Python Programlama ve Veri Bilimi

Proje Raporu

(Makine Öğrenmesi)

* Erdin Gülbudak

Mayıs 2024

# İÇINDEKİLER

[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.

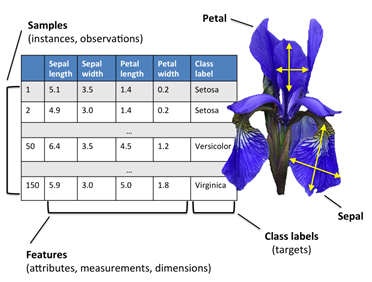
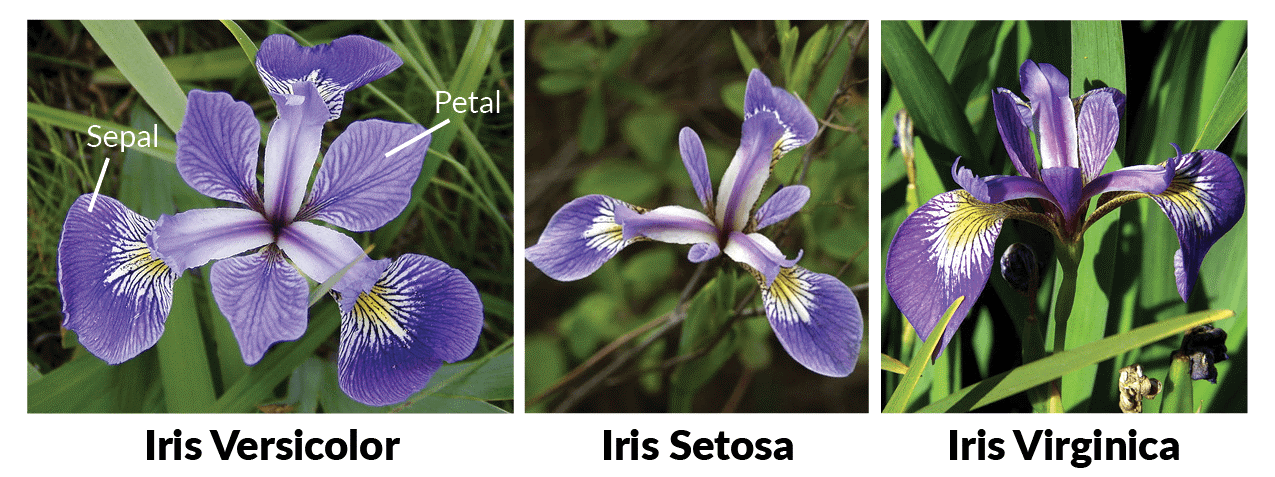
## Proje Çerçevesi

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

Bu bölüme projenin tanımından faydalanarak en genel seviyede proje amaç (lar)ı kısaca yazınız.

Bu proje, Türkiye'nin çeşitli coğrafi bölgelerinde bulunan Süsen (İris) çiçekleriyle ilgili ülke çapında bir bilimsel araştırmanın desteklenmesi amacıyla gerçekleştirilmektedir. Proje, veri bilimi ve makine öğrenmesi tekniklerini kullanarak İris çiçeklerinin türünü tahmin etmeye yönelik bir yazılım geliştirmeyi hedeflemektedir.

Bu kapsamda, proje ekibi, Python Programlama Dili ile geliştirilmiş bir makine öğrenmesi yazılımı olan Ürün-2'yi oluşturmayı planlamaktadır. Ürün-2, İris çiçeklerinin türünü belirlemede kullanılacak ve veri bilimi ile makine öğrenmesi süreçlerini bütünleştiren bir yazılım olacaktır.

Proje, İris çiçeklerinin türlerini ve özelliklerini içeren veri setlerini kullanarak şu ana başlıklar altında faaliyet gösterecektir:

1-) Veri Bilimi:

* Verinin Hazırlanması: Veri setlerinin temizlenmesi, birleştirilmesi, dönüştürülmesi ve görselleştirilmesi gibi etkinlikler.
* Keşfe Yönelik Veri Analizi: Veri setlerinin derinlemesine incelenmesi ve çeşitli veri analizi teknikleriyle keşfedilmesi.

