

BİL3013 Veri Madenciliğine Giriş – Ödev
Dersin Öğr. Elemanı: Doç. Dr. Mete Eminağaoğlu

Ödevin son teslim tarihi: 5 Aralık 2022 Pazartesi, 22:00 (TSİ)

(Ödev teslimi için **ek süre kesinlikle verilmeyecektir**. Herhangi bir nedenle **zamanında iletilmeyen ödevler, hiçbir mazeret kabul edilmeden 0 (sıfır) olarak değerlendirilecektir.**)

DİKKAT: Bu ödev **tek kişiliktir**. Her bir öğrenci, bu ödevi **tek başına** yapmak zorundadır. En ufak bir yardım, vb. **kopya / intihal olarak** değerlendirilecektir ve **yardım veren / alan öğrencilerin hepsi** bu ödevden **0 (sıfır) alacaktır**.

GENEL HATIRLATMA:

Aşağıdaki yazılanları baştan sona okumanız ve eksiksiz uymanız zorunludur. Ödevle ilişkin kodlamayı tam ve hatasız yapıp, zamanında iletip, hatasız çalıştırsanız bile, aşağıdaki yazılı kural, talimat, vb. bir tanesine bile uyulmazsa, ödevinizden puan / puanlar kırılacaktır.

Ödevin Teslim Şekli:

DEÜ Sakai sistemindeki ders sayfasında açılacak olan ödev yükleme (assignment) alanına; tüm dosyalar (tüm kaynak kodları, bağlantılı kütüphane, dizinler, proje dosyaları, vb.), **zip / rar şeklinde sıkıştırılmış tek bir dosya olarak** yüklenecektir.

Ödev Konusu:

Sakai sistemindeki Assignments - Ödev kısmında size iletilmiş olan “trainSet.csv” adlı dosyada bulunan eğitim (train) ve “testSet.csv” adlı dosyada bulunan test verileri kullanılarak **Naive Bayes** algoritması kodlanacak ve ikili sınıflandırma yapılacaktır.

- **Veri Seti ile ilgili bilgiler:**
 - Her değişken, (virgül işareti) ile ayrılmıştır. Bazı değişkenler kayıp değer içermekte olup bu değerler ? ile gösterilmiştir.
 - En sondaki “class” adlı değişken, ikili sınıflandırmaya ilişkin sınıf değişkenidir. “good” etiketi pozitif (birincil öncelikli), “bad” etiketi negatif (ikincil öncelikli) olanları gösterir.
 - Eğitim setinde 750 adet, test setinde 250 adet kayıt bulunmaktadır.
- **Test verisinin sınıflandırmasının **Accuracy, True Positive Rate, True Negative Rate, True Positive Adedi, True Negative Adedi** ölçümleri ve hesaplamalarını öğrenciler kendileri kodlayacaktır. Bu işlemler için hazır fonksiyon, kütüphane, hazır araç, vb. kullanımı kesinlikle yasaktır. Kullanılması durumunda ödev notundan 40 puan düşürülür.**
- Öte yandan dosya açma, veri okuma, Naive Bayes algoritması kullanımı kısımlarında hazır araç, kütüphane, vb. kullanabilirsiniz.
- Kodlama kısmında, **sadece aşağıdaki programlama dillerinden istediğiniz bir tanesini** seçip kullanabilirsiniz: **C, C++, C#, .Net, Java, Python.**

Program, önce eğitim dosyasını açarak karar ağacını oluşturacak, sonra da oluşan karar ağacı modeline göre, test veri setini kullanarak **Accuracy, True Positive Rate, True Negative Rate, True Positive Adedi, True Negative Adedi** ölçümlerinin sonucunu ekrana aşağıdakine benzer biçimde yazdıracaktır.

Test sonucu:

Accuracy (Doğruluk):	0.753
TP rate (Gerçek Pozitif oranı):	0.908
TN rate (Gerçek Negatif oranı):	0.322
TP (Gerçek Pozitif) adedi:	227
TN (Gerçek Negatif) adedi:	29

Ödevde Teslim Edilecekler:

- 1-Programın tüm kaynak kodları, bağlantılı kütüphane, dizinler, vb.
- 2-Kullanılan yöntemler, işlemler, vb. ile ilgili kısa bilgiler / notlar (kaynak kod içine kısa açıklamalar olarak eklenmelidir).