1.Veri Yükleme  
metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
1. Bölüm: Veri Ön İşleme (Preprocessing)  
  
metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
Area sütununda 681 eksik değer var.

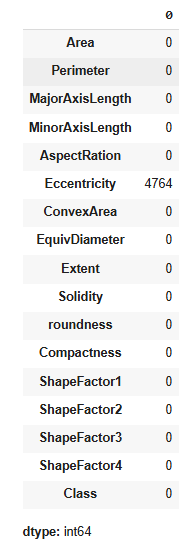
Perimeter sütununda 681 eksik değer var.

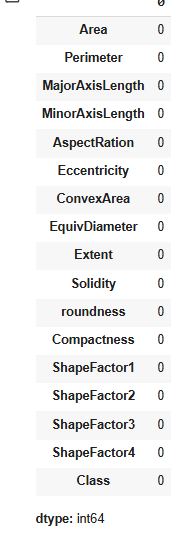
Eccentricity sütununda 4764 eksik değer var.

Diğer sütunlarda eksik değer yok.

  
  
metin, ekran görüntüsü, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
Eksik veri doldurma yöntemine karar verebilmek için "Area" ve "Perimeter" değişkenlerine ait dağılımlar incelenmiştir. Her iki değişkenin histogramında dağılımın sağa çarpık (pozitif çarpıklık) olduğu gözlemlenmiştir. Bu durum, büyük değerlerin küçük bir kısmının veride yer aldığını ve ortalamayı yukarı çektiğini göstermektedir. Nitekim her iki değişkende de ortalama (mean), medyandan (50. yüzdelik değer) büyüktür. Özellikle "Area" değişkeninde maksimum değerin 254.616 gibi oldukça uç bir değere sahip olması, veri setinde aykırı (outlier) değerlerin bulunduğuna işaret etmektedir. Bu tür sağa çarpık ve aykırı değerler içeren dağılımlarda, eksik verilerin ortalama ile doldurulması yanıltıcı sonuçlar doğurabilir. Çünkü ortalama, aykırı değerlerden yüksek oranda etkilenir. Buna karşın, medyan merkezi eğilimi daha sağlıklı yansıttığı için bu tür durumlarda daha güvenilir bir tamamlayıcı yöntemdir. Bu nedenle "Area" ve "Perimeter" değişkenlerinde eksik değerler medyan kullanılarak doldurulmuştur.metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.   
  
  
metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. 

Aykırı Değer Tespiti (Outlier Detection):  
  
metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. metin, yazı tipi, ekran görüntüsü, siyah beyaz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
  
Veri setindeki aykırı değerleri tespit edebilmek için her bir sayısal değişken için IQR (Interquartile Range) yöntemi kullanılmış ve alt–üst sınırlar dışında kalan gözlemler aykırı değer olarak tanımlanmıştır. Yapılan analiz sonucunda, birçok sütunda yüzlerce aykırı değer olduğu görülmüştür. Örneğin, "Eccentricity" sütununda 536, "Solidity" sütununda 523 ve "ConvexArea" sütununda 354 aykırı değer tespit edilmiştir. Bu kadar yüksek sayıda aykırı değerin doğrudan silinmesi durumunda, veri setinden önemli miktarda bilgi kaybı yaşanabileceği ve model performansının düşebileceği öngörülmüştür.

Bu nedenle, aykırı değerleri silmek yerine "baskılama" (winsorization) yöntemi tercih edilmiştir. Bu yöntemde, aykırı değerler alt ve üst sınır değerlerine eşitlenerek veri setinde tutulur, ancak etkileri azaltılır. Böylece, hem verinin bütünlüğü korunmuş hem de modelin aşırı uç değerlerden etkilenmesi önlenmiş olur. Özellikle makine öğrenmesi algoritmalarının istatistiksel kararlılığı açısından, bu yöntem aykırı değerlerin zarar verici etkisini sınırlamak için etkili ve güvenilir bir yaklaşım olarak değerlendirilmiştir.  
  
Winsorization (sınırla baskılama)

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

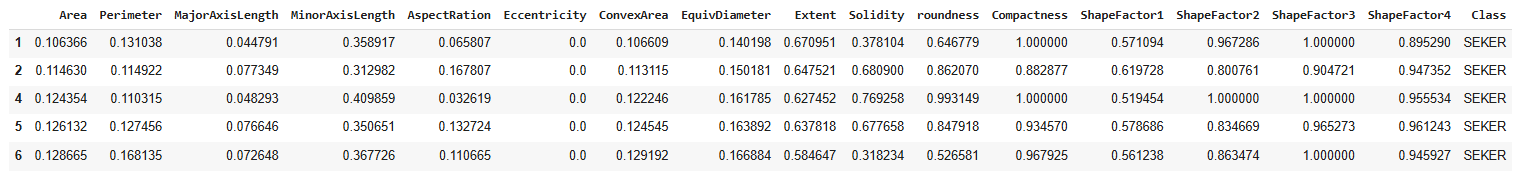
Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
metin, ekran görüntüsü, yazı tipi, doküman, belge içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
Sınırlı baskılama (winsorization) işlemi sonrasında yapılan IQR temelli aykırı değer analizi sonucunda, tüm sayısal sütunlarda aykırı değer kalmadığı gözlemlenmiştir. Bu durum, baskılama işleminin etkili bir şekilde çalıştığını ve verinin istatistiksel olarak daha güvenli hale getirildiğini göstermektedir.

