

Projeye Genel Bakış:

Bu proje, meme kanseri tespiti için bir makine öğrenmesi modeli geliştirerek, tümörlerin iyi huylu (benign) veya kötü huylu (malignant) olarak doğru bir şekilde sınıflandırılmasını amaçlamaktadır. Veri kümesi üzerinde kapsamlı özellik analizi ve mühendisliği uygulanarak, çok boyutlu verilerden anlamlı bilgiler çıkarılması ve bu bilgilerin modelin performansını optimize etmesi hedeflenmektedir.

Proje, makine öğrenmesi yöntemlerini kullanarak veri dengesizliklerini gidermek, önemli özellikleri belirlemek ve etkili bir çözüm sunmak için tasarlanmıştır. Bu amaç, kanser tespiti için erken ve doğru bir tanıyı mümkün kılıp, sağlık hizmetlerinin etkinliğini artırmayı hedefler.

Hedeflenen Sonuçlar:

Doğru ve Kararlı Bir Model Geliştirme:

- Tümör sınıflandırmasında yüksek doğruluğa sahip bir sınıflandırma modeli oluşturmak.
- Modelin, hem eğitim hem de test veri setlerinde dengeli ve tutarlı performans göstermesini sağlamak.

Veri İşleme ve Analiz:

- Veri setindeki anlamsız veya gereksiz bilgileri temizlemek.
- Yeni özellikler üretmek ve hedef değişkenle olan korelasyonlarını analiz ederek en etkili olanları seçmek.

Sınıf Dengesizliğinin Giderilmesi:

- Sınıf dengesizliği problemi varsa, SMOTE gibi yöntemlerle veri setini dengelemek ve her iki tümör sınıfında da başarılı tahminler yapılmasını sağlamak.

Veri Görselleştirme:

- Veri kümelerinin ve özelliklerin hedef değişkenle olan ilişkisini görselleştirerek, karar sürecine şeffaflık kazandırmak.

Makine Öğrenmesi Modelleri

1. Modellerin Tanımlanması:

- DecisionTree, RandomForest, SVM ve LogisticRegression gibi modeller seçilerek optimize edilmeye hazır hale getirildi.

2. Hiperparametre Optimizasyonu:

- Modellerin hiperparametreleri, GridSearchCV ile optimize edilerek en iyi parametre değerleri belirlendi.
- 3. Overfitting Kontrolü:
 - Eğitim ve test doğruluk oranları karşılaştırılarak overfitting riskleri analiz edildi. Riskin yüksek olduğu modeller için yeni hiperparametre ayarları belirlendi ve yeniden eğitim yapıldı.
- 4. Performans Değerlendirmesi:
 - Modeller, test seti üzerinde Accuracy, Precision, Recall, F1-Score ve ROC-AUC metrikleri ile değerlendirildi.
- 5. Görselleştirme:
 - Modellerin Precision, Recall, F1-Score değerleri ikili grafiklerle; Accuracy ve ROC-AUC değerleri yan yana grafiklerle sunuldu.