



Feature Engineering

İş Problemi

Özellikleri belirtildiğinde kişilerin diyabet hastası olup olmadıklarını tahmin edebilecek bir makine öğrenmesi modeli geliştirilmesi istenmektedir. Modeli geliştirmeden önce gerekli olan veri analizi ve özellik mühendisliği adımlarını gerçekleştirmeniz beklenmektedir.



Veri seti ABD'deki Ulusal Diyabet-Sindirim-Böbrek Hastalıkları Enstitüleri'nde tutulan büyük veri setinin parçasıdır. ABD'deki Arizona Eyaleti'nin en büyük 5. şehri olan Phoenix şehrinde yaşayan 21 yaş ve üzerinde olan **Pima Indian kadınları** üzerinde yapılan diyabet araştırması için kullanılan verilerdir.

Hedef değişken "**outcome**" olarak belirtilmiş olup; **1** diyabet test sonucunun **pozitif** oluşunu, **0** ise **negatif** oluşunu belirtmektedir.

9 Değişken

768 Gözlem

24 KB

Pregnancies	Hamilelik sayısı
Glucose	Oral glikoz tolerans testinde 2 saatlik plazma glikoz konsantrasyonu
Blood Pressure	Kan Basıncı (Küçük tansiyon) (mm Hg)
SkinThickness	Cilt Kalınlığı
Insulin	2 saatlik serum insülini (mu U/ml)
DiabetesPedigreeFunction	Fonksiyon (Oral glikoz tolerans testinde 2 saatlik plazma glikoz konsantrasyonu)
BMI	Vücut kitle endeksi
Age	Yaş (yıl)
Outcome	Hastalığa sahip (1) ya da değil (0)



Proje Görevleri

Görev 1 : Keşifçi Veri Analizi

Adım 1: Genel resmi inceleyiniz.

Adım 2: Numerik ve kategorik değişkenleri yakalayınız.

Adım 3: Numerik ve kategorik değişkenlerin analizini yapınız.

Adım 4: Hedef değişken analizi yapınız. (Kategorik değişkenlere göre hedef değişkenin ortalaması, hedef değişkene göre numerik değişkenlerin ortalaması)

Adım 5: Aykırı gözlem analizi yapınız.

Adım 6: Eksik gözlem analizi yapınız.

Adım 7: Korelasyon analizi yapınız.

Görev 2 : Feature Engineering

Adım 1: Eksik ve aykırı değerler için gerekli işlemleri yapınız. Veri setinde eksik gözlem bulunmamakta ama Glikoz, Insulin vb. değişkenlerde 0 değeri içeren gözlem birimleri eksik değeri ifade ediyor olabilir. Örneğin; bir kişinin glikoz veya insulin değeri 0 olamayacaktır. Bu durumu dikkate alarak sıfır değerlerini ilgili değerlerde NaN olarak atama yapıp sonrasında eksik değerlere işlemleri uygulayabilirsiniz.

Adım 2: Yeni değişkenler oluşturunuz.

Adım 3: Encoding işlemlerini gerçekleştiriniz.

Adım 4: Numerik değişkenler için standartlaştırma yapınız.

Adım 5: Model oluşturunuz.

miuul

miuul.com