



T.C
SAKARYA ÜNİVERSİTESİ
BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ BÖLÜMÜ
YAPAY ZEKAYA GİRİŞ
ÖDEVİ

HAZIRLAYAN

FURKAN TOP

B201200372

FARUK ERATA

B201200026

DANIŞMAN

Prof. Dr. İSMAİL HAKKI CEDİMOĞLU

05.01.2023

Proje Açıklama

Kanser hücrelerinin iyi huylu mu yoksa kötü huylu mu olduğunu tahmin edip sınıflandıran projedir.

Projede Kullanılan Yöntem, Algoritma ve Teknolojiler

Projemizi anaconda spyder python geliştirme ortamını hazırladık.

Veri setimizi kaggle sitesinden indirdik.

Aykırı durumları gözlemlemek için local outlier factor yöntemini kullandık.

Local outlier factor yöntemini verilerimiz secuve data olduğu için tercih ettik.

Girdiğimiz verilerle ne kadar doğru sonuç elde ettiğimizi öğrenmek için train test yöntemini kullandık.

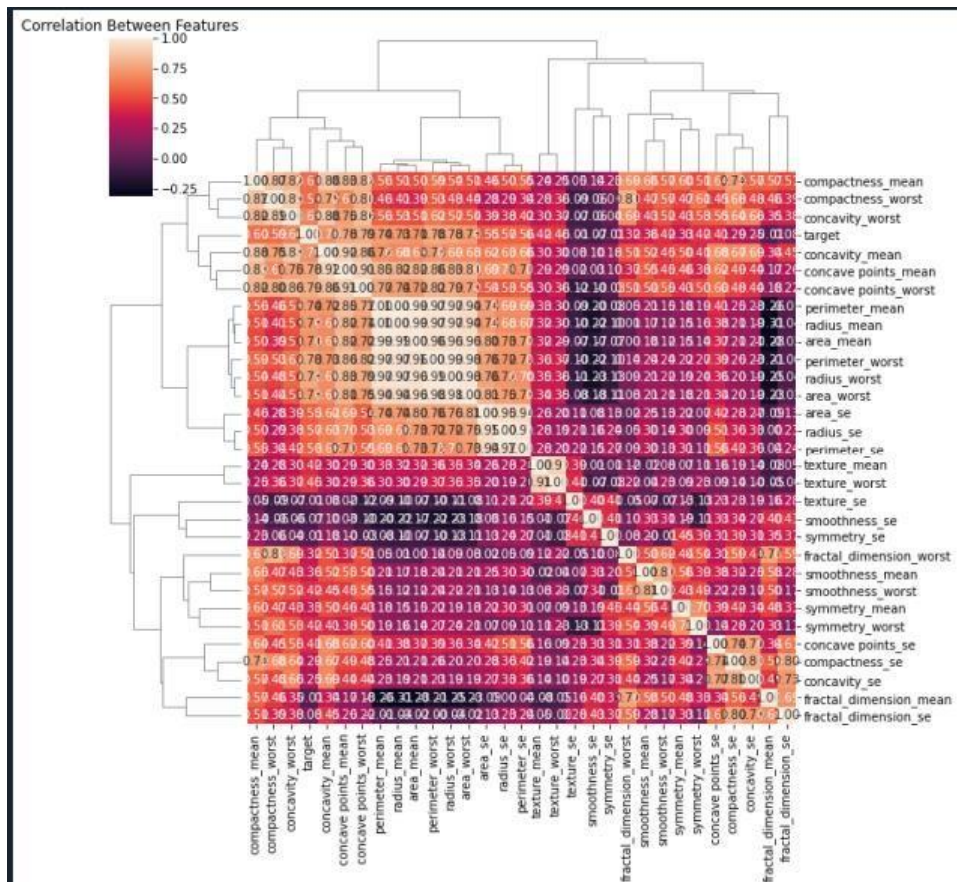
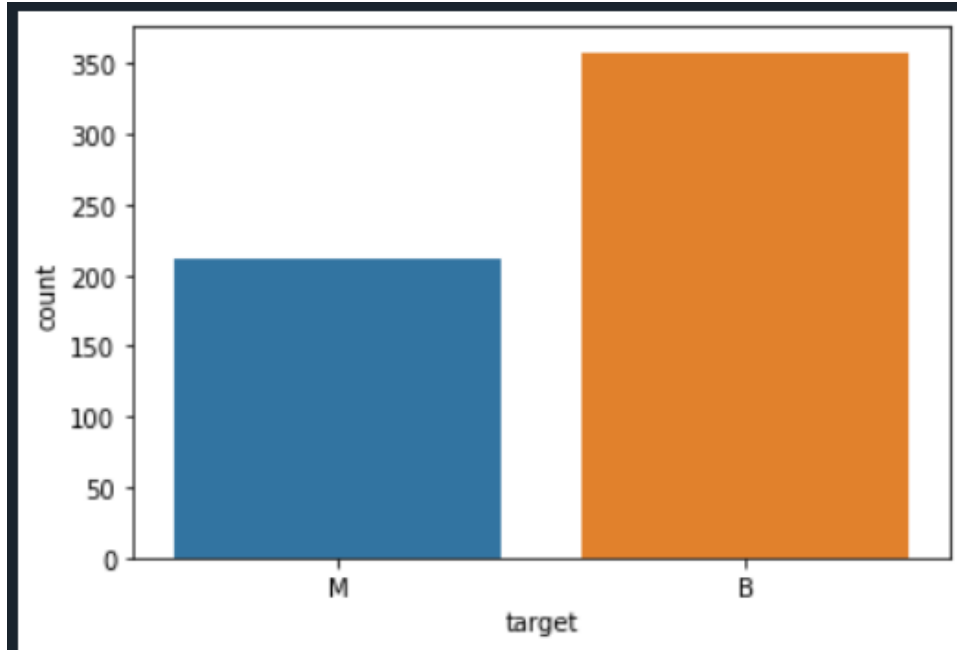
Aynı zamanda projemizde standardization ve normalization yaptık.

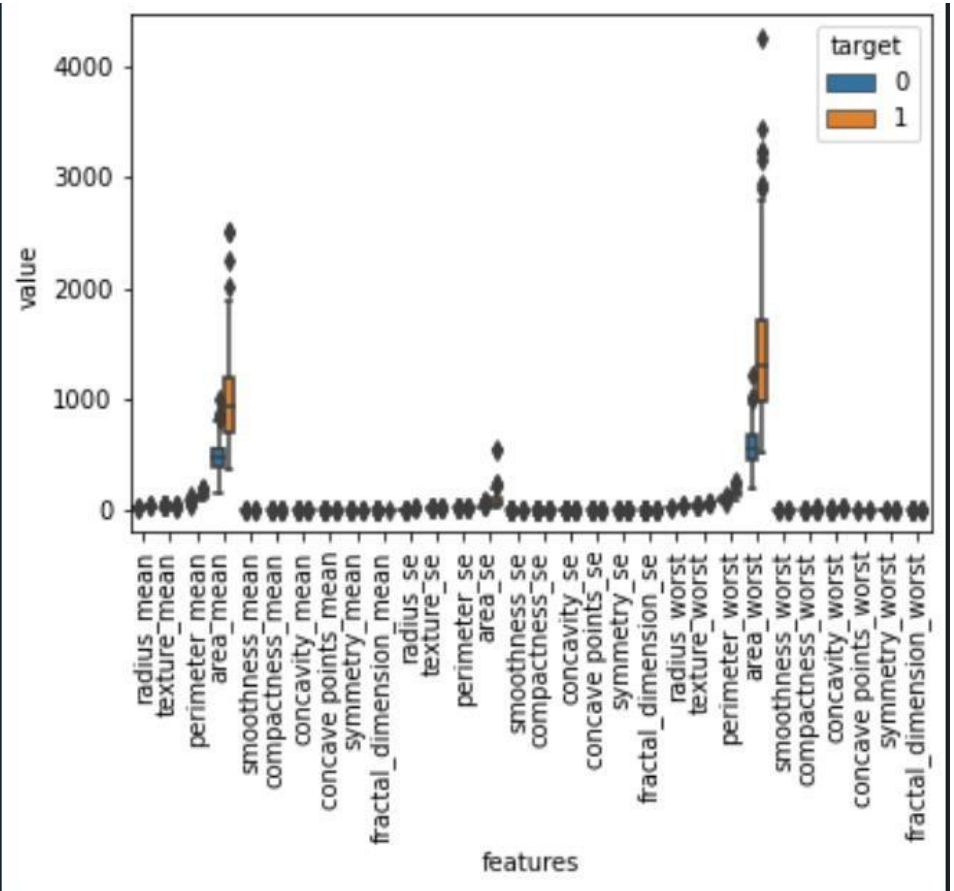
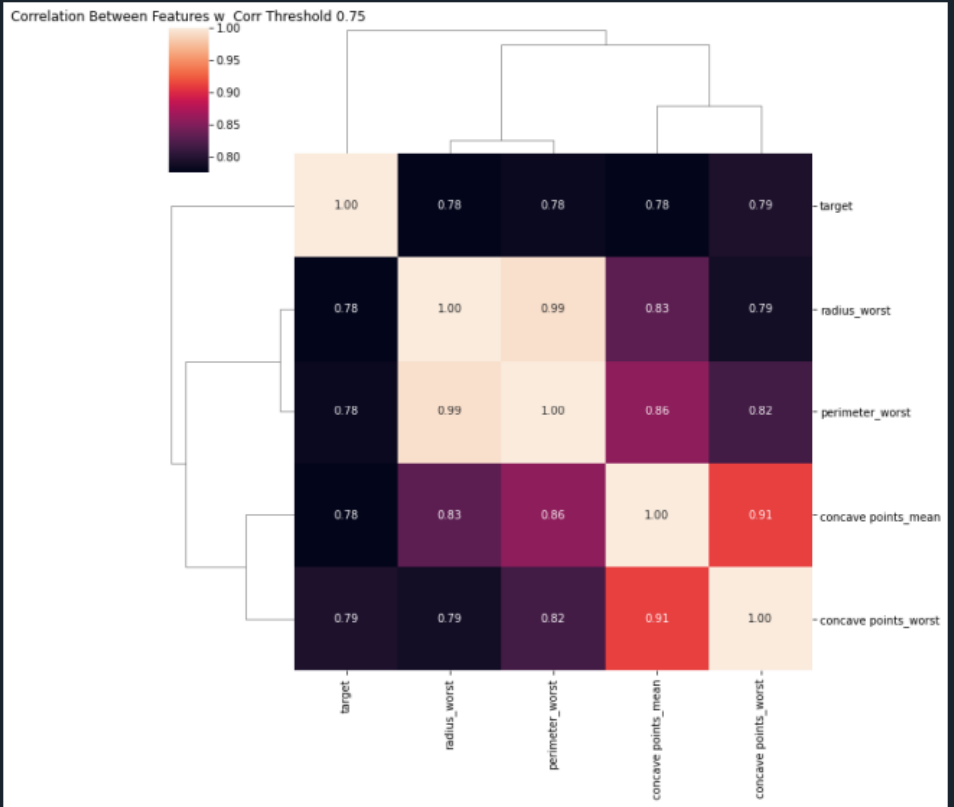
Sınıflandırma için KNN algoritmasını kullandık ve grid search cross validation yöntemini kullanarak knn algoritmasındaki en iyi parametreleri bulduk.

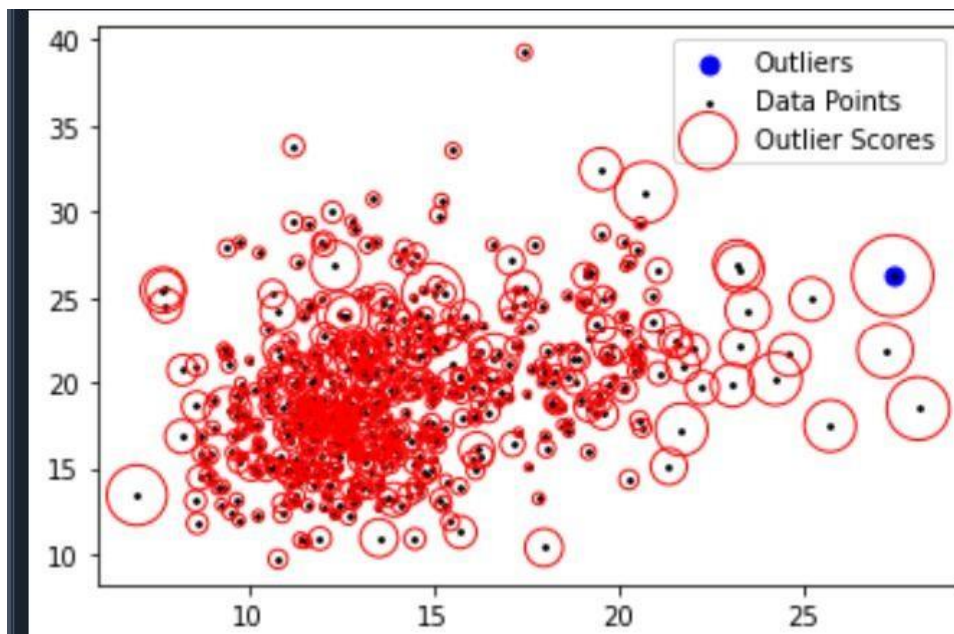
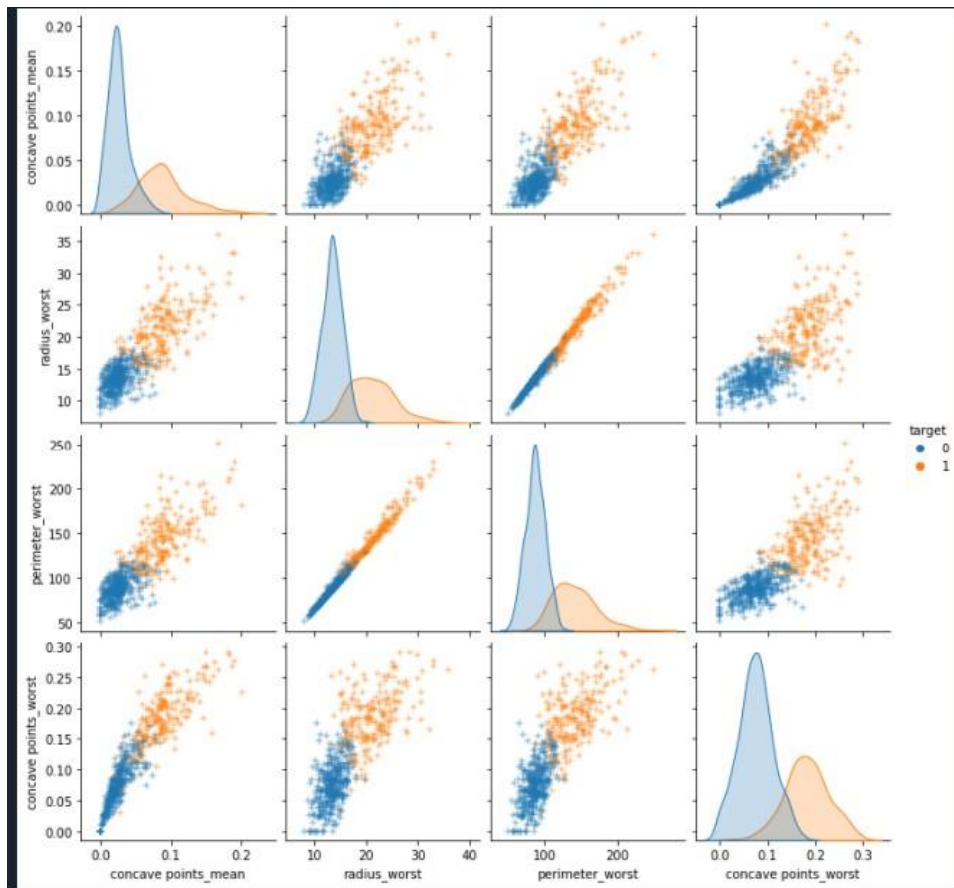
Veri boyutunu azaltmak için PCA yöntemini kullandık.

