

Fakülte-Okul: Lisansüstü Eğitim Enstitüsü	Akademik Yıl: 2023-2024	Eğitim Dönemi: x Güz <input type="checkbox"/> Bahar <input type="checkbox"/> Yaz
Dersin Kodu ve Adı: YZM 511 İstatistiksel Yapay Öğrenme	Öğretim Üyesi: Emre Topalgökçeli	
Sınav Türü: <input type="checkbox"/> Vize <input type="checkbox"/> Final x Diğer	Sınav Tarihi: 22/02/2024	Sınav Süresi: dakika
<p>*Sınav süresince cep telefonu, bilgisayar, tablet vb. iletişim araçlarını yanınızda bulundurmuyunuz.</p> <p>*Sınav kağıdınızın değerlendirilmesi için soru ve cevap kağıtlarındaki kişisel bilgilerinizin (Ad, Soyad, Öğrenci No, Bölüm/Program) tam olması gerekmektedir. <u>İmzası olmayan sınav kağıtları geçersizdir.</u></p> <p>*Sınavda kopyaya teşebbüs etmek, kopya çekmek veya çektirmek, tehditle kopya çekmek, kopya çeken öğrencilerin sınav salonundan çıkarılmasına engel olmak, kendi yerine başkasını sınava sokmak veya başkasının yerine sınava girmek kesinlikle yasaktır. Tespiti hâlinde Yükseköğretim Kurumları Öğrenci Disiplin Yönetmeliği'nin ilgili maddelerine göre işlem yapılacaktır.</p> <p>*Size verilen zamandan başka herhangi bir ilave zaman için talepte bulunmayınız. Sınavın ilk 15 dakikasından sonra sınav salonuna girilemez. Sınavın ilk ve son 15 dakikasında sınav salonundan dışarı çıkmak yasaktır. Başarılar Dileriz.</p> <p>Sınav kurallarını OKUDUM ve ANLADIM.</p>		
Öğrencinin Adı-Soyadı:	Öğrenci No:	
Bölüm-Program:	İmza:	

SORULAR

Sınavınız 2 sorudan oluşmaktadır. **Tüm soruları cevaplamanız beklenmektedir.** Cevaplarınızı Python, R veya Stata'dan hangisini tercih ediyorsanız o dilde verebilirsiniz.

Lütfen, her bir adım için yapmış olduğunuz yorumları, ilgili kod bloklarının hemen yanında ve başında (Python ve R'de çalışacaklar için) # veya (Stata'da çalışacaklar için) * işareti kullanarak ekleyiniz. Bu, kodunuzun okunabilirliğini ve anlaşılabilirliğini artıracaktır.

Lütfen cevap kağıtlarınızı resmi cevap kağıdı şablonuna elle yazarak Enstitü tarafından belirlenen yerde ve saatte imza karşılığı teslim ediniz, ayrıca cevap kağıdınızın varsa bilgisayar ortamında yazılmış kopyasını yoksa imza karşılığı teslim ettiğiniz halini en geç sınav saatine kadar emre.topalgokceli@nisantasi.edu.tr adresine gönderiniz. Lütfen unutmayın, e-posta ile iletmeniz fiziki teslimat yükümlülüğünüzün yerine geçmez.

Hatırlatma I: Sınavda ve derslerde öğrendiğiniz yöntem ve teknikleri kullanmanız önemle rica olunur. Başlangıç seviyesinde beklenenin ötesinde tekniklerin kullanılmasının uygun olmadığını lütfen unutmayınız. Soruları kendi bilgi ve becerilerinizle çözmeniz beklenmektedir. Gerek görülen öğrenciler, 24 Şubat 2024 tarihlerinde, kodlarını çalıştırarak sınav kağıtlarını mülakat yoluyla açıklamaya davet edilebilir.

Hatırlatma II: Cevaplarınızın hangi soru ya da sorulara ait olduğunu belirtmezseniz, değerlendirmeye alınamayacağını unutmayın.

Başarılar dilerim.

