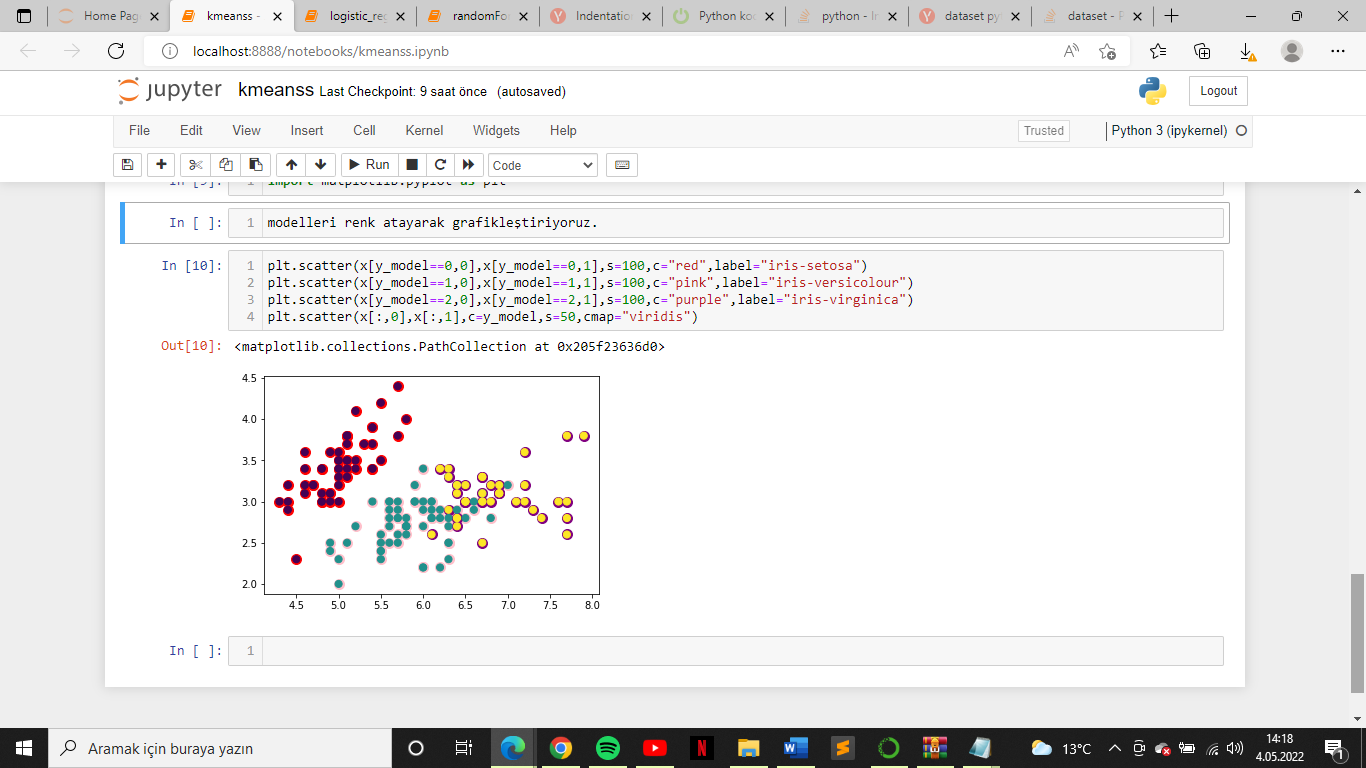
Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu



* KNN

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

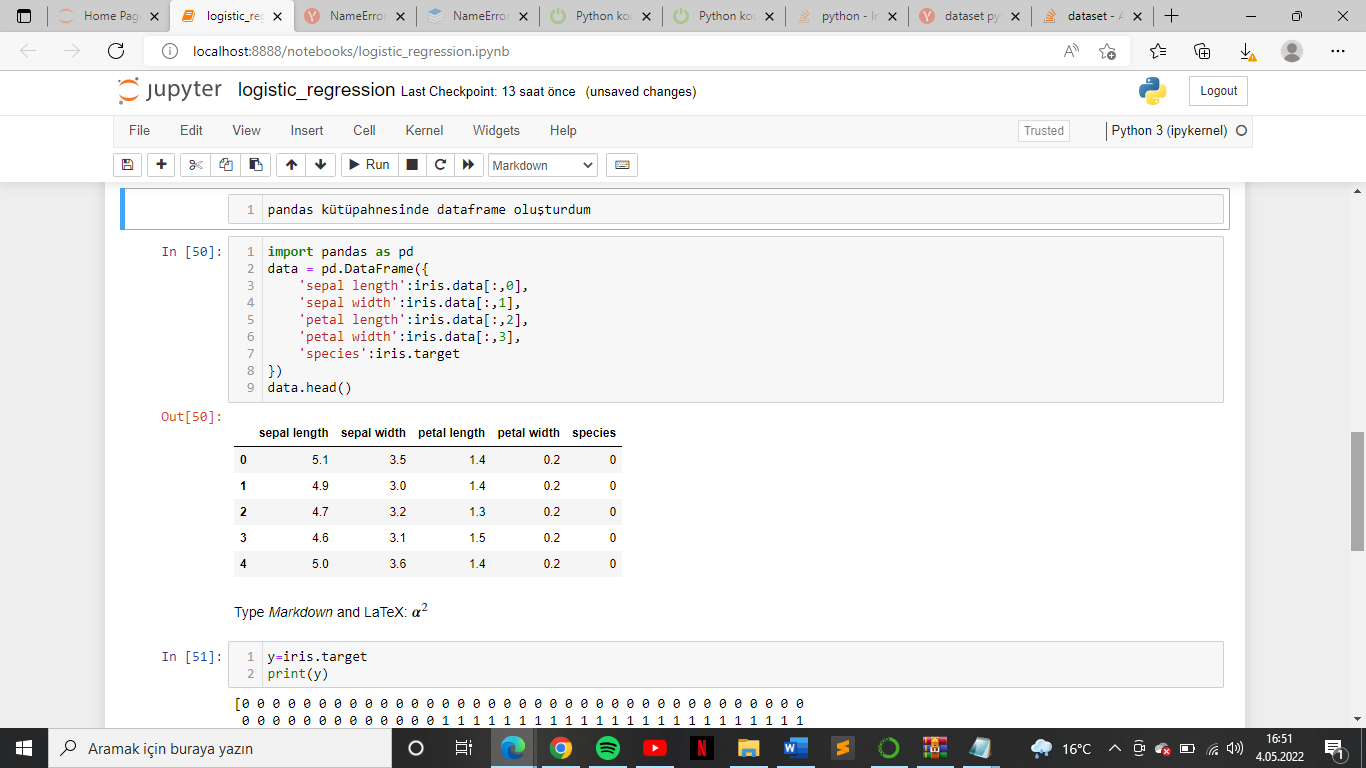
* **Logistic Regression**

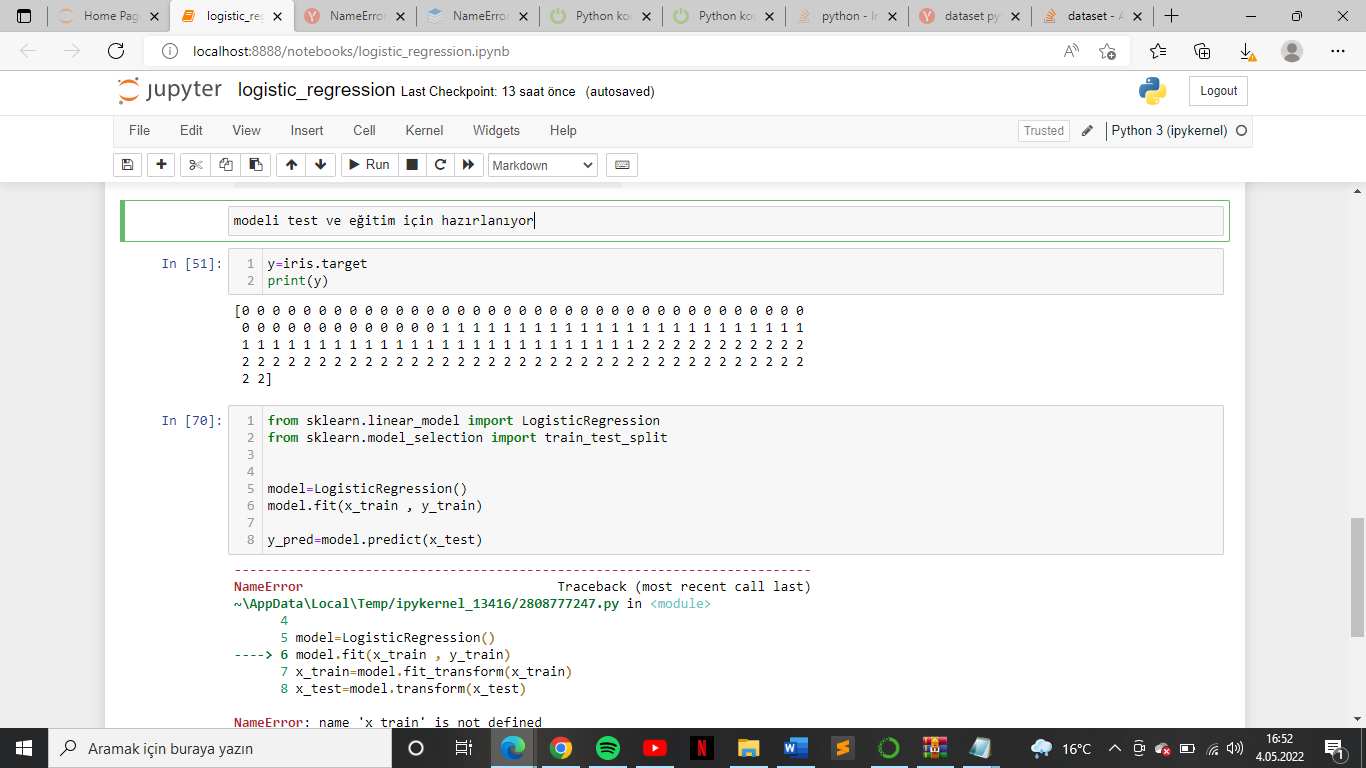
metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu





metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

## Modellerin Değerlendirilmesi

Bir önceki bölümde yapmış olduğunuz eğitim ve test işlemlerinin sonuçlarını değerlendiriniz, problemin çözümüne yönelik en uygun modeli seçiniz ve gerekçesini açıklayınız. İlgili grafikleri, program kodlarını ve ekran görüntülerini paylaşınız.

Naive Bayes:

Burda iki sonucu yan yana koydum. Proje öncesi beklentimiz %90 doğruluktu %93 ile bu doğruluğu kazandık ancak görüldüğü üzere ne kadar veri aktarımı alırsak doğruluk o kadar yükseliyor. Verilerin 4 sütununu da alırsak %93lük bir doğruluk payı ile karşılaşıyoruz. Verilerin 1 sütununu alırsak %73 lük doğruluk payı ile karşılaşıyoruz.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

