A 32

Data Mining

Sol

Quiz No.1

Roll No:_____

Section:_____

Date: 12-02-2024

Question No.1

(4 Marks)

Discuss whether or not each of the following activities as a data mining task. If the answer is yes then also specify which one of the following categories it will belong to

i. Classification
ii. Association analysis
iii. Clustering
iv. Regression
v. Anomaly detection

a) Sorting a student database based on student identification number.

No

b) By looking at a CT scan, a doctor wants to identify if a patient has a cancer or not. There are a lot of labelled CT scans that the doctor will use for making the decision.

Yes, classification

c) Extracting the frequencies of a sound wave.

Yes, but none of These Feature extraction Task

d) Predicting the future stock price of a company using historical data.

Yes, Regression.

Question No.2

(4+4+4=12 Marks)

Suppose that the data for analysis include the attribute age. The age values for the data tuples are (sorted)

13, 15 ,16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70

i. What is the mean value of the data? What is mode of the data? Comment of data modality as (bimodal, trimodal, etc...)
ii. Can you find the first quartile(Q1) and the third quartile(Q3) of the data?
iii. Calculate variance of the above data

Roll N

Que

Disc
also

I   Mean $= 809/27 \Rightarrow \boxed{29.96}$

Mode $= 25, 35$

$\overline{Data\ modality}$ Bimodal

II

1st Quartile $= (n+1) \times \dfrac{1}{4} = 7^{th}$ value $= \boxed{20}$

3rd Quartile $= (n+1) \times \dfrac{3}{4} = 21^{th}$ value $= \boxed{35}$

III variance $= \dfrac{\sum (x - \bar{x})^2}{n-1}$

$= \boxed{185} \boxed{167.5}$

$-16.96 - 14.96 - 13.96 - 13.96 - 10.96 - 9.96 - 9.96 - 8.96 - 7.96 - 7.96 - 4.96 - 4.96 - 4.96$

$-4.96 + 3.04 + 3.04 + 5.04 + 5.04 + 5.04 + 5.04 + 6.04 + 10.04 + 15.04 + 16.04 + 22.04$

$40.04 / 287.64 + 223.8016 + 10.$