SORU 1 (50 puan) emlak.csv isimli veri setini kullanarak bir lineer regresyon modeli oluşturmanız, eğitmeniz ve tahminler yapmanız beklenmektedir. Lütfen aşağıdaki adımları takip edin:

- Veri setinizdeki değişkenleri inceleyiniz, bağımsız ve bağımlı değişken olarak ayırınız.
- Bağımsız değişkenlerinizi 1'e normalize edin. Normalizasyonu gerekli gördüğünüz değişkenler ve seçiminizin nedenleri hakkında kısa bir açıklama yapın.
- Scikit-learn kullanarak bir lineer regresyon modeli oluşturun. Modelinizle tahmin yapın ve ortalama mutlak hata (MAE) değerini hesaplayın.
- Veri setinizi eğitim ve test alt kümelerine ayırın. Ayırırken ihtiyaç duyacağınız oranı, öğrenci numaranızdaki rakamların toplamını yüzde cinsinden ifade ederek belirleyin. Örneğin, öğrenci numarası "2022150099" olan bir öğrenci için:

Rakamların toplamı: $2 + 0 + 2 + 2 + 1 + 5 + 0 + 0 + 9 + 9 = 30$

Test seti oranı: %30 (yani 0.30 olarak kullanılır).

- Scikit-learn kullanarak bir lineer regresyon modeli oluşturun. Modelinizi eğitim verileri üzerinde eğitin ve test verileri üzerinde tahminler yapıp ortalama mutlak hata (MAE) değerini hesaplayın.
- Adım c ile e'de elde ettiğiniz katsayıları ve hata değerlerini içeren karşılaştırmalı bir DataFrame oluşturun. Beklenen çıktı aşağıdaki gibidir.

	Katsayi_model1	Katsayi_model2
OrtGel	6.332140	6.337709
Bina_Yasi	0.481225	0.489885
OrtOda	-15.139162	-14.371667
Ort_Y	21.760216	20.610196
Nufus	-0.141874	-0.188255
Ort_Hane	-4.705313	-4.297939
Enlem	-3.964568	-3.959521
Boylam	-4.362518	-4.359519
hata_model1	0.531164	NaN
hata_model2	NaN	0.539517

İpuçları:

- MinMaxScaler sınıfını kullanarak bağımsız değişkenleri normalize edebilirsiniz.
- train_test_split fonksiyonu ile veriyi bölebilir ve eğitim verisini elde edebilirsiniz.
- LinearRegression sınıfını kullanarak bir lineer regresyon modeli oluşturabilirsiniz.
- mean_absolute_error fonksiyonu ile ortalama mutlak hata hesaplayabilirsiniz.
- Veri setindeki bağımlı değişkeni yukarıdaki DataFrame'e bakarak çıkarmanız beklenmektedir.

SORU 2 (50 puan): veri_seti_soru_2 adlı veri setini kullanarak aşağıda belirtilen algoritmalarla sınıflandırma modelleri oluşturmanız, bu modelleri eğitmeniz ve tahminler yaparak sonuçları toplu olarak raporlamanız beklenmektedir.

- Veri ön işleme için gerekli işlemleri (normalize etme gibi.) yapınız.
- Veri setinizi eğitim ve test alt kümelerine ayırınız. Ayırırken ihtiyaç duyacağınız oranı, **egitim_test_orani_belirleme** isimli aşağıda yer alan fonksiyonu kullanarak belirleyiniz. Fonksiyonu çalıştırmadan önce kendi soy isminizi yazmayı unutmayınız.

```
# Fonksiyon
def egitim_test_orani_belirleme(soyisim):
    import random
    deger= len(soyisim)
    random.seed(deger)
    egitim_test_orani = round(random.uniform(0.1, 0.8), 2)
    return egitim_test_orani

# Örnek kullanım:
soyisim = "Topalgokceli"

print(f"Sayın {soyisim}, eğitim-test oranınız:",
      egitim_test_orani_belirleme(soyisim))
```

Yukarıdaki örnek için beklenen çıktı:

Sayın Topalgokceli, eğitim-test oranınız: 0.43

- Support Vector Machine Modeli kullanarak tahmin yapınız ve accuracy değerlerini elde ediniz.
- Random Forest Modeli kullanarak tahmin yapınız ve accuracy değerlerini elde ediniz.
- Gradient Boosting Modeli kullanarak tahmin yapınız ve accuracy değerlerini elde ediniz.
- Decision Tree Modeli kullanarak tahmin yapınız ve accuracy değerlerini elde ediniz.
- Elde ettiğiniz accuracy değerlerini ilgili model isminin karşısında olacak şekilde büyükten küçüğe doğru ve aşağıdaki örnekte olduğu gibi bir DataFrame içinde sıralayınız.

Model	Accuracy
Random Forest	0.566667
Gradient Boosting	0.488889
Logistic Regression	0.455556
Support Vector Machine	0.444444
Decision Tree	0.366667
KNN	0.355556

- Oluşturduğunuz modellerden Random Forest modelini joblib formatında kaydederek dışa aktarınız ve ödevin eki olarak gönderiniz.

İpuçları:

- from sklearn.svm import SVC
- from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier