****

**AD SOYAD: SELİN ESİN KÖKÇE**

**NUMARA: 201713709025**

**Cinsiyet Sınıflandırma Projesi**

**Bu projemde saç uzunluğu, burun ve alın genişliği gibi yüz yapısı özelliklerini içeren dataseti kullanarak cinsiyet tahmini yaptırdım. Projemi Jupyter Notebook üzerinden python programlama dili ile geliştirdim.**

**Projede kullandığım veri seti:** gender\_classification\_v7.csv

**Projede kullandığım algoritmalar:**

Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM),  
Naive Bayes, Decision Tree (Karar Ağacı), Random Forest Classification

**Bu veri seti 8 sütundan oluşmaktadır. Veri setinde 5000 satır, 8 sütun vardır ve boş değer yoktur.**

**long hair:** Boolean özellik. Eğer saç uzun ise 1, değilse 0.  
**forehead\_width\_cm:** Alın genişliğinin cm türünden ölçüsü.  
**forehead\_height\_cm:** Alın uzunluğunun cm türünden ölçüsü.  
**nose\_wide:** Boolean özellik. Eğer burun geniş ise 1, ince ise 0.  
**nose\_long:** Boolean özellik. Eğer burun uzun ise 1, kısa ise 0.  
**lips\_thin:** Boolean özellik. Eğer dudak ince ise 1, kalın ise 0.  
**distance\_nose\_to\_lip\_long:** boolean özellik. Eğer dudak ile burun arasındaki mesafe kısa ise 0, uzun ise 1.  
**gender:** Erkek veya kadın

**Not:** Cm türünden değerler popülasyondaki ilgili parametrelerin ortalama hesaplamasından sonra tanımlanacaktır.

Veri setine genel bakış:

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

## **Cinsiyet**

Projemde tahmin ettirmek istediğim değer cinsiyettir. Her bir cinsiyetten kaçar tane örnek olduğunu görelim.



## **Cinsiyete Göre Ortalama Bulma**

Tahmin yaparken özelliklerin ortalama değerleri baz alınacaktır. Bu kodda ortalama değerleri iki cinsiyet için de görebiliriz:

