Yasin KARAÇAM

Makine Öğrenmesi Ödevi: Veri Seti Analizi

Veri seti adı : Anaemia Prediction

Veri seti linki : <https://www.kaggle.com/datasets/humairmunir/anaemia-prediction>

Veri seti sahibi : <https://www.kaggle.com/humairmunir>

Veri seti kaynağı : Different Public Resourses

Veri seti amacı : Veri seti sağlık sektörüne yönelik olup anemi teşhisi ve anemiye neden olabilecek faktörlerin analiz edilmesi amacıyla toplanmıştır. İçerisinde cinsiyet (Gender), hemoglobin seviyeleri (Hb), ve görüntüleme sonuçları (piksel verileri) gibi çeşitli biyometrik ve tıbbi bilgiler yer alır. Bu veriler bir bireyin anemi olup olmadığını belirlemek için kullanılır.

Bu veri seti aşağıdaki sorulara cevap olmak için hazırlanmıştır

* Bir bireyin hemoglobin seviyesi anemi teşhisinde ne kadar etkilidir?
* Cinsiyet (Gender) anemi olma olasılığını nasıl etkiler?
* Görüntüleme verileri (kırmızı, yeşil, mavi pikseller) anemi teşhisinde nasıl kullanılabilir?

Veri seti açıklaması : Bu veri seti, aneminin tahmini amacıyla makine öğrenimi modellerinin geliştirilmesi ve test edilmesi için kullanılabilir. Aynı zamanda, hemoglobin seviyeleri ile görüntülerdeki renkli piksel dağılımı arasındaki korelasyonun analizi için de uygundur. Veri ön işleme, özellik mühendisliği ve model oluşturma süreçlerini öğrenmek amacıyla eğitim amaçlı kullanım da öngörülmektedir (Humair M.)

Veri setindeki veriler: Number , Gender , %Red Pixel , %Green Pixel , %Blue Pixel , Hb , Anaemic

Veri Türleri

1. Number

Veri Türü: Sürekli

Veri Setinde Oynadığı Rol: Tanımlayıcı özellik (Id)

Açıklama: Bu değişken, her bir gözlemi benzersiz bir şekilde tanımlamak amacıyla kullanılan bir index numarasıdır

1. Gender (Cinsiyet)

Veri Türü: Kategorik

Veri Setinde Oynadığı Rol: Bağımsız değişken

Açıklama: Bu değişken, bireylerin cinsiyetini belirtmektedir. "M" (Erkek) ve "F" (Kadın) olmak üzere iki kategori içermektedir. Cinsiyetin anemi durumu üzerinde etkili olabileceği düşünülerek modelde bağımsız değişken olarak yer alıyor

1. %Red Pixel, %Green Pixel, %Blue Pixel

Veri Türü: Sürekli (Numerik)

Veri Setinde Oynadığı Rol: Bağımsız değişkenler

Açıklama: Bu üç değişken bireyin tıbbi görüntüleme veya biyometrik analizlerden elde edilen kırmızı, yeşil ve mavi piksel oranlarını ifade eder. Görüntü işleme teknikleriyle elde edilen bu değerler anemi teşhisine yardımcı olabilir. Sayısal ve sürekli değişkenler olup sınıflandırma modellerinde bağımsız değişkenler olarak kullanılır.

1. Hb (Hemoglobin)

Veri Türü: Sürekli

Veri Setinde Oynadığı Rol: Bağımsız değişken

Açıklama: Kişinin kanındaki hemoglobin seviyesini ifade eden bu değişken, anemi teşhisinde kritik bir rol oynar. Hemoglobin düzeyi düşük olan bireylerde anemi olasılığı daha yüksektir. Sürekli bir değişken olan hemoglobin değeri bağımsız değişken olarak modele dahil edilir.

1. Anaemic (Anemik Durum)

Veri Türü: Kategorik

Veri Setinde Oynadığı Rol: Bağımlı değişken (hedef değişken)

Açıklama: Bu değişken, bireyin anemik olup olmadığını gösterir. "Yes" (Anemik) ve "No" (Anemik değil) olmak üzere iki kategori içerir. Anemi durumu sınıflandırma modelinin tahmin etmeye çalıştığı hedef değişkendir.