

# EGZERSIZ VERİMLİLİĞİ İÇİN KALORI TAHMIN MODELİ: AKILLI SPOR TAKİBİ

Kişisel verileri(cinsiyet, yaş, boy, kilo) bilinen bir kullanıcının yaktığı kalori miktarının hesaplanması.

Berkan Demiral

## Veri ile İlk Tanışma

Bu kısımda ilgili kütüphaneler (Pandas, numpy, sklearn, seaborn) import edildi ve veri dataframe'e aktarıldı.



## Keşifçi Veri Analizi

Veriyle tanışma kısmını bir kenara bırakarak onu daha iyi anlamaya ve analiz etmeye odaklanıldı. Değişkenleri arasında ilişkiler kuruldu.



## Özellik Mühendisliği

Kategorik değişkenlerin label encoder ile sayısallaştırılması ve aykırı değer analizleri yapıldı. Veri seti daha anlamlı hale getirildi.



## Makine Öğrenmesi

Makine Öğrenmesi modelleri ile bağımlı değişkene(yakılan kalori) en etkili değişkenler bulundu. Parametre optimizasyonu yapıldı ve model oluşturuldu.



# VERİ İLE TANIŞMA

Veri Seti olarak Kaggle'da bulunan "Gym Member Exercise Dataset" kullanıldı. Bu dataset genel olarak; bir kişinin spor salonundaki yaptığı egzersiz türü, süresi, nabız verileri ile yaktığı kalori miktarını vermektedir.



# VERİ SETİNE GENEL BAKIŞ

	Age	Gender	Weight (kg)	Height (m)	Max_BPM	Avg_BPM	Resting_BPM	Session_Duration (hours)	Calories_Burned	Workout_Type	Fat_Percentage	Water_Intake (liters)	Workout_Frequency (days/week)	Experience_Level	BMI
0	56	Male	88.3	1.71	180	157	60	1.69	1313.0	Yoga	12.6	3.5	4	3	30.20
1	46	Female	74.9	1.53	179	151	66	1.30	883.0	HIIT	33.9	2.1	4	2	32.00
2	32	Female	68.1	1.66	167	122	54	1.11	677.0	Cardio	33.4	2.3	4	2	24.71
3	25	Male	53.2	1.70	190	164	56	0.59	532.0	Strength	28.8	2.1	3	1	18.41
4	38	Male	46.1	1.79	188	158	68	0.64	556.0	Strength	29.2	2.8	3	1	14.39

- **Age (Yaş):** Spor salonu üyesinin yaşı.
- **Gender (Cinsiyet):** Spor salonu üyesinin cinsiyeti (Erkek veya Kadın).
- **Weight (kg) (Ağırlık - kg):** Üyenin kilosu, kilogram cinsinden.
- **Height (m) (Boy - m):** Üyenin boyu, metre cinsinden.
- **Max\_BPM (Maksimum Nabız):** Egzersiz sırasında ulaşılan maksimum kalp atış hızı (dakikada atış).
- **Avg\_BPM (Ortalama Nabız):** Egzersiz sırasında ortalama kalp atış hızı.
- **Resting\_BPM (Dinlenme Nabızı):** Egzersiz öncesindeki dinlenme halindeki kalp atış hızı.
- **Session\_Duration (hours) (Seans Süresi - saat):** Her bir egzersiz seansının süresi, saat cinsinden.
- **Calories\_Burned (Yaktırılan Kalori):** Her egzersiz seansında yakılan toplam kalori.
- **Workout\_Type (Egzersiz Türü):** Yapılan egzersizin türü (Örneğin, Kardiyo, Kuvvet, Yoga, HIIT).
- **Fat\_Percentage (Yağ Oranı):** Üyenin vücut yağ oranı.
- **Water\_Intake (liters) (Su Tüketimi - litre):** Egzersiz sırasında günlük su tüketimi.
- **Workout\_Frequency (days/week) (Egzersiz Sıklığı - gün/hafta):** Haftada yapılan egzersiz seansı sayısı.
- **Experience\_Level (Deneyim Seviyesi):** Üyenin deneyim seviyesi, başlangıç (1) ile uzman (3) arasında.
- **BMI (Vücut Kitle İndeksi):** Boy ve kilo hesaplamasına dayalı olarak hesaplanan vücut kitle indeksi.

