



Evin AYDİN ÜLGEN / 19360859066

# **KARAR AĞACI TABANLI MEME KANSERİ SINIFLANDIRMASI**

# İÇİNDEKİLER

---

Projenin Amacı ve Kapsamı

---

Literatür Taraması

---

Veri Seti

---

Veri Setinin Yapısal Analizi

---

Sınıflandırma Yöntemi: Decision Tree

---

Performans Metrikleri

---

Görselleştirme Araçları: Confusion Matris

---

Görselleştirme Araçları: ROC Eğrisi

---

Görselleştirme Araçları: Recall-Precision Eğrisi

---

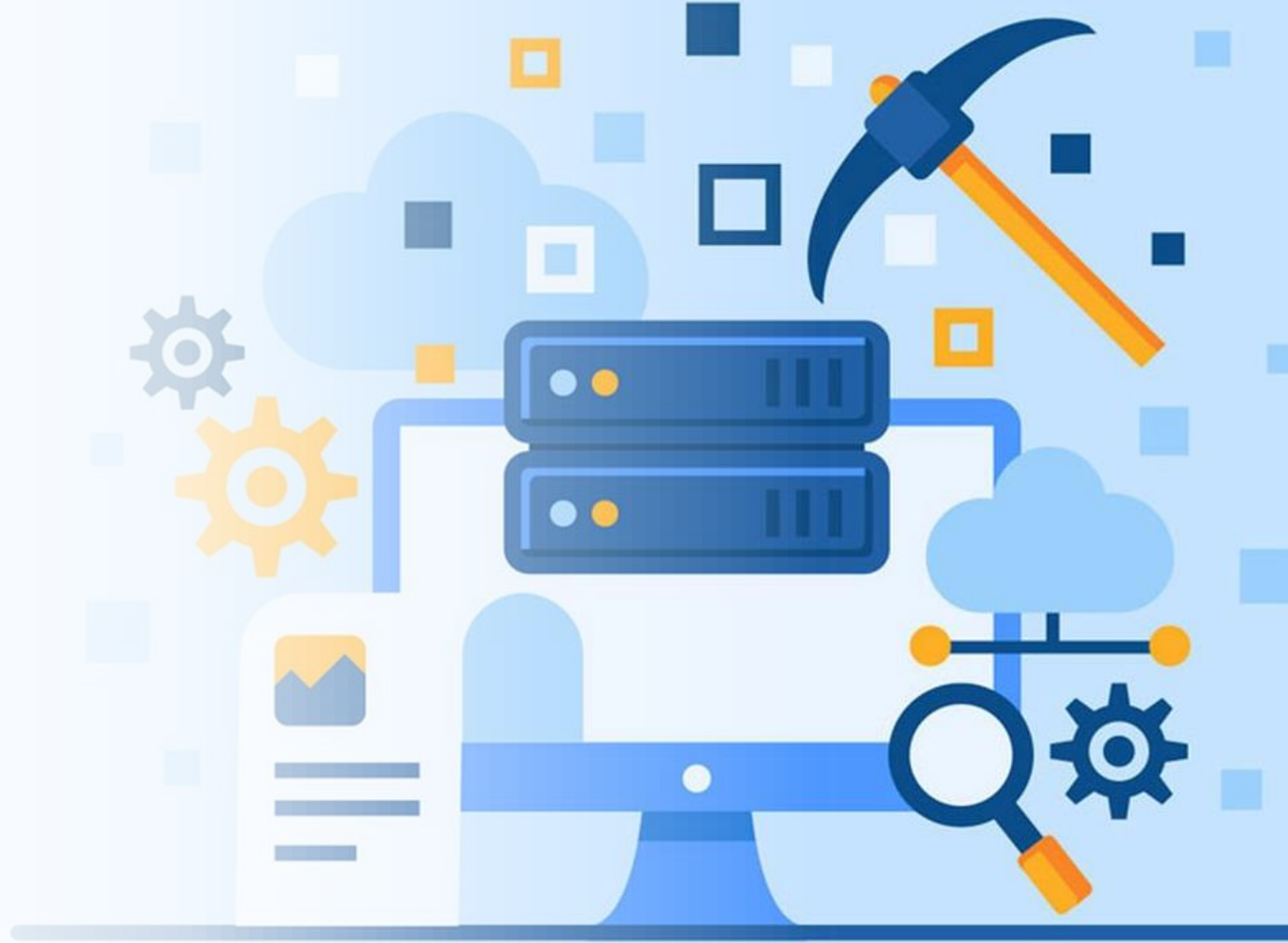
Sonuçların Karşılaştırılması

