

BİL 3013 Veri Madenciliğine Giriş – Final Sınavı (Proje)

Doç. Dr. Mete Eminağaoğlu

Proje için son teslim tarihi: 13 Ocak 2023 Cuma, 20:00

(ek süre kesinlikle verilmeyecektir. Herhangi bir nedenle zamanında iletilmeyen projeler, hiçbir mazeret kabul edilmeden 0 (sıfır) olarak notlandırılacaktır.)

Sınavın Teslim Şekli:

DEÜ Sakai sistemindeki ders sayfasında açılacak olan ödev yükleme (assignment) alanına; tüm dosyalar, rapor, vb. **zip / rar sıkıştırılmış tek bir dosya olarak yüklenecektir.**

Bu sınav tek kişiliktir. En ufak bir yardım, Internet'ten kod kopyalanması, vb eylemler kopya / intihal olarak değerlendirilecektir ve bunu yapan öğrenciler sınavdan ya da sınavın ilgili bölümünden **0 (sıfır) alacaktır.**

Bu final sınavının olduğu “Ödevler” kısmında yer alan **veri kümesini (“Final-data.csv”)** kullanılarak (toplam 214 kayıt ve 9 sayısal değişken), **k-means** algoritması ile kümeleme (clustering) analizi yapan bir program geliştireceksiniz. **Ekrandan girilen “k” değerine göre**, k adet kümeye ayıracak ve veri dosyasındaki her bir kaydı ilgili kümeye hesaplayarak atayacaktır.

Dikkat: Kümeleme işlemi öncesi, tüm değişkenlerde normalizasyon işlemi yapılacaktır.

Programınız her bir kaydın hangi küme (cluster) ‘a atandığı ve genel sonucu programınız tarafından oluşturulacak bir **“sonuc.txt” dosyasına kaydedecektir. Ayrıca, bu dosyanın sonunda WCSS (Within Cluster Sum of Squares), BCSS (Between Cluster Sum of Squares) ve Dunn Index değerleri de hesaplanıp yazılacaktır.**

Örnek bir “sonuc.txt” dosyası içeriği (ekrandan k = 3 girildiği varsayılarak) aşağıda verilmiştir (sayılar gerçek değildir, fikir vermesi için yazılmıştır).

Kayıt 1: Küme 1
Kayıt 2: Küme 1
Kayıt 3: Küme 2
...

Kayıt 121: Küme 3
Kayıt 122: ...
...
Kayıt 214: Küme 2

Küme 1: 158 kayıt
Küme 2: 35 kayıt
Küme 3: 21 kayıt

WCSS: 16.24
BCSS: 286.45
Dunn Index: 10.73

Programınızda **“veri görselleştirme”** kısmı da olacaktır. **“veri görselleştirme”** adlı bir seçenek / tuşa tıklandığında, k-means ile elde edilen kümelerin ekrana çizilebilmesi için x eksenini ve y eksenine hangi değişkenlerin geleceği seçilecek (veri setindeki “a1,a2,a3,a4,a5,a6,a7,a8,a9” değişkenlerinden ikisi), sonra da buna göre ekranda tüm kayıtlar ait olduğu küme farklı görsel (farklı renk veya şekil) olacak şekilde grafik çizimi yapılacaktır.

- (Her ne kadar önerilmese de), hazır kütüphane, hazır araç, vb. öğrenciler eğer istiyorsa kullanabilir.

- Kodlama kısmında, sadece, **aşağıdaki programlama dillerinden birini kullanabilirsiniz, aşağıdakiler harici bir dil kullanılamaz:**
C, C++, C#, .Net, Java, Python.

Projede Teslim Edilecekler:

- 1-Programın tüm kaynak kodları, bağlantılı kütüphane, dizinler, vb.
- 2-Kullanılan algoritmalar, vb. ile ilgili kısa bilgiler / notlar (istenirse kaynak kod içine de açıklamalar olarak eklenebilir).
- 3- sonuc.txt dosyası