# VERİ SETİNE GENEL BAKIŞ

	Data Type	Missing	Duplicate	Unique	MIN	MAX	AVG	STD-DEV
Age	int64	0	0	42	18.000000	59.000000	38.683453	12.180928
Gender	int64	0	0	2	0.000000	1.000000	0.525180	0.499622
Weight (kg)	float64	0	0	519	40.000000	126.400000	72.898356	20.013424
Height (m)	float64	0	0	51	1.500000	2.000000	1.722580	0.127720
Max_BPM	int64	0	0	40	160.000000	199.000000	179.883864	11.525686
Avg_BPM	int64	0	0	50	120.000000	169.000000	143.766701	14.345101
Resting_BPM	int64	0	0	25	50.000000	74.000000	62.223022	7.327060
Session_Duration (hours)	float64	0	0	147	0.500000	2.000000	1.256423	0.343033
Calories_Burned	float64	0	0	610	303.000000	1587.000000	896.542652	259.721982
Workout_Type	int64	0	0	4	0.000000	3.000000	1.494347	1.125472
Fat_Percentage	float64	0	0	239	10.000000	35.000000	24.976773	6.259419
Water_Intake (liters)	float64	0	0	23	1.500000	3.700000	2.626619	0.600172
Workout_Frequency (days/week)	int64	0	0	4	2.000000	5.000000	3.321686	0.913047
BMI	float64	0	0	749	12.320000	41.030000	24.386670	5.830857

# KEŞİFÇİ VERİ ANALİZİ



## Sayısal Değerler Gerçekten Sayısal Mı?

Kategorik ve sayısal değişkenler belirlenirken sayısal gibi görünen ama kategorik kimliğe sahip değişkenler açığa çıkarıldı :)



## Sonuç

Workout\_Frequency (days/week)

ve Experience\_Level

değişkenleri artık kategorikler sınıfına dahil edildi.



## Boş ve Ayrıntılı Değerler Var Mı?

Boş değerler, unique değerler ve ayrıntılı değerler analiz edildi. Görselleştirmeler yapıldı.