---

Kaynakça

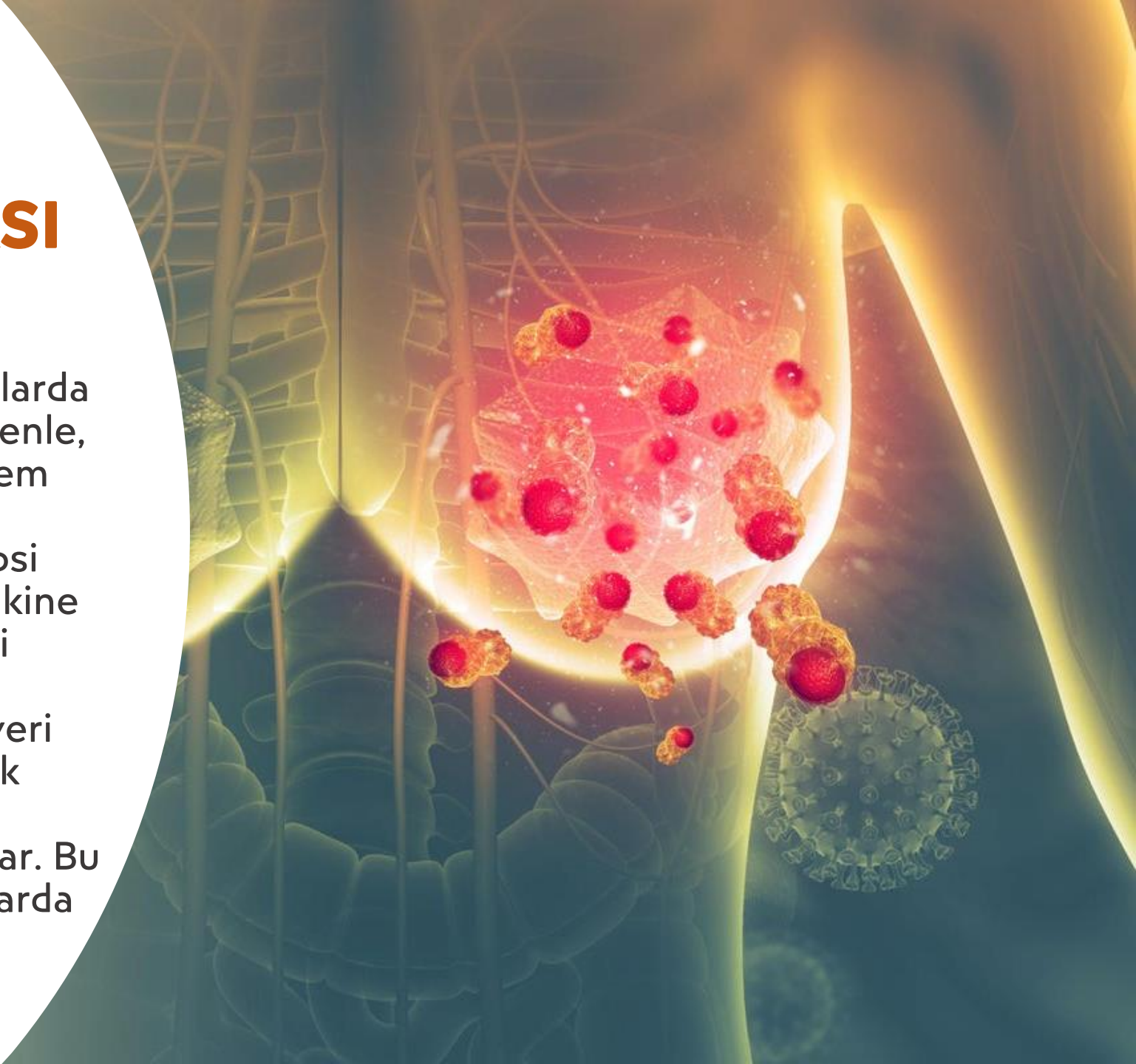
# PROJENİN AMACI VE KAPSAMI

- Araştırmanın konusu meme kanseri teşhisi için decision tree sınıflandırma yöntemini kullanarak veri analizi yapmaktır. Amacımız, meme kanseri teşhisinde en önemli faktörlerin belirlenmesi ve bu faktörlere dayalı olarak doğru teşhislerin yapılabilmesi için bir model geliştirmektir. Bu modelin başarısını, farklı performans ölçütleri kullanarak değerlendireceğiz ve sonuçlarımızı referans bir raporla karşılaştıracacağız.