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

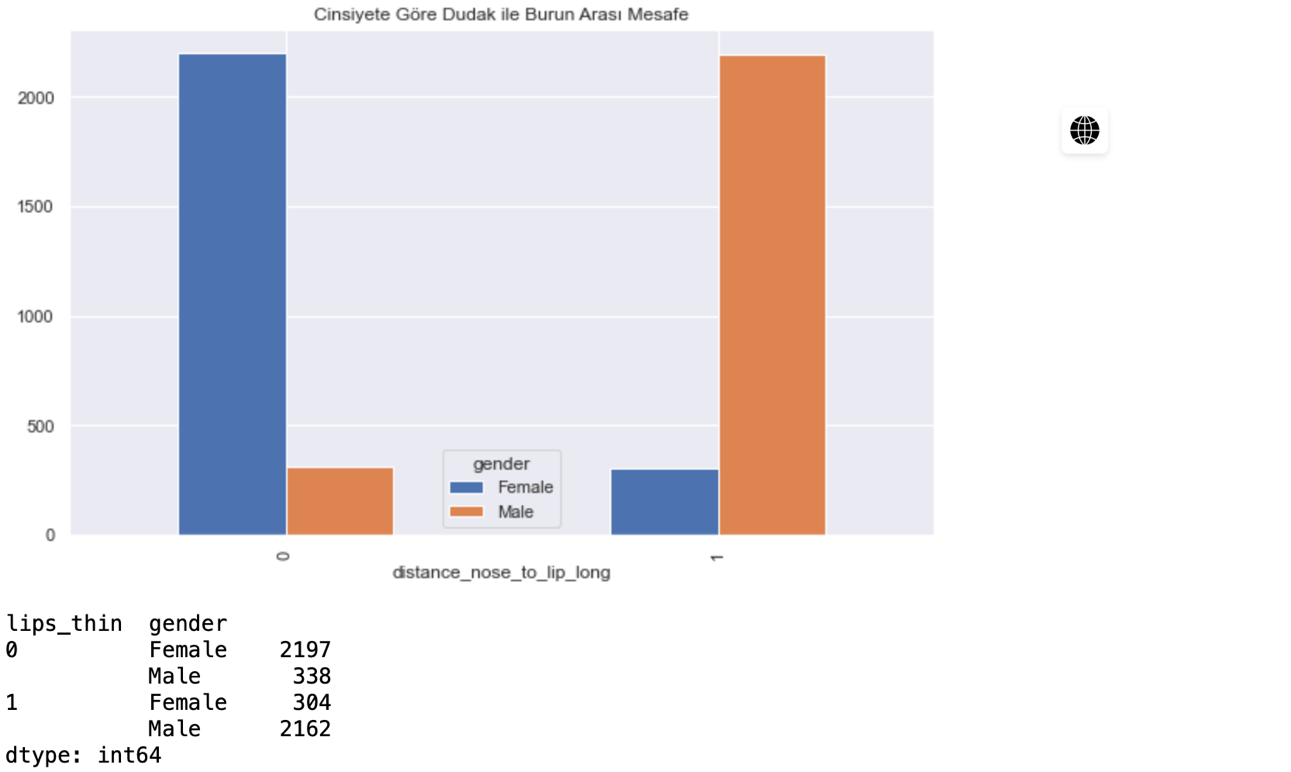
## **Görselleştirme**

Matplotlib kullanarak özellikleri görselleştirelim.



**Tüm özelliklerin (saç uzunluğu hariç) erkeklerde kadınlara göre daha yüksek değerler aldığını görebiliriz. Daha detaylı incelersek:**

### Dudak Kalınlığının ve Dudak-Burun Mesafesinin Cinsiyete Göre Dağılımı:[¶](#Dudak-Kalınlığının-ve-Dudak-Burun-Mesaf)



tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

\* **lips\_thin:** Boolean özellik. Eğer dudak ince ise 1, kalın ise 0.

**Tablolara bakarak kadınların erkeklere göre daha geniş dudaklara ve daha kısa burun-dudak arası mesafeye sahip olduğunu görebiliriz.**