## Sonuç

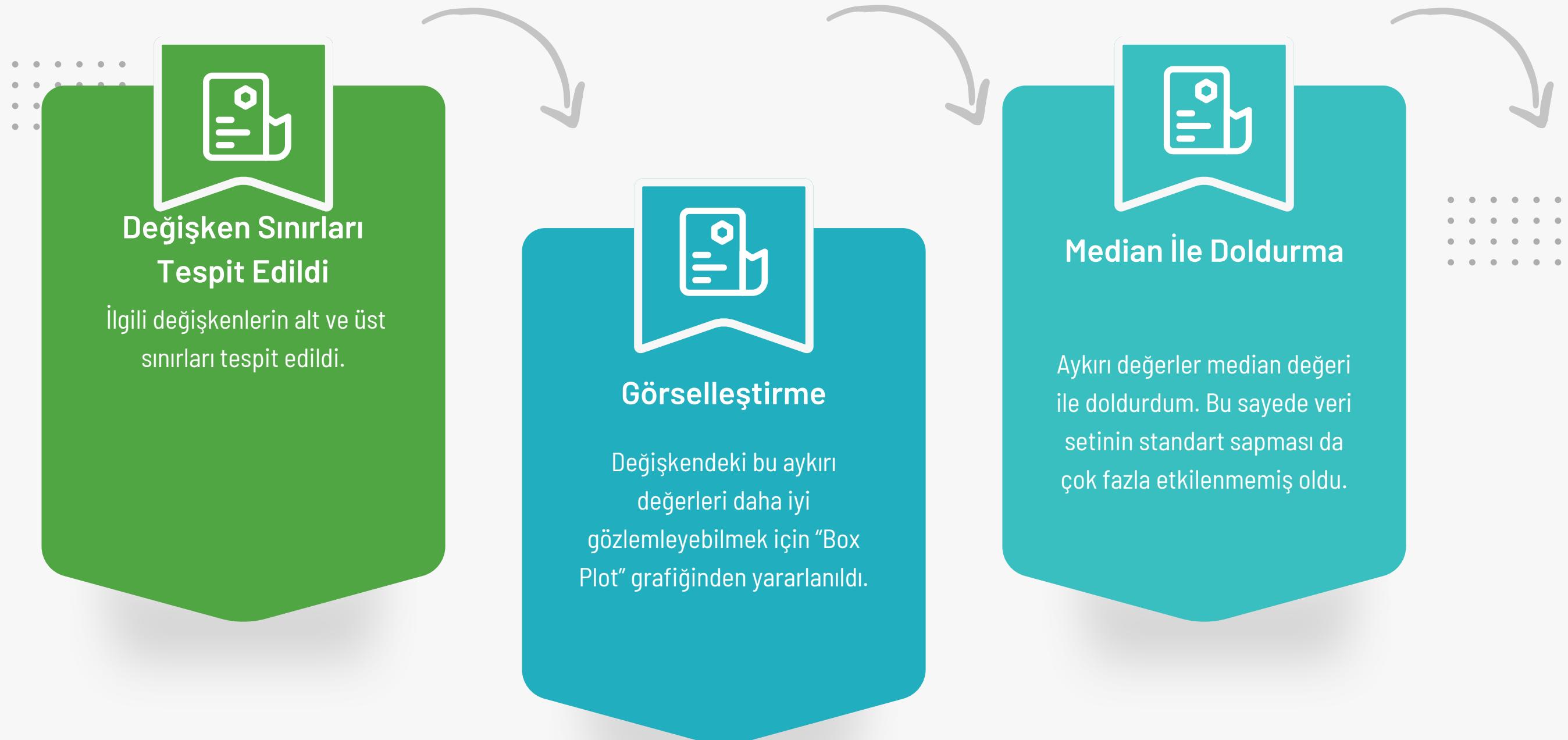
Veri setinde boş değer tespit edilmedi.

Weight değişkeninde 9

Calories\_Burned değişkeninde 10

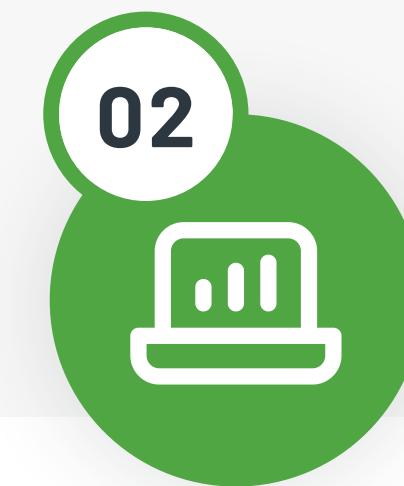
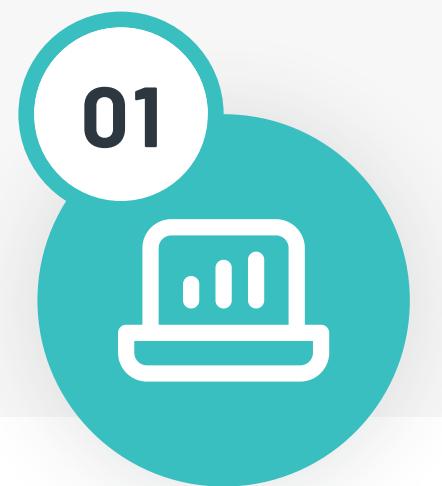
BMI değişkeninde 25 ayrıntılı değer tespit edildi.

