Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S8 (S) Examination September 2025 (2019 Scheme)

Course Code: CST466

Course Name: DATA MINING

Max. Marks: 100 **Duration: 3 Hours**

PART A

Answer all questions, each carries 3 marks. Marks

1		List out any three applications of data mining in day-to-day life.	(3)
2		Explain Roll up and Drill down with suitable examples..	(3)
3		Perform data smoothing by bin boundaries on 3 equi-width bins. Data: [24,27,29,16,17,31,33,29,36,37,35,44]	(3)
4		Explain the role of concept hierarchy.	(3)
5		Given two objects are represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8). Compute(i)Euclidean distance ii)Manhattan distance.	(3)
6		For a spam email classifier, is precision or recall more important evaluation measure. Justify your answer.	(3)
7		Describe any three methods to improve the efficiency of the Apriori algorithm.	(3)
8		Explain dynamic item set counting technique.	(3)
9		Explain any two web usage mining activities.	(3)
10		Illustrate the focused crawling and regular crawling techniques.	(3)

PART B

Answer any one full question from each module, each carries 14 marks.

Module I

11	a)	Explain the knowledge discovery process(KDD)in databases for finding useful Information and patterns in data.	(7)
	b)	Explain the three-tier architecture of the data warehouse with a neat diagram.	(7)

OR

12	a)	Consider a data warehouse to represent sales data with dimensions customer, product, date and region and two measures quantity and sales amount. Draw snow flake schema diagram for the data warehouse.	(6)

b) Illustrate different OLAP operations in multi-dimensional data model with examples. (8)

Module II

13 a) Explain the various techniques used for handling missing data. (6)

b) Discuss the significance of data discretization in data mining. List and explain any two data discretization strategies. (8)

OR

14 a) Write about any two data reduction techniques. (8)

b) Why do we need data transformation? What are the different ways of data transformation? (6)

Module III

15 a) Consider the following data set for a binary classification problem. (7)

A	B	Class label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

Calculate the gain of attributes A and B. Which attribute would be selected as root node while constructing the decision tree?

b) Explain categorical clustering algorithm or ROCK. (7)

OR

16 a) Find the first splitting attribute for the decision tree by using the ID3 algorithm with the following dataset. (8)

Age	Competition	Type	Class (profit)
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

b) Explain the working of SLIQ algorithm. (6)

Module IV

17 a) Illustrate the working of Pincer Search algorithm with an example. (8)

b) Describe partition algorithm. (6)

OR

18 a) State the Apriori principle in candidate generation. Find out the frequent item sets with minimum support of 2 using Apriori for the following data. (8)

TID	ITEMSETS
T1	A, B
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

b) Describe the working of dynamic item set counting technique with suitable example. (6)

Module V

19 a) What are the various data structures used for web usage mining process? Explain. (8)

b) Explain HITS algorithm with an example. (6)

OR

20 a) Describe different Text retrieval methods. Explain the relationship between text mining, information retrieval and information extraction. (8)

b) Write and explain CLEVER algorithm for web structure mining. (6)

****