## **Alın Genişliği ve Uzunluğunun Cinsiyete Göre Dağılımı:**



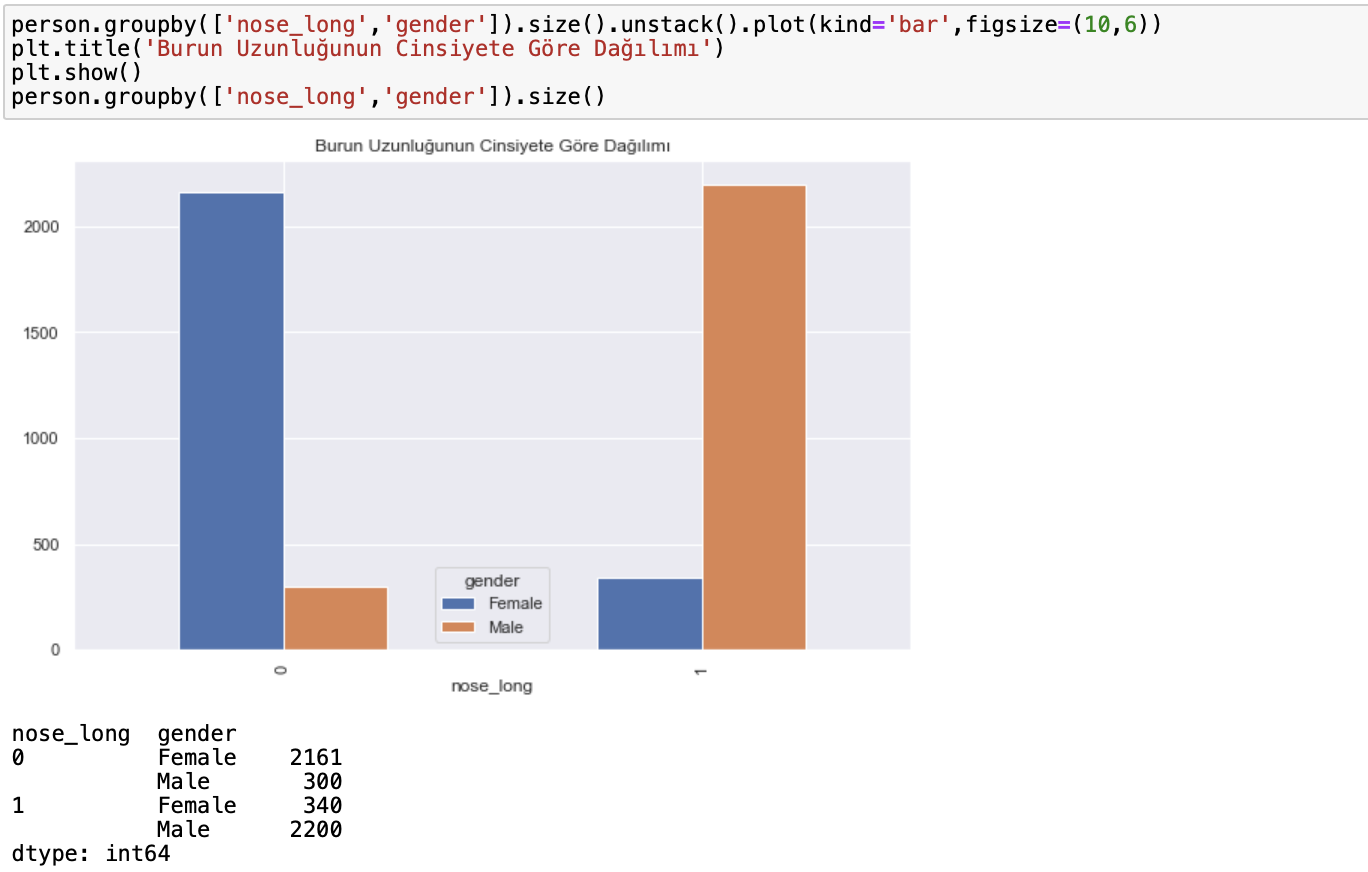
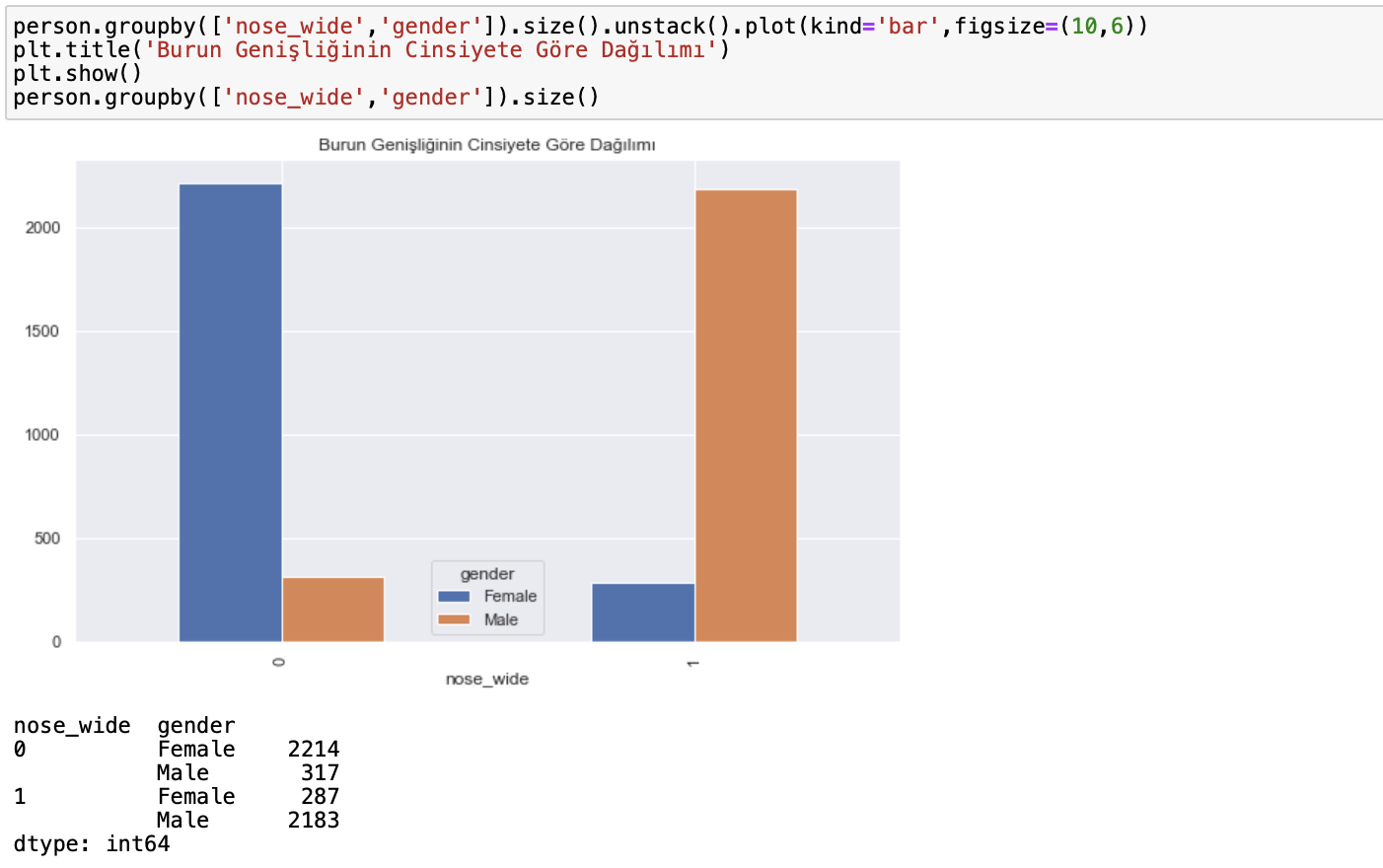
* Erkeklerin kadınlara göre daha geniş ve uzun alına sahip olduğunu görebiliriz.