# AYKIRI DEĞERLERE YAKLAŞIM



# KATEGORİK DEĞİŞKENLER İÇİN LABEL ENCODER

2 Farklı kategorik değişkenim var. Bunlar için label encoder kullandım. Bu işlemi direkt olarak LabelEncoder kütüphanesi ile yaptım. Alfabetik bir sıralama yapıldı.



Gender

MALE

FEMALE

Encoding

# Male ->1

# Female ->0

Workout\_Type

Cardio

HIIT

Strength

Yoga

Encdoing

# Cardio -> 0

# HIIT -> 1

# Strength -> 2

# Yoga ->3

# HESAPLANAN DEĞİŞKENLER



## Değişken Analizi

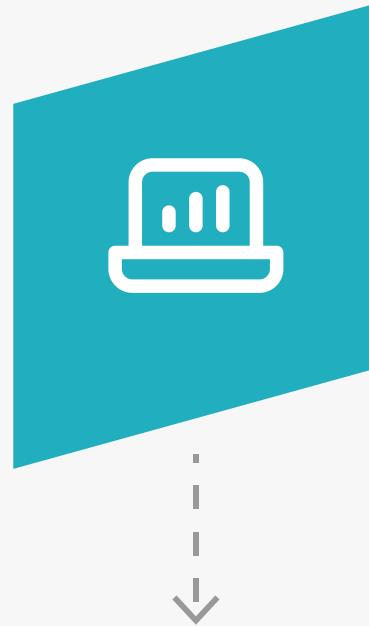
Veri setindeki BMI ve Fat\_Percentage değerleri belirli bir hesaplama ile bulunuyor. Bu değerlerin korelasyon analizleri yapıldı. Örnek bir modelde testleri yapıldı ve modele dahil edilmelerine karar verildi.

$$\text{BMI} = \text{Height (m)}^2 / \text{Weight (kg)}$$

$$\text{Fat Percentage} = 1.20 \times \text{BMI} + 0.23 \times \text{Age} - 16.2 \text{ (Erkek)}$$
$$\text{Fat Percentage} = 1.20 \times \text{BMI} + 0.23 \times \text{Age} - 5.4 \text{ (Kadın)}$$

# MAKİNE ÖĞRENMESİ - CROSS VALIDATION

Bu kısımda CrossValidation yaparak 3 farklı model oluşturarak bu modellerin başarısını karşılaştırdım. Satır sayısının çok fazla olmaması ve daha basit bir model çıkarmak istediğim için "Linear Regression", Birbiri ile ilişkili ve karmaşık veriler olduğu için "KNN" ve Parametre optimizasyonu ile iyi bir sonuç alacağımı düşündüğüm için "Gradient Boosting" algoritmalarını kullandım.



## Linear Regression

Linear Regression RMSE: 75.44

### Parametre Optimizasyonu Sonrası

Linear Regression RMSE: 76.46

**R2 Değeri: 0.90**



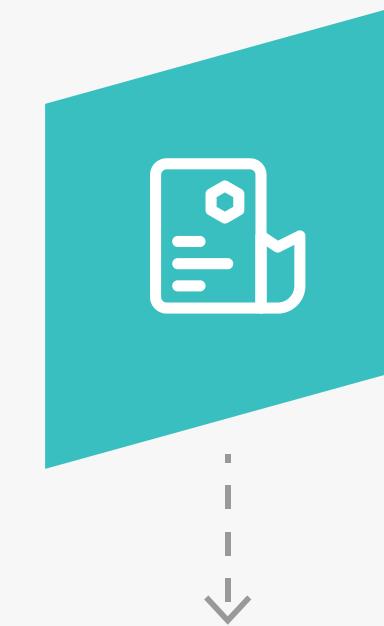
## SVM

KNN RMSE: 68.45

Parametre Optimizasyonu Sonrası

KNN RMSE: 66.50

**R2 Değeri: 0.75**



## Gradient Boosting

Gradient Boosting RMSE: 59,18

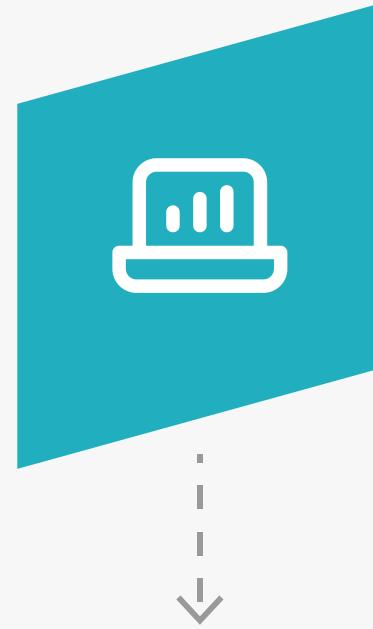
Parametre Optimizasyonu Sonrası

Gradient Boosting RMSE: 59,27

**R2 Değeri: 0.93**

# MAKİNE ÖĞRENMESİ - STANDART SCALER

Veri setimde değer aralıkları çok az olan ve çok büyük olan sayısal değerler var. Örneğin cinsiyet, yakılan kalori, ağırlık vb. değerlerinin aralıkları çok bağımsız ve farklı. Önceki modelimde de RMSE ve R2 değeri çok yüksek çıktıgı için veri setini normalize etmeye karar verdim ve böyle bir model oluşturdum.



## Linear Regression

Linear Regression RMSE: 0.36

Parametre Optimizasyonu Sonrası

Linear Regression RMSE: 0.332

**R2 Değeri:0.89**



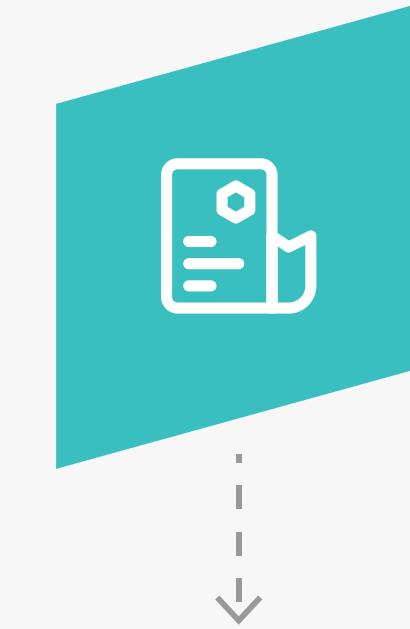
## SVM

KNN RMSE: 0.31

Parametre Optimizasyonu Sonrası

KNN RMSE: 0.3090

**R2 Değeri:0.74**



## Gradient Boosting

Gradient Boosting RMSE: 0.282

Parametre Optimizasyonu Sonrası

Gradient Boosting RMSE: 0.262

**R2 Değeri:0.92**

# KAPANIŞ VE ÖZET

## Ölçeklenebilirlik

Proje kapsamında model verileri kullanıcidan manuel alınabilir veya bir akıllı saat verisinden toplanan veriler ile genişletilebilir.



## Farklı Veri Setleri ile Kullanım

Model sonucunda yakılan kalori miktarını tahminlemektedir. Bu kısımda bir diyet programı, kalori açığı takibi, farklı bir egzersiz veri seti ile kilo verme odaklı egzersiz planlaması gibi modellere yaygınlaştırılabilir.



## Model Başarısı

Eğitim süresinde en başarılı sonuç 5 katmanlı CrossValidation uygulanmış "Gradient Boosting" algoritması ile elde edildi

## Modelin Kullanılabilirliği

Modelin kullanımı için streamlit üzerine bir arayüz hazırlandı ve kullanıcidan alınacak veriler ile tahminleme yapılması sağlandı.