# LİTERATÜR TARAMASI

- Meme kanseri, dünya genelinde kadınlarda en sık görülen kanser türüdür. Bu nedenle, meme kanseri teşhisi için birçok yöntem geliştirilmiştir. Geleneksel yöntemler arasında mamografi, ultrason ve biyopsi bulunmaktadır. Ancak, son yıllarda makine öğrenimi yöntemleri de meme kanseri teşhisi için kullanılmaya başlanmıştır. Decision tree sınıflandırma yöntemi, veri集中的 özelliklerin etkisi incelenerek meme kanseri teşhisinde en önemli faktörlerin belirlenmesine olanak sağlar. Bu yöntemin, daha önce yapılan çalışmalarda da etkinliği kanıtlanmıştır.





Class, Age, Menopause, Tumor Size, Inv Nodes, Node Caps, Deg Malig, Breast, Breast Quad, Irradiat

no-recurrence-events, 30-39, premeno, 30-34, 0-2, no, 3, left, left\_low, no  
no-recurrence-events, 40-49, premeno, 20-24, 0-2, no, 2, right, right\_up, no  
no-recurrence-events, 40-49, premeno, 20-24, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 60-69, ge40, 15-19, 0-2, no, 2, right, left\_up, no  
no-recurrence-events, 40-49, premeno, 0-4, 0-2, no, 2, right, right\_low, no  
no-recurrence-events, 60-69, ge40, 15-19, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 50-59, premeno, 25-29, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 60-69, ge40, 20-24, 0-2, no, 1, left, left\_low, no  
no-recurrence-events, 40-49, premeno, 50-54, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 40-49, premeno, 20-24, 0-2, no, 2, right, left\_up, no  
no-recurrence-events, 40-49, premeno, 0-4, 0-2, no, 3, left, central, no  
no-recurrence-events, 50-59, ge40, 25-29, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 60-69, lt40, 10-14, 0-2, no, 1, left, right\_up, no  
no-recurrence-events, 50-59, ge40, 25-29, 0-2, no, 3, left, right\_up, no  
no-recurrence-events, 40-49, premeno, 30-34, 0-2, no, 3, left, left\_up, no  
no-recurrence-events, 60-69, lt40, 30-34, 0-2, no, 1, left, left\_low, no  
no-recurrence-events, 40-49, premeno, 15-19, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 50-59, premeno, 30-34, 0-2, no, 3, left, left\_low, no  
no-recurrence-events, 60-69, ge40, 30-34, 0-2, no, 3, left, left\_low, no  
no-recurrence-events, 50-59, ge40, 30-34, 0-2, no, 1, right, right\_up, no  
no-recurrence-events, 50-59, ge40, 40-44, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 60-69, ge40, 15-19, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 30-39, premeno, 25-29, 0-2, no, 2, right, left\_low, no  
no-recurrence-events, 50-59, premeno, 40-44, 0-2, no, 2, left, left\_up, no  
no-recurrence-events, 50-59, premeno, 35-39, 0-2, no, 2, right, left\_up, no  
no-recurrence-events, 40-49, premeno, 25-29, 0-2, no, 2, left, left\_up, no  
no-recurrence-events, 50-59, premeno, 20-24, 0-2, no, 1, left, left\_low, no  
no-recurrence-events, 60-69, ge40, 25-29, 0-2, no, 3, right, left\_up, no  
no-recurrence-events, 40-49, premeno, 40-44, 0-2, no, 2, right, left\_low, no  
no-recurrence-events, 60-69, ge40, 30-34, 0-2, no, 2, left, left\_low, no  
no-recurrence-events, 50-59, ge40, 40-44, 0-2, no, 3, right, left\_up, no  
no-recurrence-events, 50-59, premeno, 15-19, 0-2, no, 2, right, left\_low, no  
no-recurrence-events, 50-59, premeno, 10-14, 0-2, no, 3, left, left\_low, no

# VERİ SETİ

- Breast Cancer veri seti, meme kanseri teşhisi konan hastaların klinik verilerini içeren bir veri setidir. Toplam 286 örnekten oluşmaktadır. Örneklerin 201'i tekrarlayan olayların olmadığı (no-recurrence-events), 85'i ise tekrarlayan olayların olduğu (recurrence-events) hastaları temsil etmektedir.
- Öznitelikler arasına yaş, menopoz durumu, tümör boyutu, invaziv nod sayısı, nodül kapsülü varlığı, tümörün derecesi, meme (sol veya sağ), meme çeyreği ve radyasyon tedavisi alma durumu yer almaktadır. Bu özniteliklerin bazıları sayısal, bazıları ise kategorik verileri temsil etmektedir.
- Bu veri seti, meme kanseri teşhisinde karar ağacı yöntemi gibi makine öğrenimi tekniklerinin kullanılması için yaygın bir veri setidir.

