

Talent Academy Case Study – EDA & Veri Ön İşleme Raporu

Hazırlayan: Habibe Bayındır

E-posta: Habibebayindir55@gmail.com

Tarih: 06.09.2025

1. Giriş

Bu çalışma, fizik tedavi ve rehabilitasyon alanındaki bir veri seti üzerinde gerçekleştirildi. Veri seti 2.235 gözlem ve 13 değişkenden oluşmaktadır. Çalışmanın amacı, veri seti üzerinde **Keşifsel Veri Analizi (EDA)** yapmak ve veriyi **potansiyel makine öğrenmesi modellerine** hazır hale getirmektir.

Hedef değişken (**target**) olarak **TedaviSuresi** belirlenmiştir. Bu raporda veri temizleme, ön işleme ve analiz adımları özetlenmiştir.

2. Veri Ön İşleme Adımları

- **Eksik değerler temizlendi:**
 - Kategorik değişkenlerde eksik değerler "**Bilinmiyor**" etiketi ile dolduruldu.
 - Sayısal değişkenlerde eksik değerler **medyan** ile dolduruldu.
- **Tekrar eden satırlar silindi.**
- **Süre değişkenleri sayısallaştırıldı:**
 - TedaviSuresi ve UygulamaSuresi sütunları içerisindeki değerler (örn. "30 dk", "10 seans") düzenlenip yalnızca sayısal kısımlar alınarak float tipine dönüştürüldü.
- **Kategorik değişkenler kodlandı:**
 - Modelleme için uygun hale getirmek amacıyla kategorik değişkenler One-Hot Encoding yöntemi ile sayısallaştırıldı.

3. Keşifsel Veri Analizi (EDA)

Verinin yapısını daha iyi anlamak için çeşitli görselleştirmeler ve istatistiksel analizler yapıldı:

- **Tanımlayıcı istatistikler** çıkarıldı (ortalama, medyan, min, max, std).
- **Histogramlar** ile tedavi süresi dağılımı incelendi.
- **Boxplot** ile uç değerler gözlemlendi.
- **Bar grafikleri** ile en çok seans alan 10 hasta belirlendi.
- **Scatter plot** ile Tedavi Süresi – Uygulama Süresi ilişkisi incelendi.
- **Korelasyon matrisi & heatmap** oluşturularak değişkenler arasındaki ilişkiler görselleştirildi.

4. Modelleme Denemeleri (Opsiyonel)

Asıl amaç veri hazırlığı olmakla birlikte, örnek olarak bazı basit tahmin modelleri denendi:

- **Doğrusal Regresyon (Linear Regression):** Tedavi Süresi → Uygulama Süresi ilişkisi incelendi.
- **Random Forest Regressor:** Daha karmaşık ilişkileri yakalamak için deneme yapıldı.

Sonuçlar, hedef değişkenin tahmin edilmesinin zor olduğunu gösterse de, bu kısım veri hazırlama sürecini destekleyici amaçla eklenmiştir.

5. Sonuç

- Veri seti **eksiksiz, temiz ve tutarlı** hale getirildi.
- Kategorik ve sayısal değişkenler modele uygun formatta hazırlandı.
- EDA ile verinin yapısı, dağılımı ve ilişkileri ortaya konuldu.
- Bu aşamadan sonra, veri seti üzerinde daha gelişmiş modelleme ve analiz yapılabilir.