2-) Makine Öğrenmesi:

* Modellerinin Belirlenmesi ve Uygun Modelin Seçimi: İris çiçeklerinin türlerini tahmin etmek için kullanılacak algoritmaların seçilmesi ve uygun modelin belirlenmesi.
* Modellerin Eğitimi ve Test Edilmesi: Seçilen modelin veri seti üzerinde eğitilmesi ve performansının test edilmesi.
* Modellerin Değerlendirilmesi ve İyileştirilmesi: Eğitilen modellerin performansının değerlendirilmesi ve gerekirse iyileştirilmesi.
* Sonuç olarak, proje sonuçları raporlanacak ve sunulacak, İris çiçeklerinin türlerini tahmin etmek için geliştirilen Ürün-2'nin başarıları ve geliştirme süreci detaylı bir şekilde belgelenerek paylaşılacaktır.

## Proje Başlatma Belgesi

Bu bölümde bir proje başlatma belgesinde bulunması gereken maddelerden önemli olduğu değerlendirilenler belirtilecektir.

1. \*\*Proje Tanımı ve Amaçları: \*\*

- Bu proje, Türkiye'nin çeşitli coğrafi bölgelerinde bulunan İris çiçekleriyle ilgili ülke çapında bir bilimsel araştırmanın desteklenmesi amacıyla gerçekleştirilmektedir. Projenin amacı, İris çiçeklerinin türlerini belirleme yeteneğine sahip bir makine öğrenmesi modeli geliştirmektir.

2. \*\*Proje Kapsamı: \*\*

- Proje, İris çiçeklerinin türlerini tahmin etmek için kullanılabilecek bir makine öğrenmesi modelinin geliştirilmesini içermektedir. Veri setinin hazırlanması, modelin seçimi ve eğitimi, performans değerlendirmesi ve sonuçların raporlanması projenin kapsamına dahildir.

3. \*\*Proje Paydaşları: \*\*

- Onur Kaan Hazer, Meryem Dönmez, Zeynep İrem Can, Erdin Gülbudak

4. \*\*Proje Süreci ve Adımları: \*\*

- Proje süreci, veri toplama ve hazırlama, model seçimi ve eğitimi, model performansının değerlendirilmesi ve sonuçların raporlanması adımlarını içerecektir.

5. \*\*Proje Kaynakları: \*\*

- Proje için gerekli kaynaklar insan kaynakları (veri bilimciler, yazılım geliştiriciler), finansal kaynaklar, veri setleri, yazılım ve donanımı içerecektir.

6. \*\*Proje Takvimi ve Mile Stone’lar: \*\*

- Proje başlangıç tarihi: [11.05.2024]

- İlk prototipin tamamlanması: [11.05.2024]

- Son versiyonun teslimi: [15.05.2024]

- Proje sonuç raporunun tamamlanması: [15.05.2024]

7. \*\*Riskler ve Risk Yönetimi Planı: \*\*

- Veri setinin kalitesi, modelin performansı, zaman ve bütçe kısıtlamaları gibi riskler tanımlanacak ve bu risklerin nasıl yönetileceği planlanacaktır.

8. \*\*İletişim Planı: \*\*

- Haftalık toplantılar, e-posta ile iletişim ve belirli aralıklarla raporlama gibi iletişim yöntemleri belirlenmiştir.

9. \*\*Kalite Standartları ve Kontrol Süreçleri: \*\*

- Proje boyunca kullanılacak kalite standartları belirlenecek ve model performansı düzenli olarak kontrol edilecektir.

10. \*\*Onay ve Kabul Kriterleri: \*\*

1-) \*\*Model Performansı: \*\*

Model doğruluğu %90'ın üzerinde olmalıdır.

2-) \*\*Uygulama Testi: \*\*

Model, test veri setinde başarılı bir şekilde çalışmalıdır.

3-) \*\*Kullanıcı Geri Bildirimi: \*\*

Kullanıcılar, modelin kullanımıyla ilgili geri bildirimlerde bulunmalıdır.

4-) \*\*Belgelendirme ve Sunum: \*\*

Proje sonuçları belgelenmeli ve paydaşlara sunulmalıdır.

5-) \*\*Proje Bütçesi ve Süresi: \*\*

Proje, belirlenen bütçe ve süre içinde tamamlanmalıdır.