# VERİ SETİNİN YAPISAL ANALİZİ

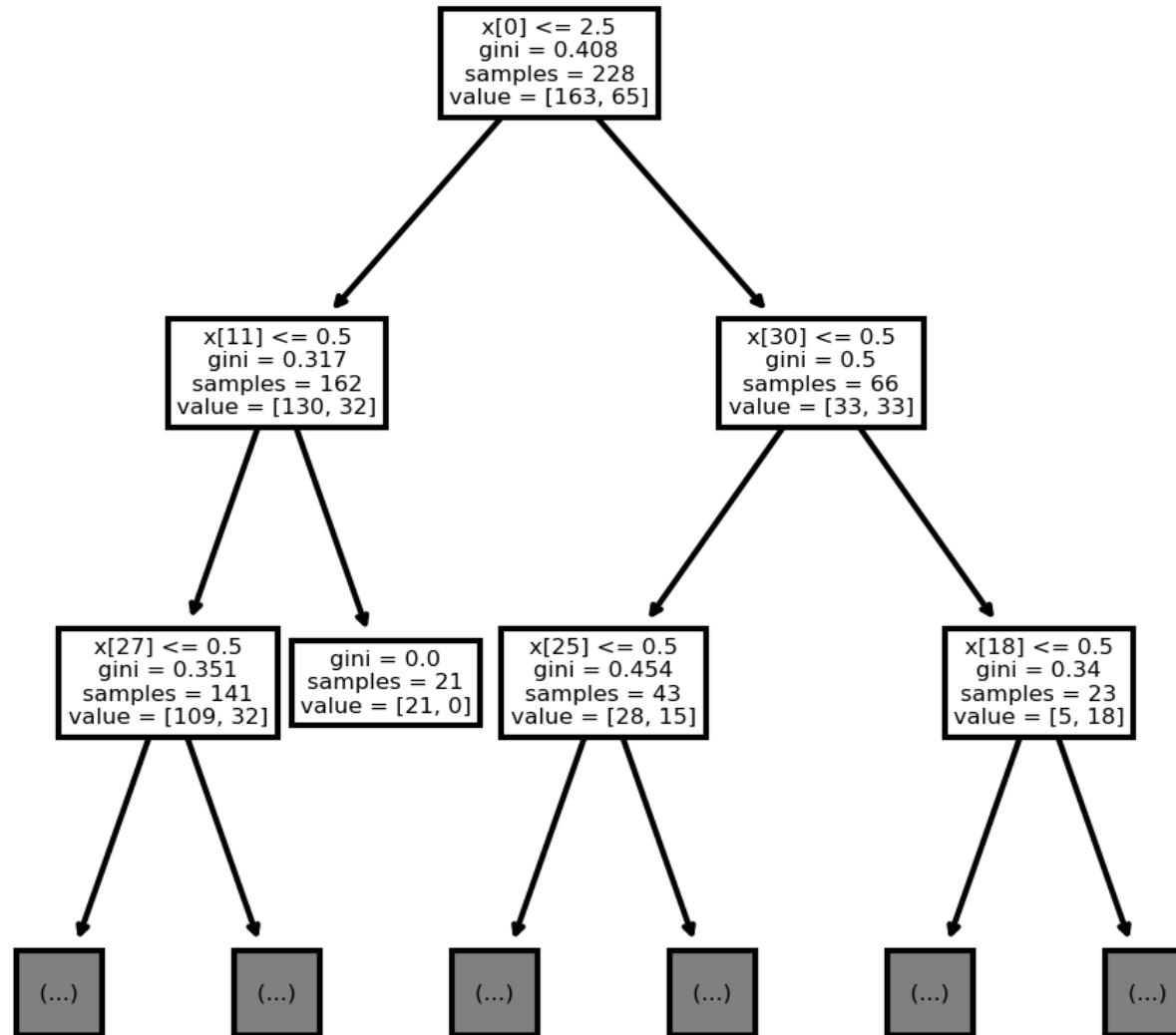
```
1 import pandas as pd
2
3 # Veri setini yükleme
4 data = pd.read_csv(r'C:\Users\90546\OneDrive\Masaüstü\veriMadenciligi\breast-cancer.csv', header=0)
5 # Veri setinin boyutunu (sıra, sütun) görüntüleme
6 print("Veri setinin boyutu:", data.shape)
7 # Sütun adlarını görüntüleme
8 print("Sütun adları:", data.columns)
9 # Sütunların veri tiplerini göster
10 print("Veri tipleri:\n", data.dtypes)
11 # Her sütundaki eksik değer sayısını kontrol etme
12 print("Sütunlardaki eksik değer sayısı:\n", data.isnull().sum())
13 # Sütunlardaki benzersiz değerleri göster
14 for column in data.columns:
15     unique_values = data[column].unique()
16     print("Benzersiz değerler:\n", f"{column}: {unique_values}")
17
```



