


Name: Enrolment No:	
--------------------------------------	--

UPES
End Semester Examination, December 2023

Course: Data Warehouse and Data Mining
Semester: VII
Program: B.Tech - CSE
Course Code: Minor

Time : 03 hrs.
Max. Marks: 100

Instructions:

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Define data mining? In your answer, address the following: (a) Is it another hype? (b) Is it a simple transformation of technology developed from databases, statistics, and machine learning? (c) Explain how the evolution of database technology led to data mining. (d) Describe the steps involved in data mining when viewed as a process of knowledge discovery.	4 marks	CO1
Q 2	(1) Movie Recommendation system is an example of: Classification Clustering Reinforcement Learning Regression (a) 2 only (b) 1 and 2 (c) 1 and 3 (d) 2 and 3 (e) 1, 2 and 3 (f) 1, 2, 3, and 4 (2) Sentiment Analysis is an example of: Classification Reinforcement Learning Clustering Regression (a) 1 only (b) 1 and 2 only (c) 1, 2 and 3 (d) 1, 2 and 4	4 marks	CO4

	<p>(e) 1,2, 3 and 4</p> <p>(3) Can Decision Trees be used for performing Clustering? (a) True (b) False</p> <p>(4) What is the minimum number of variables/features required to perform clustering? (a) 0 (b) 1 (c) 2 (d) 3</p>		
Q 3	Explain Supervised and Unsupervised learning techniques.	4 marks	CO2
Q 4	Comparison between OLAP and OLTP on features (i) Database design (ii) Characteristic (iii) Orientation (iv) User	4 marks	CO1
Q 5	Describe ETL process.	4 marks	CO2
SECTION B (4Qx10M= 40 Marks)			
Q 6	Assume your task as a software engineer at <i>Big University</i> is to design a data mining system to examine the university course database, which contains the following information: the name, address, and status (e.g., undergraduate or graduate) of each student, the courses taken, and the cumulative grade point average (GPA). Describe the <i>architecture</i> . you would choose. What is the purpose of each component of this architecture?	10 marks	CO2
Q 7	Assume that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. (a) What is the mean of the data?What is the median? (b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.). (c) What is the midrange of the data? (d) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data? (e) Give the five-number summary of the data. (f) Show a boxplot of the data. (g) How is a quantile-quantile plot different from a quantile plot?	10 marks	CO3

Q 8	Briefly outline the major steps of K-NN. Illustrate with example.	10 marks	CO4
Q 9	Explain Datawarehouse, its components with examples.	10 marks	CO2
SECTION-C (2Qx20M=40 Marks)			
Q 10	<p>(a) Suppose that in a set of examples there are two classes, with 150 examples in the + class and 50 examples in the – class. What is the entropy of the class variable?</p> <p>(b) For this data, suppose the Color attribute takes on one of 3 values (red, green, and blue), and the split into the two classes across red/green/blue is + : (120/10/20) and – : (0/10/40). Write down an expression for the class entropy in the subset containing all green examples. Is this entropy greater or less than the entropy in the previous question?</p> <p>(c) Is Color a good attribute to add to the tree? Explain your answer</p> <p>(d) What is the information gain for a particular attribute if every value of the attribute has the same ratio between the number of + examples and the total number of examples?</p>	20 marks	CO5
Q 11	<p>(a) Explain the difference between biological neuron and ANN with diagrams.</p> <p>(b) Define epoch, iterations, batches, learning rate, loss function in neural network with example.</p>	20 marks	CO5