

# Breast Cancer Wisconsin (Diagnostic) Dataset Projesi ve Sunumu

---

## Proje Açıklaması:

**Proje Amacı:** Bu projede öğrencilerin meme kanseri tanısında kullanılan bir veri seti üzerinde makine öğrenimi algoritmaları ile çalışmaları, iyi huylu (benign) ve kötü huylu (malignant) tümörleri sınıflandırmaları beklenmektedir.

## Kullanılacak Veri Seti:

- Kaggle - Breast Cancer Wisconsin Dataset
  - Veri setinde meme kanseri ile ilgili hücre çekirdeklerinin özellikleri bulunmaktadır ve tanıyı belirlemek amacıyla kullanılacaktır.
- 

## Adım Adım Proje Yönergeleri:

### 1. Kaggle Üzerinden Proje Yükleme:

Proje, Kaggle platformunda bir notebook olarak gerçekleştirilecektir. Her adımı detaylı bir şekilde Kaggle Notebook'unuzda açıklamanız gerekmektedir.

### 2. Notebook Adımları:

- **Giriş:** Meme kanseri, veri seti ve projenin amacı hakkında kısa bir bilgi verin.
- **Kütüphanelerin Yükleneşi:** Gerekli Python kütüphanelerini yükleyin ve not defterinizde açıkça belirtin.
- **Veriyi Yükleme ve İlk İnceleme:** Veri setini yükleyin, ilk incelemeyi yapın ve veri seti hakkında genel bir bilgi sağlayın (ilk 5 satır, özet istatistikler).

- **Eksik Veri Kontrolü:** Veri setinde eksik veri olup olmadığını kontrol edin ve varsa bunları uygun şekilde işleyin.
  - **Etiketleme (Label Encoding):** Diagnosis sütununu (B: 0, M: 1) sayısal değerlere dönüştürün.
  - **Özellik Mühendisliği:** Gereksiz sütunları çıkarın ve veriyi eğitim için hazırlayın.
  - **Veriyi Eğitim ve Test Olarak Bölme:** Veriyi eğitim ve test setlerine ayırın.
  - **Veri Ölçeklendirme:** Standart hale getirmek için veriyi ölçeklendirin.
  - **Model Geliştirme:** Random Forest veya başka bir modelle veriyi eğitin.
  - **Model Performansı:** Test setinde model performansını doğruluk, hata matrisi, sınıflandırma raporu gibi metriklerle değerlendirin.
  - **ROC Eğrisi ve AUC Skoru:** ROC eğrisi çizerek modelin performansını analiz edin.
  - **Sonuçlar ve Yorumlama:** Sonuçlarınızı özetleyin ve modelin başarısını tartışın.
- 

## **Sunum Hazırlığı:**

Öğrencilerden, yaptıkları bu projeyi bir sunum haline getirmeleri de istenecektir. Sunumda dikkat etmeleri gerekenler:

### **1. Kapak Sayfası:**

- Projenin adı, kendi isimleri ve tarih.

### **2. Giriş:**

- Meme kanseri hakkında kısa bilgi ve projenin amacı.
- Veri seti ve bu veri setinin kullanım amacını açıklayın.

### **3. Veri İncelemesi:**

- Veri setinin içeriği hakkında bilgi verin (kaç satır, kaç sütun, hangi özellikler var).
- Hangi sütunun hedef olduğunu belirtin (Diagnosis sütunu).

### **4. Veri Temizleme:**

- Eksik verileri nasıl işlediklerini ve gereksiz sütunları nasıl çıkardıklarını açıklasınlar.
- Etiketleme işlemini anlatın (B: 0, M: 1).

### **5. Model Geliştirme:**

- Hangi modeli kullandıklarını (örneğin Random Forest) ve neden bu modeli seçtiklerini açıklayın.
- Eğitim ve test verisini nasıl böldüklerini anlatın.

### **6. Performans Değerlendirmesi:**

- Hata matrisi ve doğruluk skorunu sunarak, modelin başarısını açıklayın.
- ROC eğrisini ve AUC skorunu göstermek için bir grafik ekleyin.

### **7. Sonuçlar:**

- Modelin ne kadar başarılı olduğunu ve hangi metriklerin önemli olduğunu özetlesinler.
- Modeli geliştirmek için ek olarak neler yapabileceklerini tartışın (örneğin hiperparametre ayarları).

---

### **Teslimat ve Sunum Bilgileri:**

- **Proje Teslimi:** Kaggle Notebook'un linki ile teslim edilmelidir.
- **Sunum Teslimi:** PowerPoint formatında hazırlanmış sunum, proje teslimiyle birlikte yollanmalıdır.
- **Tarih:** Teslimat için belirlediğiniz son teslim tarihini ekleyin.