- Karar ağacı yöntemi, veri setindeki özelliklerin ve sınıfların ilişkilerini anlamak için kullanılan bir yöntemdir. Bu yöntem, her özelliği ve sınıfı ağaç yapısı şeklinde temsil eder ve en önemli özellikleri belirlemek için kullanılır. Karar ağacı algoritması, tüm özelliklerin sınıflandırılmasına kadar devam eder ve özelliklerin sınıflandırma sonuçlarına göre sıralanarak en önemli özellikler belirlenir. Karar ağacı yöntemi, veri setindeki özelliklerin ilişkilerini görsel olarak ifade edebilir ve anlaşılması kolay bir modele sahiptir. Veri analizi ve sınıflandırma problemlerinde sıklıkla kullanılır.

```
29 import matplotlib.pyplot as plt
30 from sklearn import tree
31 from sklearn.tree import DecisionTreeClassifier
32 from sklearn.model_selection import train_test_split
33 from sklearn.preprocessing import LabelEncoder
34
35 # String değerleri sayısal değerlere dönüştürme
36 label_encoder = LabelEncoder()
37 y = label_encoder.fit_transform(data['Class'])
38 # Özniteliklerin dönüştürümü
39 X_encoded = pd.get_dummies(data.drop('Class', axis=1))
40 print("Öznitelikler:\n", X_encoded)
41 print("Hedef değişken y:\n", y)
42 # Eğitim ve test verisi olarak böl
43 X_train, X_test, y_train, y_test = train_test_split(X_encoded, y, test_size=0.2, random_state=0)
44 # Decision Tree modelini oluştur ve eğit
45 clf = DecisionTreeClassifier()
46 clf.fit(X_train, y_train)
47
48 # Decision Tree Çizimi
49 fig, axes = plt.subplots(nrows = 1,ncols = 1,figsize = (4,4), dpi=300)
50 tree.plot_tree(clf, max_depth = 2)
51 plt.savefig('decision_tree.png')
52
```





## DECISION TREE ÇIKTISI

```
46 # Test veri kümesi üzerinde tahmin yap
47 y_pred = clf.predict(X_test)
48
49 from sklearn.metrics import confusion_matrix
50 from sklearn.metrics import accuracy_score, confusion_matrix, classification_report, precision_score, recall_score
51 # Metrikleri hesapla
52 accuracy = accuracy_score(y_test, y_pred)
53 precision = precision_score(y_test, y_pred)
54 recall = recall_score(y_test, y_pred)
55 f1 = f1_score(y_test, y_pred)
56 report = classification_report(y_test, y_pred)
57
58 # Metrikleri yazdır
59 print("Accuracy:", accuracy)
60 print("Precision:", precision)
61 print("Recall:", recall)
62 print("F1-Score:", f1)
63 print("Classification Report:", report)
64
```