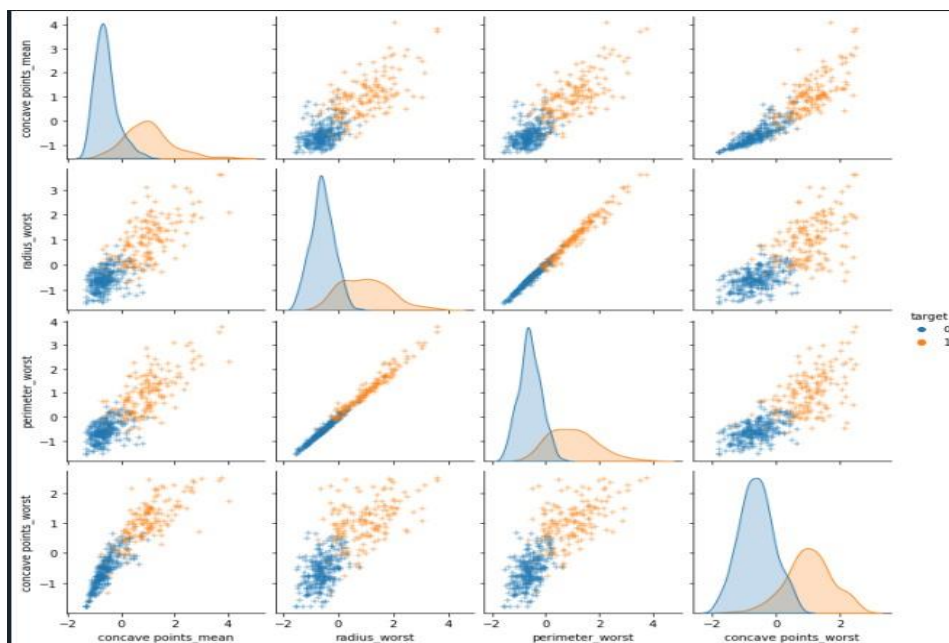
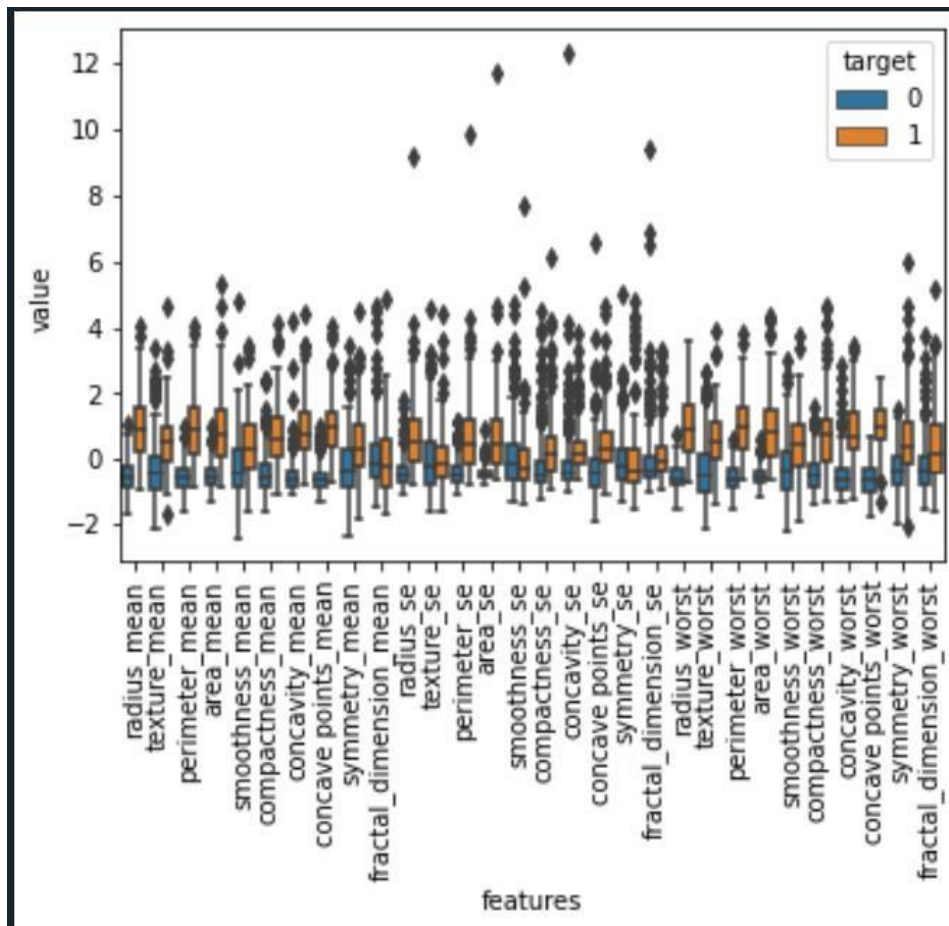
Yaptığımız tahminlerin başarısının artması için NCA yöntemini kullandık.

