

YAP 470 / BIL 570 - ÖDEV 1

24 Ekim 2022

1 Ödev İçeriği

IBM firmasında çalışan veri bilimciler tarafından yaratılan çalışanların işten ayrılması ve performansları ile ilgili veri seti üzerinde model eğitmeniz beklenmektedir.

Ödev ile ilgili yapılması gereken adımlar aşağıda listelenmiştir.

1. "IBM HR Analytics Attrition Dataset" isimli veri seti Kaggle sitesi üzerinden indirilir ¹.
2. Deneylerin tekrarlanabilirliği için SEED = 12345 olarak seçilmelidir. Bu adım **zorunlu!** [0p]
3. Exploratory Data Analysis adımı veri analizi, görselleştirmeleri gerçekleştirilmelidir. [10p]
4. Veri ön işleme adımı ile gerekli normalizasyon, öznitelik dönüşümleri yapılmalıdır. [10p]
5. Veri uygun şekilde eğitim, validasyon, test kısımlarına ayrılmalı ve K=5 olacak şekilde cross-validation uygulanmalıdır. [5p]
6. Logistic Regression, SVM, Random Forest ile modeller eğitilmeli ve sonuçlar, en az dört farklı metrik ile raporlanmalıdır. [25p]
7. Eğitilen modellerin overfit olup olmadığı kontrol edilmelidir. [10p]
8. Modeller arasında hem sınıflandırma başarımı hem de eğitim ve test süreleri karşılaştırılmalıdır. [10p]
9. Parametre optimizasyonu gerçekleştirilmeli ve bu parametre optimizasyonun başarıma etkisi gösterilmelidir. [10p]
10. En iyi 5 öznitelik istenilen yöntem ile bulunmalı ve bu 5 öznitelik ile daha önceki adımda en iyi sonucu veren model tekrar eğitilerek başarımlar, ve eğitim/test süreleri karşılaştırılmalıdır. [10p]

¹<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>

11. Tüm deneyler boyunca elde edilen en iyi model ve verinin ön işleme adımları kaydedilmelidir. (.pickle olarak) [0p] **zorunlu**
12. ogrencino_test.ipynb dosyası içerisinde kaydedilen veri ön işleme adımı ve eğitilen model ile örnek bir test işlemi gerçekleştirilmelidir. [0p] **zorunlu**
13. Bulgular bir sayfayı geçmeyecek şekilde dökümanda özetlenmelidir. (**Zorunlu adım!**) [10p]

Kullanılabilecek kütüphaneler aşağıdakiler ile sınırlıdır.

- scikit-learn
- panda
- numpy
- os
- matplotlib
- seaborn

2 Teslim Kuralları

Ödev için teslim edilmesi gereken dosyalar ve dosya isimleri ve şekli aşağıda belirtilmiştir.

- Mail başlığı ODEV1, ve ekte "ogrencino.zip" dosyası olacak şekilde **batuhanbardak@etu.edu.tr** ve **m.turhan@etu.edu.tr** adreslerine mail atılmalıdır.
- Ekte yer alacak .zip dosyası içerisinde bulunması gereken dosyalar:
 1. ogrencino.ipynb (ödevde istenen adımlar bulunmalıdır)
 2. ogrencino.html (ogrencino.ipynb dosyasının html'e dönüştürülmüş hali)
 3. requirements.txt (kütüphane ve versiyonlarını içeren dosya, pip freeze ile requirements.txt)
 4. ogrencino.pdf (1 sayfalık bulguları içeren rapor)
 5. model.pickle (elde edilen en iyi modelin binary dosyası)
 6. preprocessing.pickle (ön işleme adımlarını içeren, binary dosya)
 7. ogrencino_test.ipynb (model.pickle, preprocessing.pickle dosyalarını yükleyerek örnek test işlemini içeren dosya)
 8. ogrencino_test.html (ogrencino_test.ipynb dosyasının html'e dönüştürülmüş hali)
- Ödevde Python 3.8 versiyonu kullanılmalıdır.

- Ödev aşağıdaki durumlarda değerlendirilmeye alınmayacaktır.
 1. Yukarıda belirtilen formata uygun olmayan,
 2. ogrencino.ipynb ve test.ipynb içerisindeki kodlarda hata bulunan, çalışmayan vb.
- Ödevin geç gönderiminde, ilk gün -10, ikinci gün -25, üçüncü gün -50 puan ceza uygulanır. Daha sonra gönderilen ödevler kabul edilmez.
- Ödevin son teslim süresi **7 Kasım 2022 23:59**'dur.