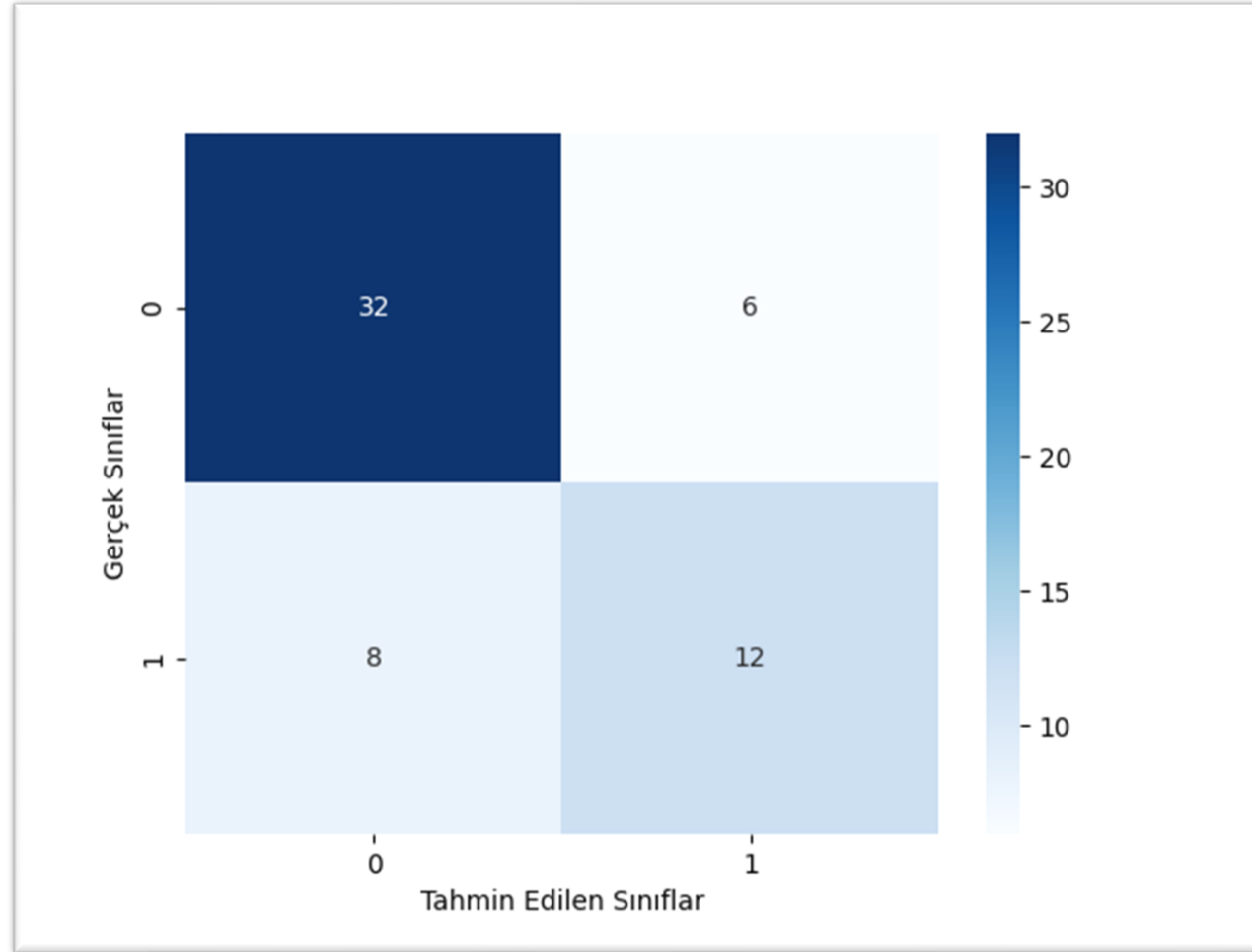
# PERFORMANS METRİKLERİ

---

Performans metrikleri , sınıflandırma modelinin performansını ölçmek için kullanılan ölçümlerdir. Örneğin, doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru, sınıflandırma modelinin performansını değerlendirmek için kullanılabilir. Bu metrikler, sınıflandırma modelinin ne kadar iyi çalıştığını ve hangi sınıfları daha iyi veya daha kötü sınıflandırdığını belirleyebilir

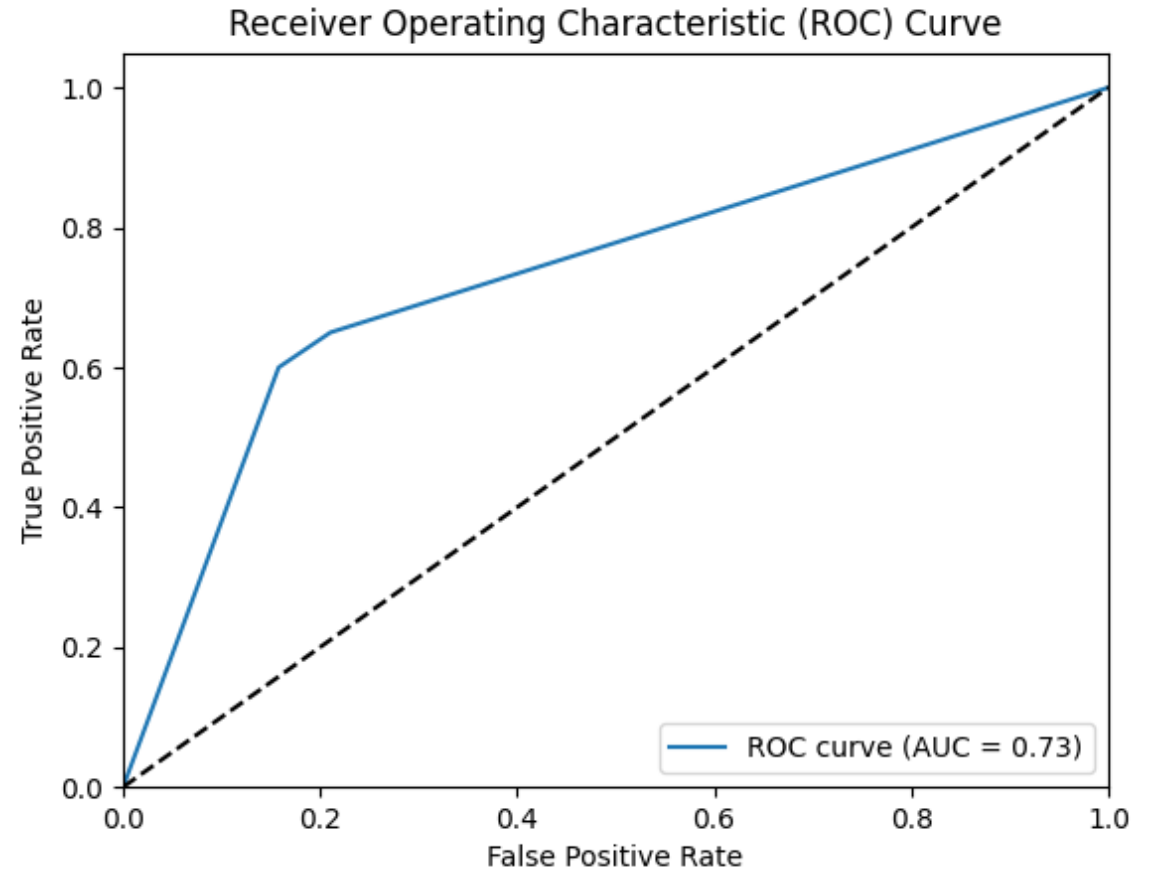
# GÖRSELLEŞTİRME ARAÇLARI: CONFUSION MATRİS

```
65 # Confusion matrix'i oluştur
66 cm = confusion_matrix(y_test, y_pred)
67 print("Confusion Matrix:\n", cm)
68
69 #Confusion matrisini görselleştirme
70 # Heatmap oluştur
71 sns.heatmap(cm, annot=True, cmap="Blues")
72 # Eksen etiketlerini ayarla
73 plt.xlabel("Tahmin Edilen Sınıflar")
74 plt.ylabel("Gerçek Sınıflar")
75 # Grafiği göster
76 plt.show()
77
```



# GÖRSELLEŞTİRME ARAÇLARI: ROC EĞRİSİ

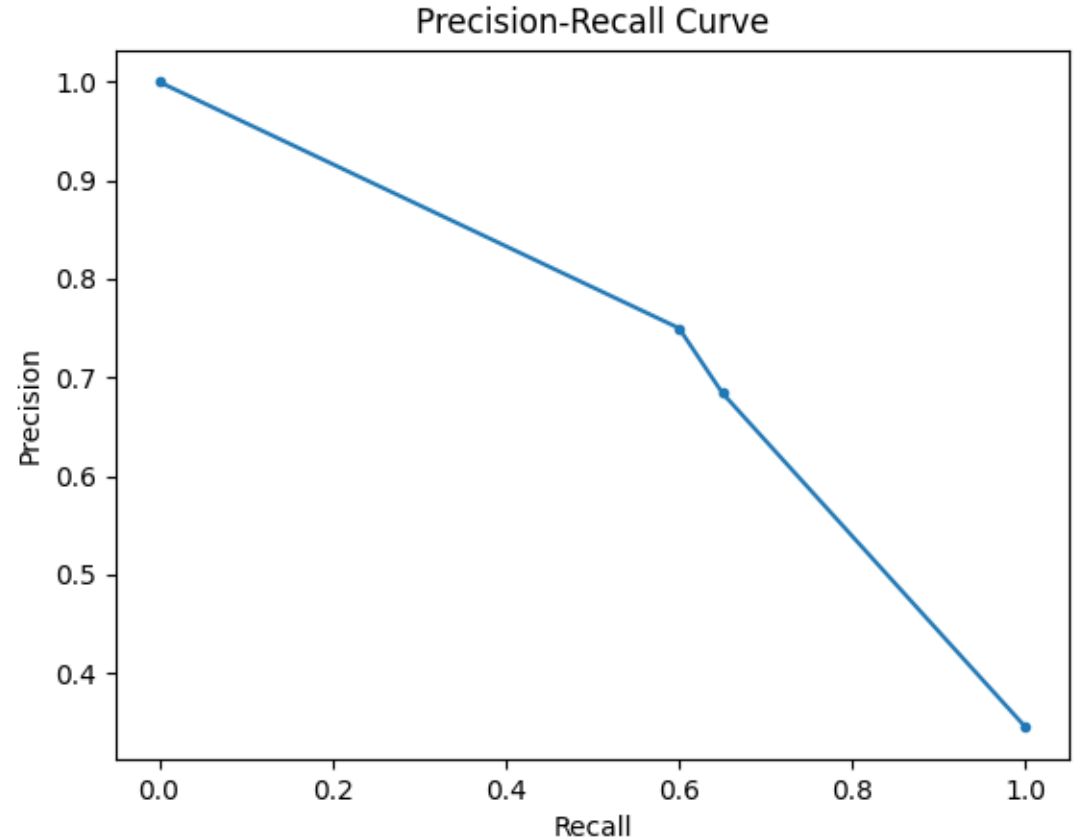
```
78 #Roc-curve grafiğini kullanarak görselleştirme
79 y_pred_prob = clf.predict_proba(X_test)[: , 1]
80 fpr, tpr, thresholds = roc_curve(y_test, y_pred_prob)
81 roc_auc = auc(fpr, tpr)
82
83 plt.plot(fpr, tpr, label='ROC curve (AUC = %0.2f)' % roc_auc)
84 plt.plot([0, 1], [0, 1], 'k--')
85 plt.xlim([0.0, 1.0])
86 plt.ylim([0.0, 1.05])
87 plt.xlabel('False Positive Rate')
88 plt.ylabel('True Positive Rate')
89 plt.title('Receiver Operating Characteristic (ROC) Curve')
90 plt.legend(loc="lower right")
91 plt.show()
92 y_pred_prob = clf.predict_proba(X_test)[: , 1]
93
```