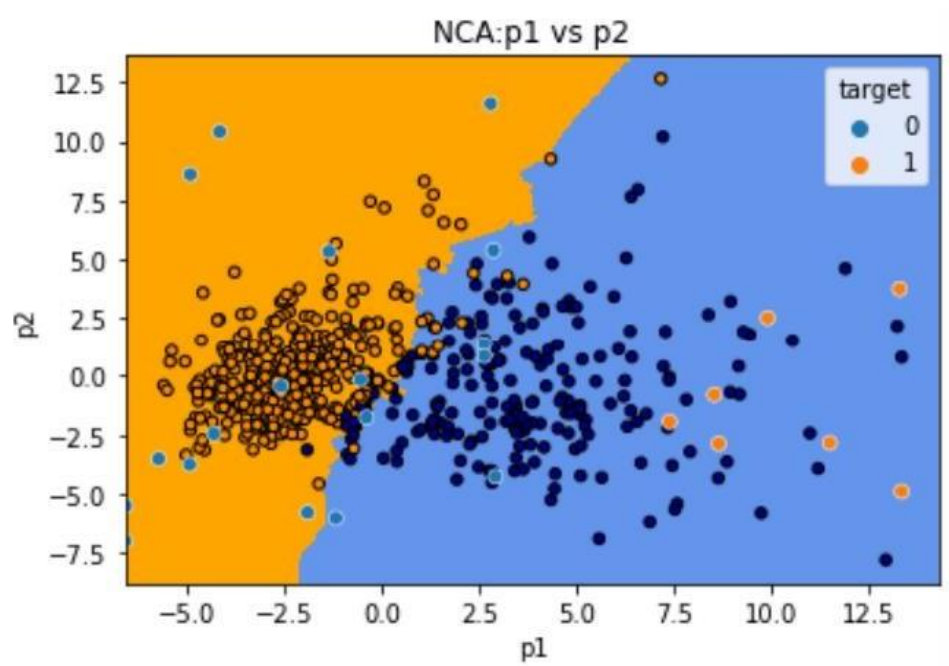
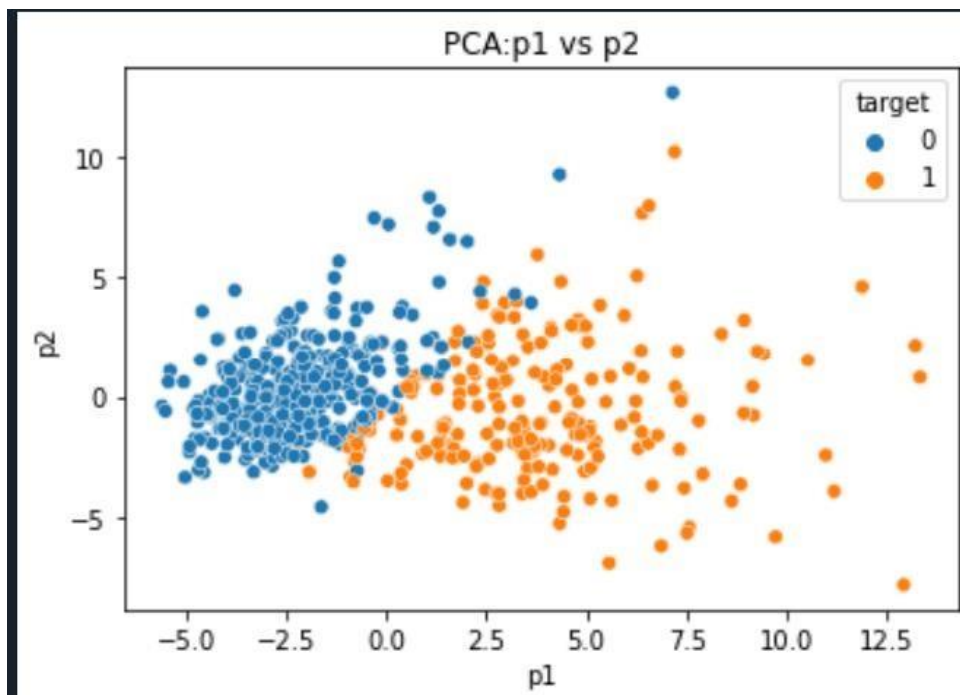
Proje Değerlendirme