## **Saç Uzunluğunun Cinsiyete Göre Dağılımı:**



* Kadınlar ile erkekler arasında saç uzunluğu dağılımının neredeyse aynı olduğunu görebiliriz.

## **Burun Genişliği ve Uzunluğunun Cinsiyete Göre Dağılımı:**



* Kadınların erkeklere göre daha ince ve kısa buruna sahip olduğunu görebiliriz.

### Veri setinde 'gender' özelliği string türündedir. Float değerine de ihtiyacımız olacağından 'gender\_code' sütunu ekledim. Erkekler için 0, kadınlar için 1 değerini verdim.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

## **Train Test Split (Eğitim/Test Seti Ayırma)**

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

## **Logistic Regression**

metin içeren bir resim

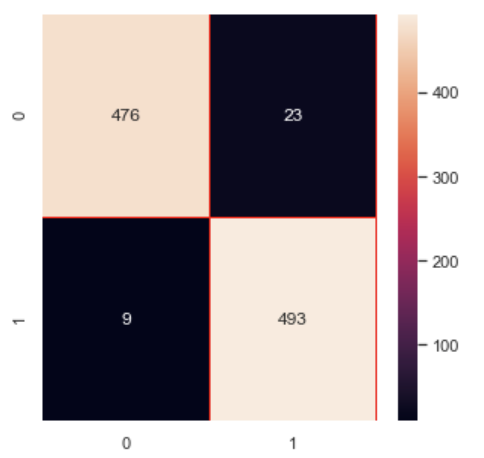
Açıklama otomatik olarak oluşturuldu

Projemde kullandığım algoritmalarda ilk olarak confusion matrix’in görselleştirmesini yaptım. Accuracy, precision, recall, f1-score, support değerlerini gösterdim. Model performansını değerlendirmek için MAE, MSE, RMSE ve Kappa Score değerlerini gösterdim.

