

# Mansurjonov Nurmuhammad

## Loyiha Maqsadi

Ushbu loyihada biz datani mse(mean\_squared\_errors) larini aniqlab uni uni darajasini tushurishga va r2\_scoreni oshirishga harakat qildik. Bu uchun biz Random Forrest Regressor, KNN, Decision Tree Regressor va Linear Regression dan foydlandik.

## Qilingan asosiy bosqichlar

- 1) Tushib qolgan malumotlarni toldirish
- 2) Claslarni ajratish va kerak emaslarni olib tashlash
- 3) Friquent va Target Encoding orqali categoricaldan numericalga otkazish
- 4) Datani Train va testing qilish (4 model yordamida)
- 5) Yangi classlar yaratish va boshqatdan train va testing qilib(mse,r2\_scorelarni tekshirish)
- 6) KFold va cross\_val\_score yordamida (Har bir foldning MSE qiymatlari:, O'rtacha MSE:, MSE standart chetlanishini korib chiqdik, 4 modelda)
- 7) Mutal Information qilib qaysi calssimizni bogliqligini tekshirib chiqdik

## Olingan Natijalar

LinearRegression:

mse:1.4002087664006476

r2:0.3728899340957652

Feature engineering:

mse:1.4053315650509366

r2:0.37059559151179833

Scaling:

mse:0.466616672249738

r2:0.533383327750262

k\_fold:

Har bir foldning MSE qiymatlari: [104.20222653 66.52074552 63.17447946 69.47370565 105.17473228]

O'rtacha MSE: 81.70917788826905

MSE standart chetlanishi: 18.870610932980682

## Decision Tree Regressor

mse:1.0540938914618394

r2:0.5279040485918931

Feature engineering:

mse:3018788373945986.5

r2:0.8647978343256193

Scaling:

mse:1.0932429870568732

r2:-0.09324298705687317

k\_fold:

Har bir foldning MSE qiymatlari: [226.22879169 94.43033405 248.89578985 61.507844 177.72695524]

O'rtacha MSE: 161.7579429650769

MSE standart chetlanishi: 72.92131830036523

## Random Forrest Regressor Natija

mse:6516201261634713.0

r2:0.7081595615821369

Feature engineering:

mse:1293643536049496.5

r2:0.9420617194653059

Scaling:

mse:1.0932429870568732

r2:-0.09324298705687317

k\_fold:

Har bir foldning MSE qiymatlari: [152.19594652 67.73452145 184.07955107 65.89136806 146.19030073]

O'rtacha MSE: 123.21833756569724

MSE standart chetlanishi: 47.82538134575722

## K-Nearest Neighbors (KNN) Natija

mse:6108388592102639.0

r2:0.7264242258381857

Feature engineering:

mse:5111685628472063.0

r2:0.7710634593730563

Scaling:

mse:1.0932429870568732

r2:-0.09324298705687317

k\_fold:

Har bir foldning MSE qiymatlari: [108.14198878 74.98410274 156.69390886 98.73930714 138.3767445 ]

O'rtacha MSE: 115.38721040536943

MSE standart chetlanishi: 28.97177607227407

## Xulosa

Ushbu loyihada biz barcha modellar orqali mse ni tushurish va  $r^2$  scoreni oshirish uchun bazi ishlarni qilib chiqdik va natijamiz shuni korsatyabtiki bizning datamiz Linear Regression orqali yahshi ishlashini kordik.