

```
# Gerekli kütüphaneler

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

# 1. Veri Setini Yükle
data = load_breast_cancer()
df = pd.DataFrame(data.data, columns=data.feature_names)
df['target'] = data.target

# 2. Veri Kümesini İncele
print("🔍 Veri Kümesi Şekli:", df.shape)
print(df.head())

# 3. Eksik Veri Kontrolü
print("\n💎 Eksik Değer Kontrolü:\n", df.isnull().sum())

# 4. Giriş ve Çıkış Değişkenlerini Ayır
X = df.drop('target', axis=1)
y = df['target']

# 5. Eğitim ve Test Verisine Ayırma (%80 - %20)
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

6. Model Kur ve Eğit

```
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)
```

7. Tahmin Yap

```
y_pred = model.predict(X_test)
```

8. Model Başarımı

```
print("\n Doğruluk (Accuracy):", accuracy_score(y_test, y_pred))
print("\n Sınıflandırma Raporu:\n", classification_report(y_test, y_pred))
print("\n Karışıklık Matrisi:\n", confusion_matrix(y_test, y_pred))
```

9. Özellik Önem Dereceleri

```
importances = pd.Series(model.feature_importances_, index=X.columns)
importances = importances.sort_values(ascending=False)
```

10. Görselleştir

```
plt.figure(figsize=(10,6))
sns.barplot(x=importances[:10], y=importances.index[:10], palette='viridis')
plt.title(" En Önemli 10 Özellik")
plt.xlabel("Önem Skoru")
plt.tight_layout()
plt.show()
```

model başarılı bir performans sergiledi:

Doğruluk (Accuracy): %96.49

Precision, Recall, F1-Score: Her iki sınıf (kanserli ve kanserli olmayan) için de yüksek değerlerde.

Ayrıca karışıklık matrisine baktığımızda:

Çok az sayıda yanlış sınıflandırma olduğunu görüyoruz.

Bu da, modelin her iki sınıfı da dengeli ve başarılı bir şekilde ayırt ettiğini gösteriyor.

Bu, tıbbi teşhis gibi kritik alanlarda güvenilir bir model için oldukça iyi bir başlangıç.

Veride Hangi Değişkenler Önemliydi?

Random Forest modelinin çıkardığı özellik önem düzeyine göre en önemli 10 özellik şunlardı:

1. worst area
2. worst concave points
3. mean concave points
4. worst radius
5. mean concavity
6. worst perimeter
7. mean area
8. worst concavity
9. mean radius
10. mean perimeter

Bu özellikler genellikle tümörün şekli ve büyüklüğüyle ilgili ölçümler. Özellikle concave points, yani tümörün girintili çıkıntılı yapısı, malign (kötü huylu) olup olmadığını anlamada belirleyici oluyor.

Daha İyi Sonuç Almak İçin Neler Yapılabilir?

Modeli daha da geliştirmek için aşağıdaki adımlar uygulanabilir:

1. Hiperparametre Optimizasyonu:

GridSearchCV ya da RandomizedSearchCV ile n_estimators, max_depth, min_samples_split gibi parametreleri ayarlayarak modeli optimize edebilirsiniz.

2. Başka Algoritmalarla Kıyaslama:

Lojistik Regresyon, XGBoost, SVM gibi farklı modeller deneyerek hangisinin daha iyi sonuç verdiğini görebilirsiniz.

3. Cross-Validation (Çapraz Doğrulama):

Veriyi farklı bölünmelerle test ederek modelin genelleme yeteneğini ölçebilirsiniz.

4. Özellik Mühendisliği:

Yeni özellikler üretmek (örneğin oranlar, çarpımlar) ya da daha az bilgi taşıyanları elemek faydalı olabilir.

5. Veri Normalizasyonu:

- Özellikle SVM veya KNN gibi algoritmalar kullanacaksan, veriyi ölçeklendirmek (standardizasyon) performansı artırabilir.

