Modeling:

1. En basta targetimi y, featureslari da X olarak atiyorum. Sonrasinda Train ve test olarak datami boluyorum.(train\_test\_split) (min %70 train, %30test olacak)

2. Elimde X\_train, y\_train, X\_test, y\_test datasi olacak.

3. Model’i linear regression olarak tanimliyorum.(model=LinearRegression()). coefficient ve intercepti cagirarak katsayilari bulduruyorum

4. X-train ile y\_train’i model fit kullanarak egitiyorum (model.fit)

5. X\_test datasetini kullanarak bir y-predict elde ediyorum (model.predict)

Bunu yaparken elde ettigim regresson line’ina (y=aX1+bX2+cX3+dX4+e) X\_test verilerini vererek y\_predict elde etmis oluyorum. Burdaki a,b,c,d,e’yi de coefficient ve interceptleri cagirarak bulmustum.

1. Sonra y-predict ile y-actual’i, yani y-testi kiyasliyorum. Residual’lari elde ediyorum. Bakiyorum hata payim ne kadar diye. Bunu da eval\_metric ile error metricleri cagirarak yapiyorum.
2. Train datasetimin error metriclerini de cagirip, test sonucunda elde ettiklerim ile kiyasliyorum.
3. Boyle boyle denedigim modeller icinde en iyi sonuc aldigim modele, final\_model olarak tum datami sokuyorum.

**Normalizasyon:**

En buyuk degeri 1, en kucuk degeri 0’a esitleyerek degerleri bunun arasinda oranlayarak 0-1 arasinda yerlestiriyor.

Buyuk degerlere bilgisayarin daha fazla onem vermesinin onune gecmek ve feature’lari ayni olcuye getirmek icin yaptigimiz bir islem. Scaling

Serilerin normal dagilima yaklasmasini da sagliyor.

Ornegin X1--> balik yasi, X2-->balik yumurtasi olsun. X1--> 1,2,3,4 diye giderken, X2--> 1milyon, 2 milyon diye gidiyor. Bilgisayar balik yumurtasi feature’nun daha onemli oldugunu sanmasin diye

Adaletin onunde esit hale getirmek gibi.

**Standardizasyon:**

Benzer bir islem yapiyor ayni mantikla normalizasyon ile.

Ortalamasi 0, standard sapmasi 1 olacak sekilde yeniden deger atiyor degerlere.

Burda da degerler kuculuyor ama -’ler de giriyor burda icine ortalamasi 0 olmasi icin.

Her ikisi de scale islemi. Usenmeyip ikisini de deneyip onun sonuclarini kullanacagiz.

Sayilari belli araliklara indirip daha iyi analiz yapmamizi sagliyor.

Bir datada normalizasyon yaparsan hepsine normalizasyon yap ki ayni dilden konussunlar. Standardizasyon yapiyorsan da digerlerine de standardizasyon yap

Hata terimlerinin karelerinin toplaminin 0’a yakin olmasini istiyorum. Verilerim de kucuk olursa, hata miktarlarim da kuculmus oluyor normalizasyon ve standardizasyon ile. Verilerimi tren yoluna hapsetmis oluyorum, aralarinda bir regresyon dogrusu cizerek

Analizin sonunda dusuk hata degerleri almis oluyorum, ama yine de yorumlayabiliyorum.

Minyaturk gibi dusun, esit olcekte kucultuyor, Selimiye’nin ayni olcegini koruyacak sekilde kucultuyor. Ayni olcekte kucultmezse zaten analizi bozmus oluyor.