# GÖRSELLEŞTİRME ARAÇLARI: RECALL- PRECISION EĞRİSİ

```
94 # precision-recall grafiğiyle görselleştirme
95 precision, recall, thresholds = precision_recall_curve(y_test, y_pred_prob)
96 # Precision-Recall eğrisini çizmek için bir çizgi grafiği oluşturma
97 plt.plot(recall, precision, marker='.')
98 # Eksen ve başlık etiketlerini ayarlama
99 plt.xlabel('Recall')
100 plt.ylabel('Precision')
101 plt.title('Precision-Recall Curve')
102 # Grafikleri göster
103 plt.show()
104
```



# SONUÇLARIN KARŞILAŞTIRILMASI

## «Breast Cancer Classification using Decision Tree Algorithms» Referans Makale

TABLE VIII. DETAILED ACCURACY BY CLASS OF DECISION TREE ALGORITHM

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.995	0.059	0.978	0.995	0.985	0.950	0.999	0.999	0
	0.941	0.005	0.988	0.941	0.964	0.950	0.999	0.996	1
Weighted Avg.	0.979	0.043	0.979	0.979	0.979	0.950	0.999	0.998	

## «Karar Ağacı Tabanlı Meme Kanseri Sınıflandırması»

Class	TP Rate	FP Rate	Precision	Recall	F-Measure
1	0.60	0.12	0.75	0.60	0.66

# SONUÇLARIN KARŞILAŞTIRILMASI: CONFUSION MATRİS

«Breast Cancer Classification using Decision Tree Algorithms» Referans Makale

TABLE IX. CONFUSION MATRIX OF DECISION TREE ALGORITHM

a	b	classified as
200	1	a = 0
5	80	b = 1

«Karar Ağacı Tabanlı Meme Kanseri Sınıflandırması»

a	b	classified as
32	6	a = 0
8	12	b = 1

×5

a	b	classified as
156	30	a = 0
40	60	b = 1

# KAYNAKÇA

- <https://i0.wp.com/rayhaber.com/wp-content/uploads/2021/06/veri-madenciligi-nedir-nasil-yapilir-veri-madenciliginin-faydalari-nelerdir.jpg?fit=1780%2C1000&ssl=1>
- <https://www.veribilimiokulu.com/siniflandirma-notlari-16-karar-agaci-python-uygulama/>
- <https://www.datascienceearth.com/python-uygulamasi-ile-karar-agaclari/>
- <https://chat.openai.com/>
- [https://t4.ftcdn.net/jpg/02/95/60/87/360\\_F\\_295608779\\_4xRpcPkb7tpIA959WkGeBqs1GNvCrI9l.jpg](https://t4.ftcdn.net/jpg/02/95/60/87/360_F_295608779_4xRpcPkb7tpIA959WkGeBqs1GNvCrI9l.jpg)
- [https://talosapp.me/mag/wp-content/uploads/2021/11/https\\_\\_specials-images.forbesimg.com\\_imageserve\\_614cb5f80cb8203307e03218\\_0x0-1.jpg](https://talosapp.me/mag/wp-content/uploads/2021/11/https__specials-images.forbesimg.com_imageserve_614cb5f80cb8203307e03218_0x0-1.jpg)