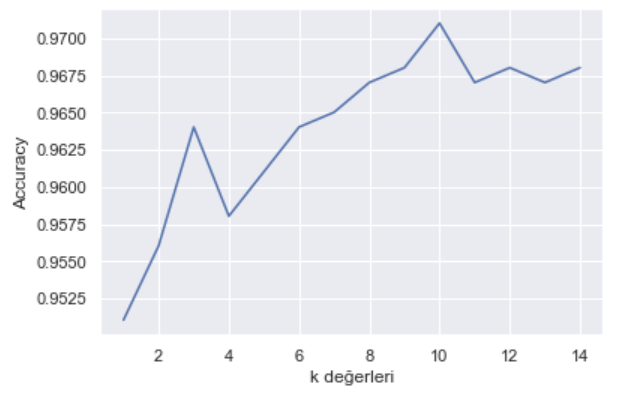
## **KNN Sınıflandırması**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

****

Her bir k değeri için accuracy değerini gösteren tablo:

****

**metin, makbuz içeren bir resim

Açıklama otomatik olarak oluşturuldu**

metin içeren bir resim

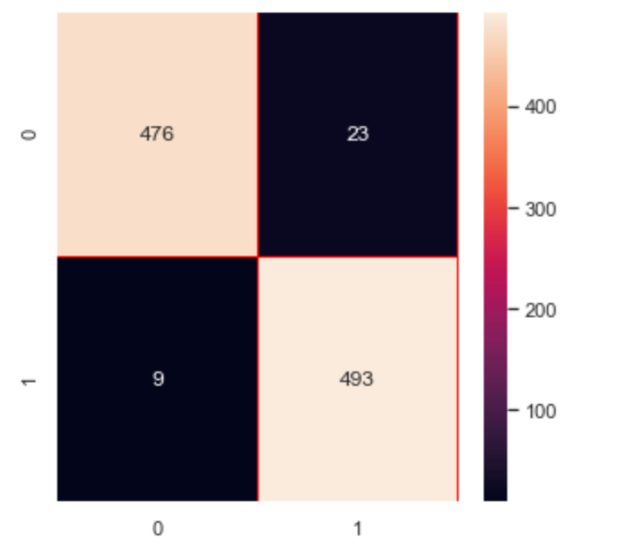
Açıklama otomatik olarak oluşturuldu

\*RMSE değerinin olabildiğince düşük olmasını isteriz. RMSE değeri düştükçe modelin tahmin yeteneği artar.

