

Veri Bilimine Giriş Dersi Proje Ödevi

Part-1: Keşifçi Veri Analizi ve Veri Görselleştirme

1- Veri Seti Seçimi

En az 5 değişken (özellik, sütun)'den oluşan ve içerisinde eksik veriler barındıran bir veri seti tercih edilmelidir.

2- Veri Seti Hikâyesi ve Görselleştirme

- Veri Seti Hikayesi Anlatılmalıdır (Veri seti bize ne anlatıyor? Nasıl oluşmuş hangi değişkenleri içeriyor her bir değişken bize ne anlatıyor?)
- Veri Yapısı Nasıldır?
 - ✓ Gözlem Sayısı (Satır Sayısı)
 - ✓ Değişken Bilgileri (Sütun Sayısı, Değişken Türleri)
 - ✓ Sayısal Değişkenler için; Betimsel İstatistikler
 - ✓ Kategorik Değişkenler için; sınıf sayıları ve her bir sınıf frekansları
 - ✓ Eksik Veri Mevcut mu ? Var ise değişken başına sayıları.
- En az 2 yöntemle veri görselleştirme (çaprazlamalı yorumları da dahil)

Part-2: Veri Ön İşleme

3- Veri Ön İşleme

- **Aykırı Değer Problemini aşağıdaki yollardan biri ile çözünüz.**

Not: Aykırı değer problemini çözerken eksik verileri görmezden gelerek çözümünüzü yapınız.

1. **Yol :** Her bir sayısal değişkene ait varsa aykırı değerler tespit edilmeli ve bu değerler ya ortalama ile doldurma ya da baskılama yöntemi ile aykırı veri problemleri her bir değişken için çözülmeli. Veri seti aykırı değer problemi çözülmüş verisyonu ile yeniden kaydedilmeli (ileri aşamalarda kullanacağımız veri seti artık aykırı değerleri çözümlenmiş veri setimiz olmalı).
2. **Yol:** Her bir değişken için ayrı ayrı incelemek yerine LOF (Local Outlier Factor) yöntemi ile Çok Değişkenli Aykırı Gözlem Analizi yaparak ya silme ya da baskılama yöntemi ile aykırı değer problemini çözünüz. Veri seti aykırı değer problemi çözülmüş verisyonu ile yeniden kaydedilmeli (ileri aşamalarda kullanacağımız veri seti artık aykırı değerleri çözümlenmiş veri setimiz olmalı).

- **Eksik Veri Problemi Çözümü:** Aykırı değer problemi çözülmüş veri setinizdeki eksik verileri saptayınız. (Hangi değişken başına kaç eksik veri var). Tespit ettiğiniz eksik veri problemini derste görmüş olduğumuz yöntemlerden yalnızca biri (silme (Bir gözlem için değişkenlerinin %50'sinden fazlası boş ise o

gözlem için silme yapabilirsiniz), değer atama, tahmin yöntemleri ile değer atama) ile çözünüz. Veri seti eksik değer problemi çözülmüş verisyonu ile yeniden kaydedilmeli.

İleri aşamalarda kullanacağımız veri seti artık hem aykırı değerleri çözümlenmiş hem de eksik veri problemi kalmamış veri setimiz olmalıdır.