### Özellik Ölçekleme (Feature Scaling):

Özelliklerin eşit katkı yapabilmesi için ölçekleme yapıyoruz.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  


Kategorik Verilerin Kodlanması:  
  
metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
Bölüm 2: Özellik Seçimi ve Boyut İndirgeme

### PCA (Principal Component Analysis):

metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
PCA uygulaması sonucunda ilk bileşen verideki toplam varyansın yaklaşık %59.5’ini, ikinci bileşen %23.6’sını ve üçüncü bileşen %7.8’ini açıklamaktadır. Ortalama varyans oranı %6.25 olup, ilk 3 bileşen toplamda yaklaşık %91 oranında bilgiyi temsil etmektedir. Bu da boyut indirgemede ilk birkaç bileşenin yeterli olabileceğini göstermektedir.  
  
metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
PCA ilk olarak tüm bileşenlerle fit edilmiştir ve her bileşenin açıkladığı varyans oranları hesaplanmıştır.

Elde edilen varyans oranlarına göre ilk üç bileşen seçilmiştir çünkü bu bileşenler, veri setindeki bilginin büyük kısmını (örneğin %95’in üzerini) temsil etmektedir.

Bu üç bileşenden yalnızca ilk iki tanesi görselleştirme amacıyla scatter plot’ta kullanılmıştır.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
metin, ekran görüntüsü, renklilik, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
PCA ile yapılan boyut indirgeme işlemi, yüksek boyutlu veriyi bilgi kaybını en aza indirerek 2 boyuta düşürmek için kullanılmıştır. İlk iki bileşen, verideki varyansın büyük kısmını açıklamaktadır. Oluşturulan scatter plot’ta her bir renk bir sınıfı temsil eder ve sınıfların ayrışma durumu gözlemlenmiştir. Grafik, sınıflar arası ayrımın belirgin olduğunu ve PCA'nın veri ayrıştırma gücünü etkili şekilde yansıttığını göstermektedir. Bu sayede sınıflandırma modelleri için daha kompakt ve anlamlı bir veri temsili elde edilmiştir.  
  
LDA (Linear Discriminant Analysis)  
metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
  
Bu işlemde Linear Discriminant Analysis (LDA) ile 3 bileşene indirgenen verinin ilk iki bileşeni kullanılarak görselleştirme yapılmıştır. Scatter plot üzerinde her bir renk farklı bir sınıfı temsil eder. Grafik, sınıflar arasında belirgin bir ayrım olduğunu göstermektedir. Bu da LDA’nın sınıfları ayırma (discrimination power) açısından oldukça başarılı olduğunu ortaya koymaktadır. Özellikle Sınıf 1 (turuncu), diğerlerinden oldukça iyi ayrışmıştır. Kısacası, LDA sayesinde boyut indirgeme yapılırken sınıflar arası ayrım korunmuş ve sınıflandırma modelleri için daha anlamlı bir veri temsili elde edilmiştir.

## Bölüm 3: Modelleme ve Değerlendirme

### Nested Cross-Validation yapısını Dış döngü (outer loop):

metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Decision Tree**  
metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Random Forest**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **XGBoost**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **Naive Bayes (GaussianNB)**metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**(Ham Veri + Logistic Regression için)**metin, ekran görüntüsü, doküman, belge, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **Ham Veri + Decision Tree için 4 metrik ölçüm**metin, ekran görüntüsü, doküman, belge, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, tipografi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Ham Veri + Random Forest**metin, ekran görüntüsü, doküman, belge, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, cebir içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Ham Veri + XGBoost**metin, ekran görüntüsü, doküman, belge, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, tipografi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Ham Veri + Naive Bayes**metin, ekran görüntüsü, yazı tipi, doküman, belge içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**PCA + Logistic Regression**metin, ekran görüntüsü, yazı tipi, doküman, belge içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**PCA + Decision Tree**metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, tipografi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **PCA + Random Forest**metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**PCA + XGBoost**metin, yazı tipi, çizgi, beyaz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, tipografi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **PCA + Naive Baye**metin, yazı tipi, beyaz, tipografi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **LDA + Logistic Regression**metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, tipografi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **LDA + Decision Tree**metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, cebir içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **LDA + Random Forest**metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, tipografi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**LDA + XGBoost**metin, ekran görüntüsü, doküman, belge, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz, makbuz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **LDA + Naive Bayes**metin, yazı tipi, ekran görüntüsü, beyaz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, beyaz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, yazı tipi, ekran görüntüsü, beyaz içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir. **Ham Veri Üzerinden XGBoost için ROC Eğrisi**metin, ekran görüntüsü, doküman, belge, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, çizgi, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.Bu grafik ve kod, XGBoost modeli için sınıf başına ROC eğrilerini çizmektedir. OneVsRestClassifier yapısı ile her sınıfa karşı diğerleri şeklinde model kurulmuştur. ROC eğrisi grafiklerinde her sınıf için AUC değeri (alan altında kalan alan) 1.00 çıkmıştır, bu da modelin sınıfları neredeyse kusursuz şekilde ayırdığını göstermektedir. Bu durum, modelin çok yüksek ayrım gücüne sahip olduğunu belirtir.