## **SVM (Support Vector Machine)**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

****

**metin, makbuz içeren bir resim

Açıklama otomatik olarak oluşturuldu**

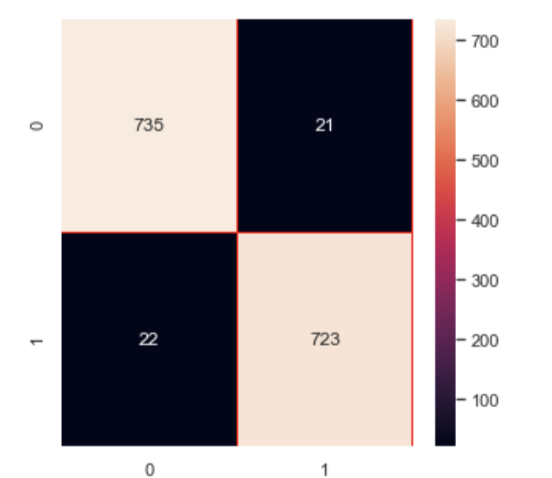
**metin içeren bir resim

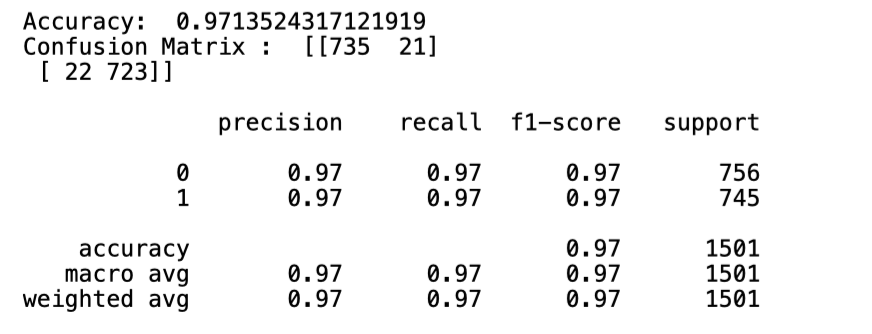
Açıklama otomatik olarak oluşturuldu**

**Naive Bayes**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

****

****

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

## **Decision Tree (Karar Ağacı)**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

****

**metin, makbuz içeren bir resim

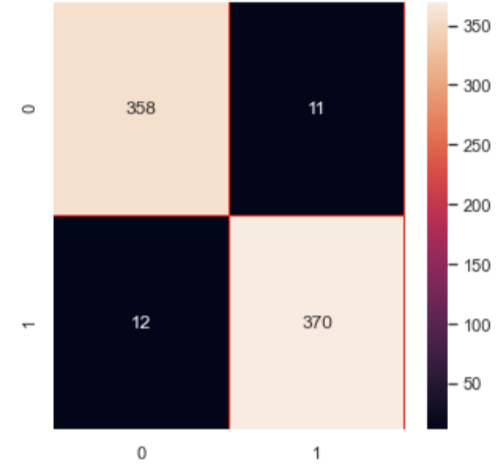
Açıklama otomatik olarak oluşturuldu**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

## Random Forest Classification

## 

****

**tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

## **Accuracy Değerlerinin Karşılaştırılması**[**¶**](#Accuracy-Değerlerinin-Karşılaştırılması)

**Logistic Regression için;**

**Test Accuracy:** 0.9675294117647059

**Train Accuracy:** 0.9707057256990679

**KNN için Accuracy:** 0.968031968031968

**Support Vector Machine (SVM) için Accuracy:** 0.9680426098535286

**Naive Bayes için Accuracy:** 0.9713524317121919

**Decision Tree (Karar Ağacı) için Accuracy:** 0.9693741677762983

**Random Forest Classification için Accuracy:** 0.966711051930759

En yüksek accuracy'nin Naive Bayes algoritması ile alındığı görülür. Naive Bayes kullanarak tahmin yaptırdığım kodlar :

## **Naive Bayes Algoritması ile Tahmin Ettirme**[**¶**](#Naive-Bayes-Algoritması-ile-Tahmin-Etti)

### Değerleri hatırlayalım:

"long\_hair","forehead\_width\_cm","forehead\_height\_cm", "nose\_wide","nose\_long","lips\_thin","distance\_nose\_to\_lip\_long"

Doğru tahmin yapıp yapmadığını kontrol edebilmek için veri setinde bulunan bir erkeğe ait değerleri girdim.

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

Girdiğim bu değerlere göre, algoritmanın cinsiyet tahmini sonucu erkek çıktı. Doğru çalışıyor.