6-) \*\*Kapsamın Karşılanması: \*\*

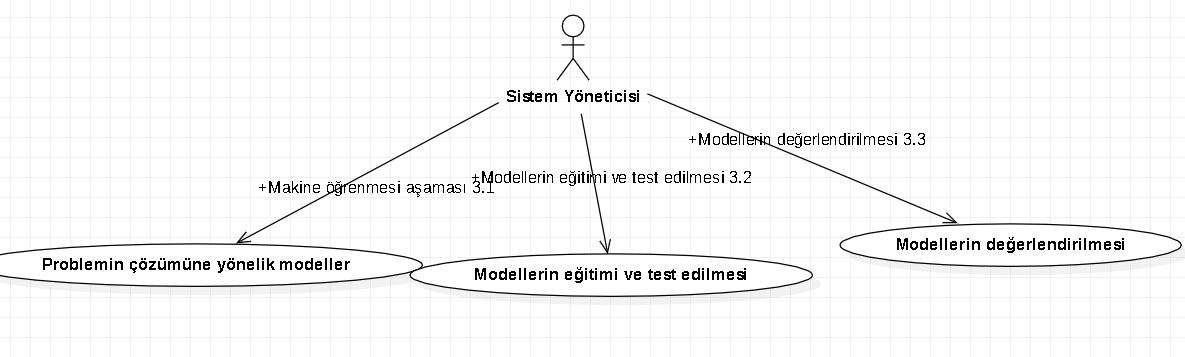
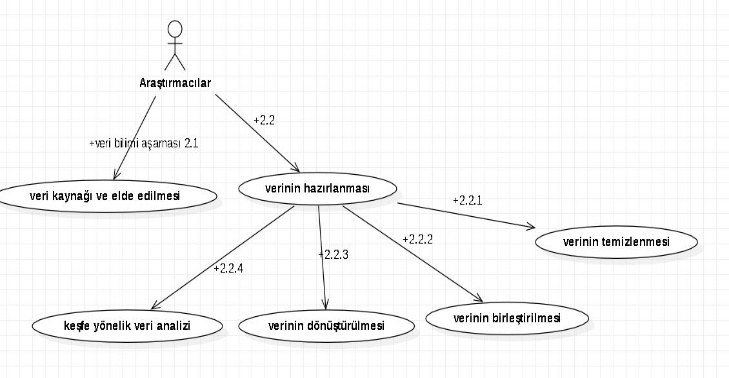
Tüm proje hedefleri ve faaliyetleri tamamlanmalıdır.

Bu kriterlerin karşılanmasıyla, proje başarıyla tamamlanmış ve kabul edilmiş sayılacaktır.

### 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 Modeli 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 | 22195689 | Onur Kaan Hazer |
| 2 | Takım Üyesi | Ü-1 | 22195250 | Meryem Dönmez |
| 3 | Takım Üyesi | Ü-2 | 22194799 | Zeynep İrem Can |
| 4 | Takım Üyesi | Ü-3 | 22296025 | Erdin Gülbudak |
| 5 | Takım Üyesi | Ü-4 |  |  |

# Veri Bilimi Aşaması

## Veri Kaynağı ve Elde Edilmesi

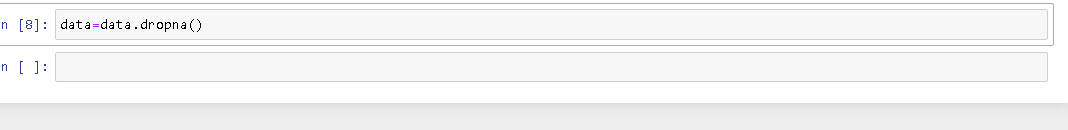
* Problemin çözümüne yönelik veri kaynaklarını, verinin nereden ve nasıl elde edileceğine ilişkin konuları belirtiniz.
* Problemin çözümüne yönelik değişkenler (özellikle) ile bunları veriyle olan ilişkisi de dikkat edilecektir.

**Önemli Not:** *Bu projede veri kaynağınız Python’un ilgili kütüphanelerinde zaten bulunmaktadır. Sizden bu bölümde beklenilen derste ve ilgili kaynaklarda görüşülen konular doğrultusunda önemli hususları genel geçer seviyede belirtmenizdir*.

## Verinin Hazırlanması

Bu bölüme veri bilimi kapsamında aşağıdaki başlıklar altına projede yapmış olduğunuz etkinlikleri belirtiniz, Python program kodlarını ver ekran görüntülerini paylaşınız

### Verinin Temizlenmesi



### Verinin Birleştirilmesi

metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

### Verinin Dönüştürülmesi

metin, ekran görüntüsü, ekran, görüntüleme, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

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

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi, ekran, görüntüleme içeren bir resim

Açıklama otomatik olarak oluşturuldu

# Makine Öğrenmesi Aşaması

## Problemin Çözümüne Yönelik Modeller (Algoritma + Veri)

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

## Modellerin Eğitimi ve Test Edilmesi

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

## Modellerin Değerlendirilmesi

Bu noktada en uygun model/modeller :RandomForest veya DesicionTree olacaktır çünkü doğruluk oranları %100 ve ana model ile aynı

ekran görüntüsü, renklilik, diyagram, metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

ekran görüntüsü, renklilik, metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

# Proje Sonlandırma Aşaması

## Proje Sonlandırma İşlemleri

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu