

## **Proje Hikayesi:**

Bu projeye başlama sürecim, sağlık alanında yakın ilgimden kaynaklanmaktadır. Sağlık verileriyle ilgili bilgim, iki yakın arkadaşım sayesinde gelişti. Kendileri hemşire olup, daha önce TÜBİTAK projelerinde çalışıyorlardı ve bu projeler kapsamında çok sayıda sağlık verisi topluyorlardı. Bu veriler, çeşitli hastalıkların yaygınlıkları ve etkileyen faktörlerle ilgiliydi. Onlara, sahip oldukları veriler hakkında bilgi alabileceğimi sordum ve onlar da bana 30 deneğin verilerini sağladılar.

Bu verileri aldıktan sonra, verilerin ne anlama geldiğini ve ne tür analizler yapabileceğimi sordum. Arkadaşlarım, KOAH (Kronik Obstrüktif Akciğer Hastalığı) hastalığının hangi faktörlerden etkilenebileceği ve risk faktörlerinin neler olduğu konusunda bana yardımcı oldular. KOAH hastalığının gelişimine katkıda bulunan birçok faktör bulunmaktaydı: yaş, cinsiyet, sigara içimi, kronik hastalıklar, ilaç kullanımı, ekonomik durum ve daha birçok etken bu hastalığın oluşumunda önemli rol oynamaktadır. Bu bilgileri aldıktan sonra, bu veriler üzerinde çalışarak bir model geliştirmeye karar verdim.

## **Yapılanlar ve Kullanılan Yöntemler:**

İlk olarak, verileri JSON formatında aldım ve gerekli ön işleme işlemlerini gerçekleştirdim. Veriler, sayısal olmayan kategorik verilere sahipti, bu nedenle her bir kategoriye karşılık gelen sayısal değerleri tanımladım ve verileri sayısal hale getirdim. Ardından, bu veriler üzerinde temel keşifsel veri analizi (EDA) yaparak eksik veya hatalı veri olup olmadığını kontrol ettim.

Sonrasında, bu veriler üzerinde risk analizi yapmak amacıyla bir model geliştirmeye başladım. Risk puanını hesaplamak için belirlediğim faktörlere göre her bir deneğin KOAH hastalığına yakalanma riskini değerlendiriyorum. Kullanıcıların yaş, cinsiyet, sigara içme durumu, kronik hastalıklar, ilaç kullanımı gibi verileri göz önünde bulundurularak her denek için bir risk puanı oluşturuluyor. Bu risk puanına göre, her bir denek için KOAH riski yüzde olarak hesaplanabiliyor.

Makine öğrenmesi yöntemleri kullanarak, bu veriler üzerinde sınıflandırma yapmak için bir model oluşturdum. Veriyi eğitim ve test setlerine ayırarak, rastgele orman (Random Forest) algoritması ile bir model geliştirdim. Modelin doğruluğunu ölçerek, farklı faktörlerin KOAH riskini nasıl etkilediği üzerine sonuçlar elde ettim.

## **Sonuçlar:**

Bu proje, sağlık verisi üzerinde yapılan analizlerin ve makine öğrenmesi tekniklerinin sağlık alanındaki problemlere nasıl uygulanabileceğini gösteriyor. KOAH hastalığının risk faktörlerine dair çıkardığım sonuçlar, özellikle hastalığın önceden tespiti ve erken müdahale için önemli olabilir. Ayrıca, bu tür projeler, sağlık alanındaki araştırmalara katkıda bulunarak daha sağlıklı toplumların oluşmasına yardımcı olabilir.