

Yapay Zeka

Ödev 2

16011038

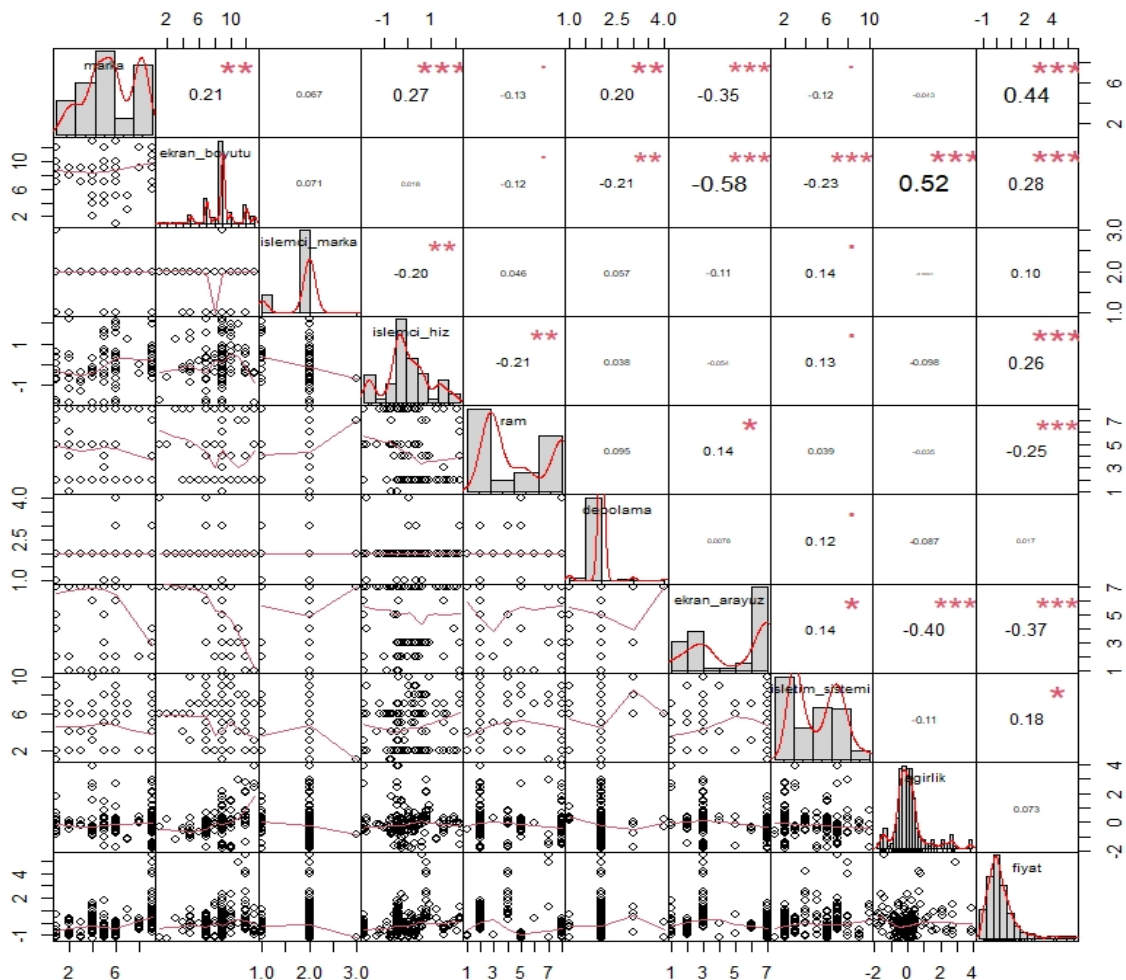
TALHA BACAK

Veri Kümesi

Veri oluşturmak için Amazon'dan veri çeken kod yazdım. Bunun için python'da selenium kütüphanesini kullandım. Amazondan bilgisayarların özelliklerini çekerek fiyatlarını tahmin ettim. 200 bilgisayarın verisini çektim. Veri kümesinde bulunan eşsiz değer sayısı şu kadar:

[illegible]

Veriler arasında şu şekilde bir korelasyon mevcuttur:



Modeller

Regresyon için şu modelleri kullandım:

- Lineer Regresyon
- Ridge Regresyon
- Lasso Regresyon
- Regresyon Tree
- Random Forest

Sonuçlar

Özellik değişimi yapılmış bağımlı değişkenin özellikleri:

Min	Median	Mean	Max
- 1.19525	- 0.29581	- 0.08626	2.57027

Sonuçlar şu şekilde:

Model	R2	RMSE	MAE
Lineer Regresyon (özellik değişimi yok)	0.2953417	0.6851206	0.5654407
Lineer Regresyon	0.3517245	0.6832572	0.5710802
Ridge Regresyon	0.3613363	0.6638286	0.5478668
Lasso Regresyon	0.3347452	0.6758463	0.553247
Regresyon Tree	0.5970499	0.4624376	0.349227
Random Forest	0.7139355	0.4064132	0.3132251

Random forest algoritması en iyi sonuçları verdi. Lineer, ridge ve lasso regresyonlar kötü sonuçlar verdi.

Kategorik özellikler için one hot encoding, sayısal özellikler için pca işlemleri yapıldı. Bu işlemler Lineer Regresyon modellerinden birinde yapılmadı.

Lasso ve ridge regresyon modellerinin lamda değerleri cross validation ile bulundu.