

1- T1=[88, 66, 95, 100, 105, 78, 20, 210, 166, 102, 75, 71, 80, 89, 121] veri setini paketleme (binning) yöntemi (3 paket, eşit derinlik, ortanca) ile düzleştirildiğinde 89'un yeni değeri ne olur?

a.89 b.90.4 c.95 d.97.5 e.100

2- Aşağıdakilerden hangisi Gözetimli öğrenmedir?

- a.Harcama alışkanlığı benzer olan kredi kartı sahiplerini bulmak.
- b.Doküman içinde geçen terimlere göre aynı konudaki dokümanları bulmak.
- c.Kızılderili toplumundan alınan örneklerde bulunan çeşitli HIV risk gruplarının belirlenmesi.
- d.Sahtekarlık olabilecek davranışları bulmak.
- e.Yazıcı ile birlikte en sık satılan ürünleri bulmak

3- Hangi veri çeşiti içerisinde diğerlerinden daha fazla bilgi barındır?

a.Ratio b.Nominal c.Ordinal d.Categorical e.Interval

4- Aşağıdakilerden hangisi olasılık tabanlı bir sınıflandırma yöntemidir?

a.K-Nearest Neighbor b.Decision Tree c.CART d.SVM e.Naive Bayes

5- V1=011001101010

V2=110101101101

V1 ve V2 ikili vektörleri için Jaccard katsayısı benzerlik ölçütü aşağıdakilerden hangisidir?

a.0.3 b.0.6 c.0.5 d.0.2 e.0.4

6-

T2			
No	X1	X2	Class
1	3	6	Yes
2	2	8	No
3	4	2	Yes
4	5	9	No
5	1	5	No
6	6		Yes
7		7	No
8	4	11	No

Tablodaki T2 veri kümesinde eksik veriler "Sınıf bazında ortalama" yöntemi ile doldurulacak olursa sırasıyla X1 ve X2 niteliklerinin eksik verileri ne olur?

- a.3.57 ve 6.86
- b.3.57 ve 4
- c.3 ve 4
- d.4 ve 3
- e.4 ve 1.6

7- Rastgeleliğin ve belirsizliğin ölçütü anlamına gelen terim aşağıdakilerden hangisidir?

- a.Information Gain **b.Entropy** c.Pruning d.Greedy e.Gini Index

8-

Gözlem	Nitelik1	Nitelik2
X1	4	2
X2	6	4
X3	5	1
X4	10	6
X5	11	8

Tablodaki veri kümesinden K-means yöntemi ($k=2$) ile $C1=(X1,X2,X4)$ ve $C2=(X3,X5)$ olarak iki küme oluştuğunda $C1$ kümesinin küme içi hatası (e^2) ne olur?

- a.42.5 b.69.17 c.9.33 **d.26.67** e. 11.83

9- Bir firmanın 2006-2020 arası yıllık ortalama satış rakamları $T1 = [88, 66, 95, 100, 105, 78, 20, 210, 166, 102, 75, 71, 80, 89, 121]$ olarak verilmiştir. $T1$ veri setinin IQR yardımı ile bulunan sapan verileri nelerdir?

- a.210 b.20,210 c.166,210 d.Sapan veri yoktur. **e.20, 166, 210**

10- $V1=011001101010$

$V2=110101101101$

$V1$ ve $V2$ ikili vektörleri için Yalın Uyum katsayısı benzerlik ölçütü aşağıdakilerden hangisidir?

- a.0.6 b.0.4 c.0.2 **d.0.5** e.0.3

11- “Araç özellikleri kullanılarak fiyat tahminlemesi yapmak” problemidir?

- a.Regresyon** b.Sınıflandırma c.Sorgulama d.Kümeleme e.Birliktelik Analizi

12- Analiz edilmek istenilen verilerdeki beklenen değerlerden sapan aykırı değerlere veya hatalara aşağıdakilerden hangi ad verilir?

- a.Büyük veri b.Eksik Veri c.Gereksiz veri d.Tutarsız veri **e.Gürültü**

13- Aşağıdakilerden hangisi makine öğrenmesi sınıflandırma yöntemi değildir?

- a.Destek vektör makineleri b.Naif Bayes Yöntemi c.Karar Ağaçları
d.Yapay sinir ağları **e.K-ortalamlar Yöntemi**

14- Aşağıdakilerden hangisi bilgi keşfi çalışmalarında bulunan bilginin sağlaması gereken özelliklerden değildir?

- a.Önceden bilinmeyen b.Basit c.Önemli d.Yararlı e.Gizli

15- Aşağıdakilerden hangisi eksik veri problemini çözen veri ön-işleme yöntemlerinden değildir?

- a.Normalizasyon ile tamamlama
b.Karar ağacı ile tamamlama
c.Ortalama ile tamamlama
d.ihmal etme
e.Regresyon eğrisi ile tamamlama

16-Aşağıdakilerden hangisi veri ön-işleme süreçlerinden değildir?

- a.Veritoplama b.Veribirleştirme c.Veridönüştürme d.Veritemizleme e.Verindirgeme

17- Aşağıdakilerden hangisi Karar Ağaçları için doğru değildir?

- a.Ağaç yaprakları sınıf etiketlerini içerir.
b.Ağaçtaki her düğüm bir nitelikteki testi gösterir.
c.Çoğunlukla farklı sınıfa ait örneklerin bulunduğu nitelikler düğüm olarak seçilir.
d.Ağacın inşasında örnekler seçilmiş özelliklere göre tekrarlamalı olarak bölünür
e.Ağaç veri setindeki örnekleri en iyi ayıran niteliği ifade eden düğüm ile başlar

18- "Farklı kümelerdeki herhangi iki eleman arasındaki en büyük uzaklık" hangi kümeler arası uzaklık ölçme metodudur?

- a.Average distance
b.Single linkage
c.Average linkage with Centroid
d.Average linkage with Medoid
e.Complete linkage