

YAP470 ÖDEV 1 RAPOR

Yapılan Adımlar :

1. İlk başta veriler görselleştirilmiştir.
2. Gereksiz görülen sütunlar çıkarılmıştır.
3. Verilerin eğimlerinin sayısal değerlerine göre log ve sqrt ile yeniden şekillendirilmiştir.
4. Stringler integera çevrilmiştir.
5. Eklemeler ve düzeltmeler sonucunda oluşan farkı görebilmek için yeniden veriler aynı şekilde görselleştirilmiştir.
6. Elimizde olan veriler test ve tarin olmak üzere ayrılmıştır. Elimizde az veri olduğu için `test_size=0.149` seçilmiştir.
7. `metrics = ['accuracy', 'precision', 'recall', 'f1_weighted']` şeklinde 4 metrik seçilmiştir.
8. GridSearch ile Logistic Regression, SVM, Random Forest için seçilen parametrelerden en iyisi bulunup sonrasında en iyi parametreler ile Cross validation score, test score süreleri bulunup yazdırılmıştır. Geçen süreler time kütüphanesi ile hesaplanmıştır.
9. Son adım olarak sonucu en iyi çıkan Logistic Regression için en iyi 5 parametre bulunup yeniden test score hesaplanmıştır.
10. Ayrıca farkı görebilmek için farklı sayıda parametre için de test score hesabı yapılmıştır.

Çıkarılan Sonuçlar :

- Grid Search sonucunda :
 - SVC için verilen parametreler içinde en iyi sonucu olan parametreler `C=1, kernel='linear'`,
 - Logistic Regression için verilen parametreler içinde en iyi sonucu olan parametreler `C=10, penalty='l1, solver = 'liblinear'`,
 - Random Forest Classifier için verilen parametreler içinde en iyi sonucu olan parametreler `max_depth=15, criterion='entropy', solver = 'liblinear'` olarak bulunmuştur.
- En iyi sonuç veren parametreler sonucunda elde edilen 4 metrik için cross validation ve test score değerleri şu şekildedir;
 - SVC, Logistic Regression, Random Forest Classifier en iyi sonuç veren metrik `accuracy`'dir.
 - En iyi `accuracy` sonucu yaklaşık 0.01 farkla Logistic Regression'a aittir. Test score 0.8909090909090909, Cross validation score 0.8728 olarak bulunmuştur.
- Elimizde çok fazla veri olmadığı için ve veriler çok dengeli olmadığı için `recall` ve `precision` sonuçları iyi sonuçlar vermedi.
- En iyi 5 öznitelik seçildiğinde test skorlarında düşüş görülmektedir. Yani en iyi 5 öznitelik yeterli değildir. 21 öznitelik ile daha iyi sonuçlar almamızı sağlar.

Hira Cansu Değirmenbaşı
191101053