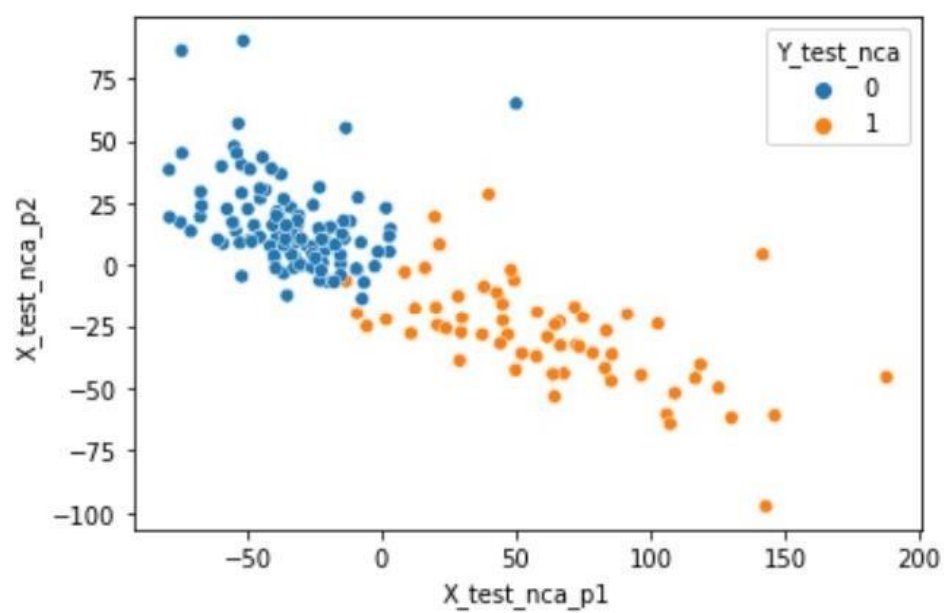
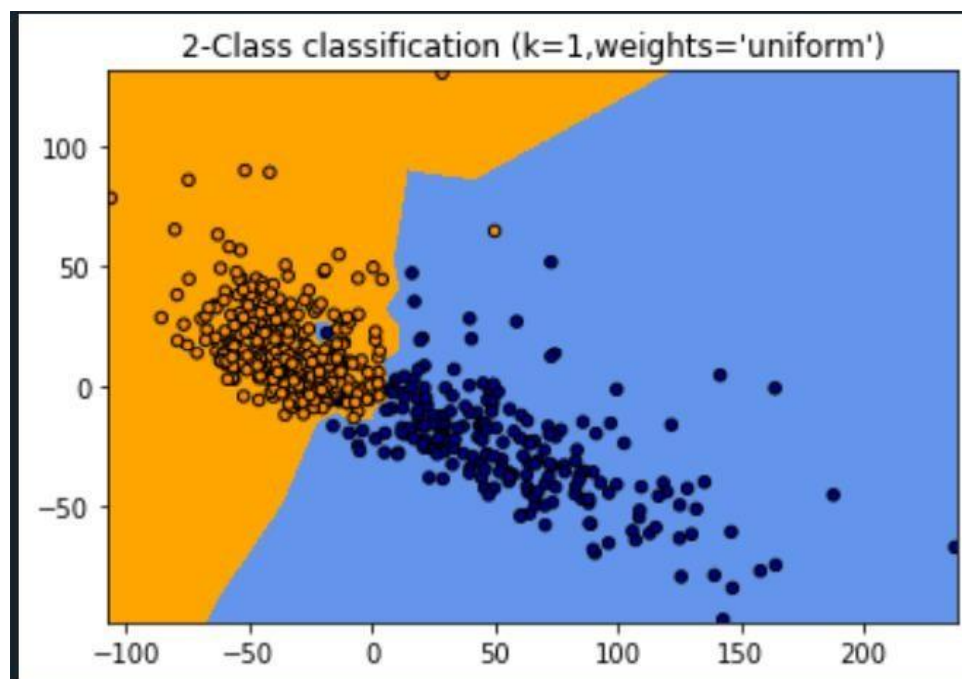













```
basic_knn_acc: 0.9532103742050039

Best training score: 0.9670512820512821 with parameters: {'n_neighbors': 4, 'weights':
'uniform'}

Test score:0.9590643274853801, Train score:0.9773299748110831

CM Test: [[107  2]
 [ 5 57]]
CM Train: [[248  0]
 [ 9 140]]

Best training score: 0.9419230769230769 with parameters: {'n_neighbors': 9, 'weights':
'uniform'}

Test score:0.9239766081871345, Train score:0.947103274559194

CM Test: [[103  6]
 [ 7 55]]
CM Train: [[241  7]
 [ 14 135]]

Best training score: 0.9873076923076922 with parameters: {'n_neighbors': 1, 'weights':
'uniform'}

Test score:0.9941520467836257, Train score:1.0

CM Test: [[108  1]
 [ 0 62]]
CM Train: [[248  0]
 [ 0 149]]
Traceback (most recent call last):

Python Console History
LSP Python: ready conda: base (Python 3.9.13) Line 194, Col 29 UTF-8 CRLF RW Mem 44%
```

```
Console 1/A X
0 target 569 non-null int64
1 radius_mean 569 non-null float64
2 texture_mean 569 non-null float64
3 perimeter_mean 569 non-null float64
4 area_mean 569 non-null float64
5 smoothness_mean 569 non-null float64
6 compactness_mean 569 non-null float64
7 concavity_mean 569 non-null float64
8 concave points_mean 569 non-null float64
9 symmetry_mean 569 non-null float64
10 fractal_dimension_mean 569 non-null float64
11 radius_se 569 non-null float64
12 texture_se 569 non-null float64
13 perimeter_se 569 non-null float64
14 area_se 569 non-null float64
15 smoothness_se 569 non-null float64