**GUI (Arayüz)**

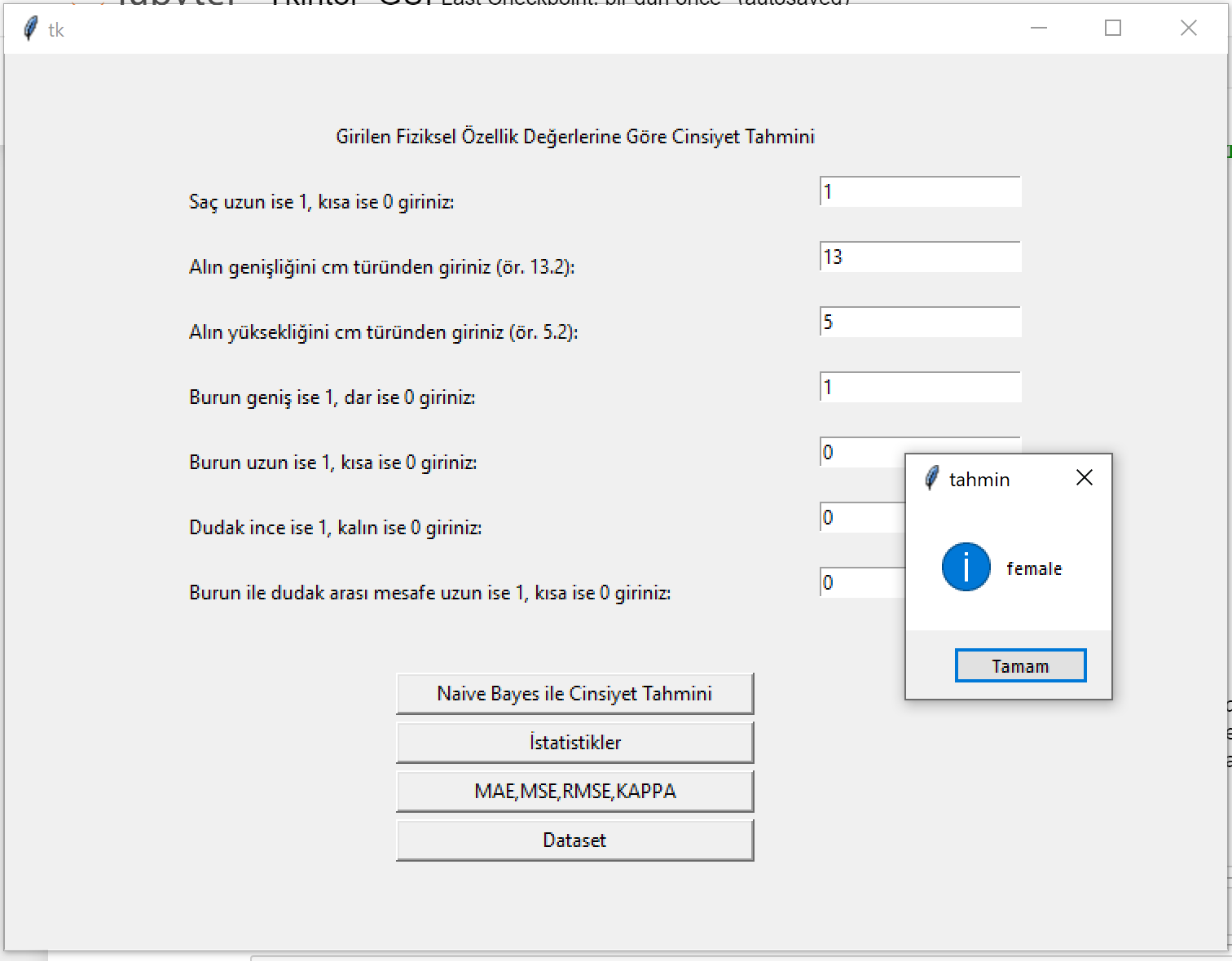
Uygulama arayüzü için Python programlama dili ile birlikte gelen grafiksel kullanıcı arayüzü aracıdır **Tkinter’i** kullandım.

Arayüzde kullanıcı girişi için entry alanları, cinsiyet tahmin butonu, istatistikleri gösteren buton, sırayla MAE,MSE,RMSE ve KAPPA’yı gösteren buton ve veri setinin genel görünümünü gösteren dataset butonu vardır.