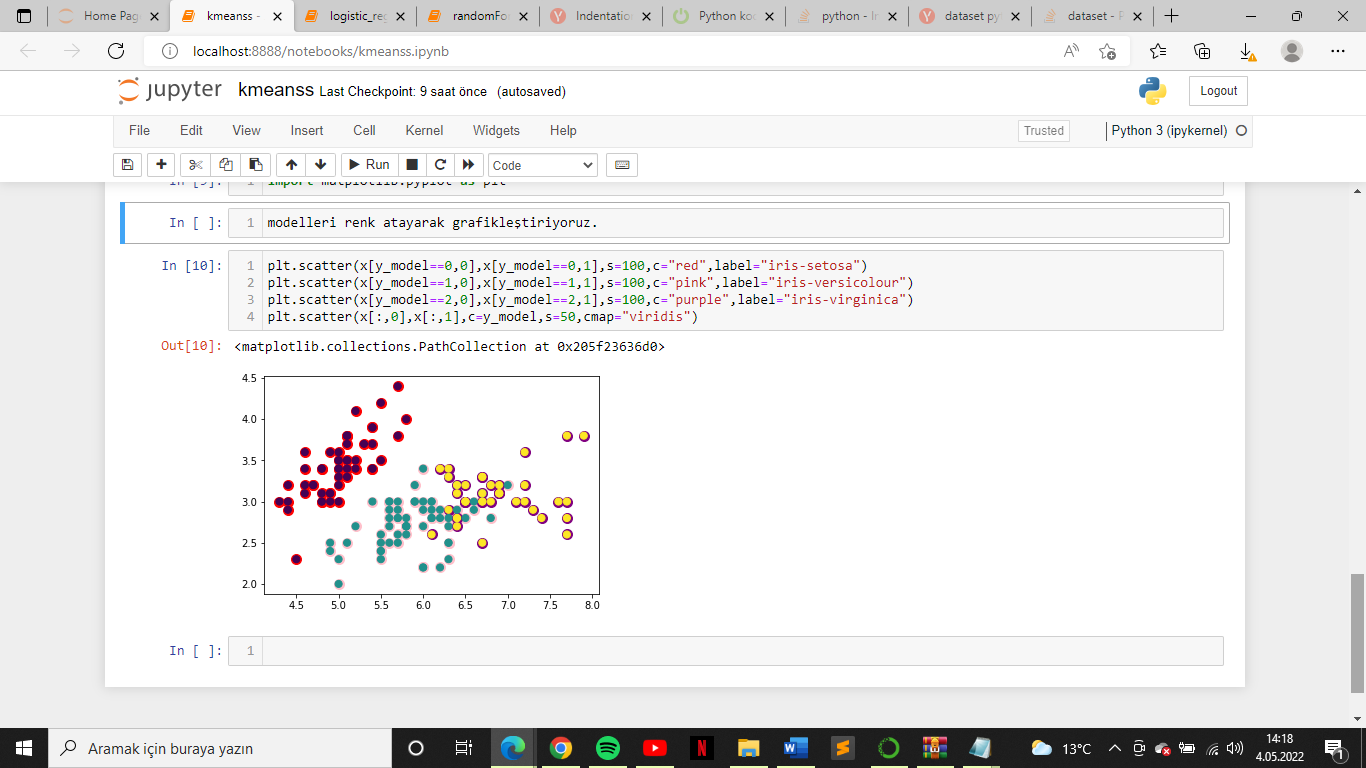
Random Forest:

Proje de %90lık doğruluk payını bekliyorduk. %93 ile yüksek bir pay yakaladık.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

K-Means:



KNN:

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Logistic regression:

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

# Proje Sonlandırma Aşaması

## Proje Sonlandırma İşlemleri

* Geliştirilen ürün ve dokümanların ihtiyaç makamına teslim edilmesi ve projenin sonlandırılması işlemlerini içermektedir.
* Bu bölüm aynı zamanda Teslim edilen ürünün izlenmesi, düzeltici ve iyileştirici yazılım bakım uygulamalarını da içermelidir.

.

## Proje Görevlerinin Takım Üyelerine Dağılımı

Aşağıdaki tabloda projeyle ilgili görev ve sorumlulukların takım üyelerine (görev kodları kullanılarak) nasıl paylaştırıldığı, dağılım oranları gösterilecektir.

| **S.**  **Nu.** | **Projedeki Görevlerin Takım Üyelerine Dağılımı** | **Yüzdelik Dağılımları (%25, %50, %75, %100)** |
| --- | --- | --- |
| **T-1**  **(%)** |
| **1.** | **Projenin Tanımı ve Başlatılma Aşaması** | |
| 1.1 | Projenin tanımı ve başlatma belgesi | 100 |
| **2** | **Veri Bilimi Aşaması** |  |
| 2.1 | Verinin hazırlanması, temizlenmesi, birleştirilmesi, görselleştirilmesi | 100 |
| 2.2 | Keşfe yönelik veri analizi (exploratory data analysis) | 100 |
| **3** | **Makine Öğrenmesi Aşaması** |  |
| 3.1 | Modellerin belirlenmesi | 100 |
| 3.2 | Modellerin eğitimi ve test edilmesi | 100 |
| 3.3 | Modellerin değerlendirilmesi | 100 |
| **4** | **Proje Sonlandırma Aşaması** |  |
| 4.1 | Proje raporunun hazırlanması | 100 |
|  | **ORTALAMA (%)** | 100 |