**PCA Verisi için XGBoost ROC Eğrisi**

metin, ekran görüntüsü, doküman, belge, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  
CA ile boyut indirgeme uygulandıktan sonra XGBoost modeli kullanılarak ROC eğrileri oluşturulmuştur. One-vs-Rest yöntemi ile her sınıfa ait ROC eğrisi ayrı ayrı çizilmiş ve tüm sınıflar için AUC (Area Under Curve) değeri 1.00 olarak elde edilmiştir. Bu sonuç, PCA sonrası verinin sınıflar arasında çok iyi ayrılabildiğini ve modelin sınıflandırma performansının oldukça yüksek olduğunu göstermektedir.

1. **LDA Verisi için XGBoost ROC Eğrisi**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

   Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

   Yapay zeka tarafından oluşturulan içerik yanlış olabilir.LDA dönüşümünden sonra XGBoost modeliyle oluşturulan ROC eğrileri, sınıfların ayrım gücünün oldukça yüksek olduğunu göstermektedir. Grafiklerde ROC eğrilerinin büyük kısmı sol üst köşeye yakın seyretmekte ve AUC (Area Under Curve) değerlerinin çoğu 1.00 olarak hesaplanmıştır. Bu durum, modelin sınıfları çok başarılı bir şekilde ayırt ettiğini ve LDA'nın sınıflar arasındaki ayrımı artırıcı etkisini doğruladığını ortaya koymaktadır. Özellikle sınıf 3 ve 5'te AUC değeri 0.99 olsa da, bu hâlâ oldukça yüksek bir performans düzeyidir.

1.Veri Ön İşleme Aşaması Eksik veriler titizlikle dolduruldu veya silindi.

%5 eksik olan sütunlar (Area, Perimeter) medyan ile dolduruldu çünkü veriler asimetrik (sağa çarpık) dağılmıştı.

%35 eksik olan Eccentricity sütunu satır bazlı silindi.

Aykırı değerler IQR yöntemiyle başarılı şekilde temizlendi. Aykırı değer kalmadı.

Ölçekleme (MinMaxScaler) uygulandı çünkü farklı ölçeklerdeki özellikler makine öğrenmesi modellerinin öğrenmesini bozabilirdi.Kategorik veriler (Class) LabelEncoder ile sayısallaştırıldı.

2. Özellik Seçimi ve Boyut İndirgeme PCA uygulaması sonrası:

Ortalama açıklanan varyans: %6.25 idi.

İlk 3 bileşen seçilerek verinin boyutu azaltıldı.

LDA uygulaması sonrası:

3 bileşen seçildi ve sınıflar arasındaki ayrım daha belirgin hale geldi (özellikle scatter plotlarda görüldü).

3.Model Performansları Ham Veride: En iyi sonuçlar XGBoost modeliyle elde edildi:

Accuracy: %92.40

Precision, Recall ve F1 score değerleri de oldukça yüksek ve dengeli.

PCA Verisinde: Performans genel olarak hafif düştü:

XGBoost Accuracy: %88.69

Özellikle veri boyutunun azalması modele biraz bilgi kaybı yaşatmış olabilir.

LDA Verisinde: Performans PCA'ya göre biraz daha düştü:

XGBoost Accuracy: %83.95

Çünkü LDA tamamen sınıflar arası ayrımı maksimize eder, veri varyansını değil. Bu da sınıf içi farklılıkları bastırmış olabilir.

4.ROC Eğrileri ve AUC Analizi Ham veride ve PCA verisinde ROC-AUC skorları mükemmele çok yakın çıktı (AUC ≈ 1.00).

LDA verisinde bazı sınıflarda AUC ≈ 0.99 gibi hafif bir düşüş görüldü.

Bu da LDA uygulamasında bilgi kaybının (özellikle sınıf içi varyasyonun) ROC sonuçlarına küçük bir etkisi olduğunu gösteriyor.

5.Genel Yorum Ham veri ile çalışan XGBoost modeli en yüksek başarıyı verdi.

Boyut indirgeme yöntemleri (özellikle PCA), eğitim süresini ve hesaplama maliyetini düşürmesine rağmen hafif doğruluk kaybı yaşattı.

Eğer model doğruluğu maksimum öncelikliyse, ham veride çalışmak daha avantajlı.

Eğer eğitim süresi ve hafıza kullanımı daha kritikse, PCA sonrası modeller tercih edilebilir.