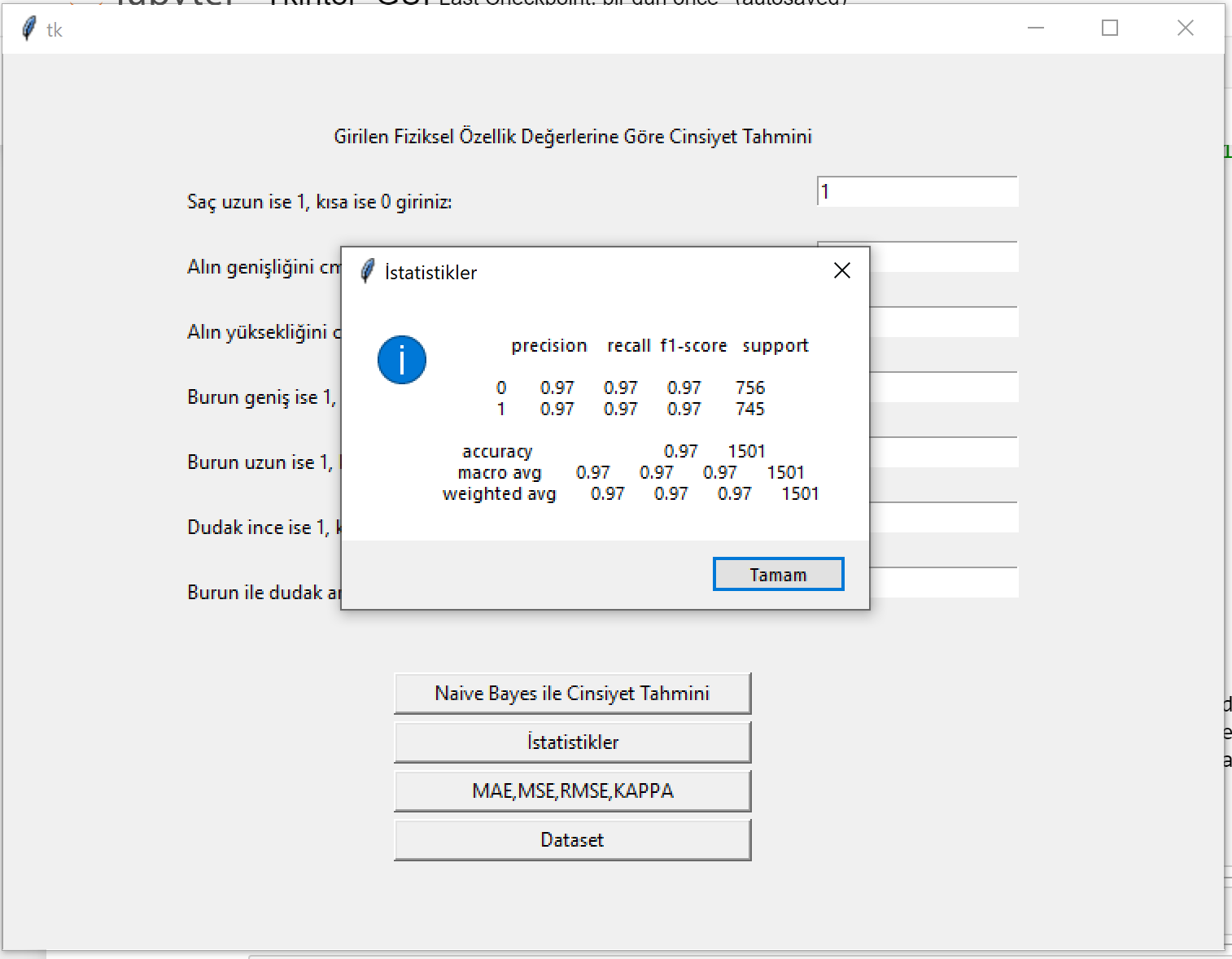
tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

Örnek cinsiyet tahmini sonucu:



Naive Bayes istatistik sonuçları:

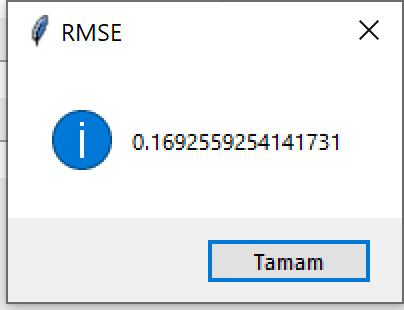


MAE, MSE, RMSE, KAPPA sonuçlarının ekranda gösterilmesi:

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

 metin içeren bir resim

Açıklama otomatik olarak oluşturuldu