**Konya Teknik Üniversitesi**

**Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı**

**Veri Ön İşleme Teknikleri Dersi Uygulama Ödevi**

**Öğrenci No :**

**Ad Soyad :**

**Size gelen E-postada belirtilen X veri kümesi ve sadece belirtilen nitelikler (attributes) için aşağıda istenenleri gerçekleştiriniz.**

**1.** X veri kümesinin e-postada belirtilen her bir niteliği (**attribute**) için *beş sayı özeti* (**five number summary**), *ortalama* (**mean**), *mod* (**mode**), *çeyrekler arası genişlik* (**InterQuartile Range**), *varyans* (**variance**), *standart sapma* (**standard deviation**) ve varsa *aykırı* (**outlier**) değerlerini hesaplayınız. Her bir nitelik için *kutu diyagramı* (**boxplot**) çizimini gerçekleştiriniz.

**2.** X veri kümesi için aşağıdaki işlemleri gerçekleştiriniz.

**a)** Veri kümesinin e-postada belirtilen her bir niteliği için ayrı ayrı [0-1] aralığına ***min-max normalizasyon*** ve ***z-score normalizasyon*** değerlerini hesaplayınız.

**b)** Ayrık olmayan nitelik değerlerini (**continuous attributes**) *n eşit genişlik* (**n equal-width**) bölme yöntemleri ile ayrık hale dönüştürünüz (n değerini kendiniz belirleyebilirsiniz). Her bir yöntem sonucu elde edilen verilerin her bir niteliği için frekans tabanlı histogram grafiklerini çiziniz. Veri kümenizde belirtilen tüm nitelikler ayrık (discrete) ise bu seçeneği ihmal ediniz. Sadece sürekli değerli nitelikler için gerçekleştirilecektir.

**3.** X veri kümesi için e-postada belirtilen her bir niteliğin bilgi kazancı (**information gain**) değeri hesaplanacaktır. Sürekli değerli (Ayrık olmayan, Continuous) olan nitelikler varsa aşağıdaki işlemleri gerçekleştiriniz. Veri kümenizin belirtilen niteliklerinin tamamı ayrık (**discrete**) ise **a** ve **b** seçeneklerini ihmal ediniz ve direkt olarak bilgi kazancı değerlerini hesaplayınız.

**a)** Veri kümesinde ayrık olmayan nitelik değerlerini *3 eşit genişlik* (**3 equal-width**) değeri olacak şekilde parçalara ayırdıktan sonra her bir nitelik için *bilgi kazancını* (**information gain**) hesaplayınız.

**b)** Veri kümesinde ayrık olmayan nitelik değerlerini *4 eşit genişlik* (**4 equal-width**) değeri olacak şekilde parçalara ayırdıktan sonra her bir nitelik için *bilgi kazancını* (**information gain**) hesaplayınız.

**Not:** Ödev raporu ve tüm uygulamalar RAR formatında sıkıştırılarak hkodaz@ktun.edu.tr adresine e-posta yoluyla gönderilecek ve lms.ktun.edu.tr adresi üzerinden dersin ödev linkinden sisteme yüklenecektir. LMS üzerinden son teslim tarihi 18 Ocak 2021 pazartesi günü saat 23:55'tir.