Jaypee Institute of Information Technology, Noida
T1 Examination, EVEN 2022-23
B.Tech (Information Technology) Semester-VI

Course Title: Data mining and Web Algorithm
Course Code: 15B22CI621

Maximum Time: 1 Hour
Maximum Marks: 20

After pursing this course, the students will be able to:

COURSE OUTCOMES		
C401.12.1	Understand the basis of data mining and pre-processing of data	Understand Level (level 2)
C401.12.2	Analyse the transactional data for finding frequent and interesting patterns using association rule mining techniques like Apriori and FP growth	Analyse Level (level 4)
C401.12.3	Apply a wide range of classification techniques like Naïve Bayes, decision tree and KNN for the numerous applications including fraud detection, target marketing. Medical diagnosis etc.	Apply Level (level 3)
C401.12.4	Cluster the similar /dissimilar objects using different methods like partitioning, hierarchical and density-based clustering.	Create Level (level 6)
C401.12.5	Analyse the link structure of web using page rank and HITS algorithms.	Analyse Level (level4)
C401.12.6	Develop recommendation systems using collaborative filtering techniques.	Create Level (level 6)

Note: Attempt all the questions:

Q1. A Market research for the real estate investments was carried on to reveal the sales figure of new houses of different prices.

Sales of new Houses	125	20	104	85	40	80	75
Price (Thousands of $)	160	280	180	200	260	240	220

Answer the following:

a. Predict the sales of new houses if the price is 180 thousand dollars using regression. (3 Marks, CO1)

b. Do the sales depend on the price and if yes how? Justify the statement with a proper statistical value (2 Marks .CO1)

Q2. Calculate the Z-score for a distribution with mean of 150 and standard deviation of 15 for the following raw score: (2 Marks, CO1)

a.135 b.140 c.170 d.185

Also, what is the advantage of Z score normalization over min Max normalization?

Q3. A random sample of 160 car accidents is selected and categorized by the age of the driver determined to be at fault. The results are listed below. The age distribution of the drivers for the given categories is 18% for the under 26 group, 39% for the 26-45 group, 31% for the 45-65 group, and 12% for the group over 65. Calculate the Chi-square statistic to test the claim that all ages have the crash rates proportional to the driving rate. (4 marks, CO1)

Age	>26	26-45	46-65	65<
No of Drivers	66	39	25	30

$X^2_{0.05}=7.814728$

Q4.

TID	Items
1	a, b
2	c, b, d
3	a, e, c, d
4	a, d, e
5	b, c, a
6	d, c, b, a,
7	a
8	a, b, c
9	b, a, d
10	c, b, e

For the above transactional dataset, calculate the following:

a. Find the Association rules, for the above transactions with minimum support threshold 30% and minimum confidence threshold 80%. (3 Marks, CO2)

b. Using FP Growth, find the frequent Itemsets for the given transactional dataset. (3Marks, CO2)

c. Using ECLAT, only find the frequent Itemsets. (3 Marks, CO2)