16 compactness_se 569 non-null float64
17 concavity_se 569 non-null float64
18 concave points_se 569 non-null float64
19 symmetry_se 569 non-null float64
20 fractal_dimension_se 569 non-null float64
21 radius_worst 569 non-null float64
22 texture_worst 569 non-null float64
23 perimeter_worst 569 non-null float64
24 area_worst 569 non-null float64
25 smoothness_worst 569 non-null float64
26 compactness_worst 569 non-null float64
27 concavity_worst 569 non-null float64
28 concave points_worst 569 non-null float64
29 symmetry_worst 569 non-null float64
30 fractal_dimension_worst 569 non-null float64
dtypes: float64(30), int64(1)

Python Console History
LSP Python: ready conda: base (Python 3.9.13) Line 194, Col 29 UTF-8 CRLF RW Mem 44%
```

Sonuç

Veri setimizi inceleyip problemimizi tanımladık.

Verilerimiz içinde bulunan aykırı durumları tespit edip verimizden çıkardık.

Verimizi eğitebilmek için train ve test split olarak ikiye bölüp standardization işlemini uyguladık.

Daha sonra basit bir KNN uygulaması yaptık ve %95 başarı sağladık.

PCA ile verimizi iki boyuta düşürdük fakat başarı oranı %92' ye kadar düştü.

NCA ile yine veriyi iki boyuta düşürdük ve tekrardan %99 başarı sağladık.

Tüm bunlar sonucunda kanser hücrelerinin iyi huylu mu yoksa kötü huylu mu olduğunu görselleştirme de yaparak ulaştık.

Kaynakça

<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>

<https://scikit-learn.org/stable/>

<https://www.udemy.com/course/python-ile-makine-ogrenmesi-yapay-zeka-projeleri-52/learn/lecture/18611120#reviews>