J. THI ENGG. COLLEGE LIBRARY
THRISSUR EDUCATIONAL TRUST
CHERUTHURUTHY

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S8 (S,FE) Examinations September 2026 (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 and explain any three applications of data mining.	(3)
2		Differentiate between operational database systems and data warehouse.	(3)
3		Explain concept hierarchy with an example.	(3)
4		Calculate the normalized value for 50 in the following set of data points using min-max and z-score normalization: 10, 20, 50, 80, 100	(3)
5		Discuss the issues regarding the implementation of decision tree.	(3)
6		What are the requirements for a good clustering algorithm?	(3)
7		Define support, confidence and frequent item set in Association rule mining context.	(3)
8		Discuss the significance of association rule mining in market basket analysis.	(3)
9		Briefly explain web mining taxonomy.	(3)
10		What are the basic measures for text retrieval?	(3)

PART B

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

Module I

11	a)	Explain different OLAP operations on multidimensional data with suitable examples.	(7)
	b)	Explain three-tier data warehouse architecture with a neat diagram.	(7)
		OR	
12	a)	Explain the different data mining functions.	(7)
	b)	Suppose that a data warehouse consists of the four dimensions date, spectator, location, and game, and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors, with each category having its own charge rate. (a) Draw a star schema diagram for the data warehouse.	(7)

(b) Starting with the base cuboids [date, spectator, location, game], what specific OLAP operations should you perform in order to list the total charge paid by student spectators at GM Place in 2010?

Module II

13 a) Describe the various techniques for numerosity reduction in data mining. (8)

b) Suppose that the data for analysis includes the attribute age. The age values For the data tuples are (in increasing order): 8, 9, 15, 16, 21, 21, 24, 26, 27, 30, 30, 34 .Use smooth by bin means and bin boundaries to smooth the above data where number of bins are 4. (6)

OR

14 a) Write about any two data reduction techniques. (7)

b) Discuss the significance of data discretization in data mining. List and explain any two data discretization strategies. (7)

Module III

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

Student	Study hours	Attendance	Sleep	Internet Use	Result
1	High	Good	Good	Low	Pass
2	High	Poor	Poor	High	Fail
3	Medium	Good	Good	Low	Pass
4	Low	Poor	Good	High	Fail
5	Medium	Good	Poor	Low	Pass
6	Low	Poor	Poor	High	Fail
7	High	Good	Good	Medium	Pass
8	Medium	Poor	Good	High	Fail
9	High	Good	Poor	Low	Pass
10	Low	Good	Good	Medium	Pass
11	Medium	Poor	Poor	High	Fail
12	High	Good	Good	Low	Pass

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

OR

16 a) What are the key parameters used in DBSCAN? Explain each parameter in detail. Also write the advantages of DBSCAN compared to partition based clustering methods. (8)

b) Explain the working of PAM algorithm with an example. (6)

Module IV

17 a) Consider a dataset containing transactions from a retail store. The dataset consists of the following transactions: (8)

Transaction 1: {milk, bread, eggs}

Transaction 2: {milk, bread, butter}

Transaction 3: {bread, butter}

Transaction 4: {milk, bread}

Transaction 5: {bread, eggs}

Using the Frequent Pattern Growth Algorithm, find all frequent itemsets with a minimum support of 40%. Then, determine the association rules with a minimum confidence of 60%.

b) Demonstrate how the Pincer Search Algorithm works with an example. (6)

OR

18 a) Discuss the partitioning algorithm for finding large itemset and compare its performance with Apriori algorithm. (6)

b) Describe the working of dynamic itemset counting technique with suitable example. Specify when to move an itemset from dashed structures to solid structures. (8)

Module V

19 a) Explain various data structures used for web usage mining process. (6)

b) Describe different Text retrieval methods. Distinguish text mining, information retrieval and information extraction.. (8)

OR

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

b) Term frequency matrix given in the table shows the frequency terms per document. (8)

Document/ Terms	T1	T2	T3	T4	T5	T6	T7
d1	0	4	10	8	0	5	0
d2	5	19	7	16	0	0	32
d3	15	0	0	4	9	0	17
d4	22	3	12	0	5	15	0
d5	0	7	0	9	2	4	12

Calculate the TF-IDF value for the term T6 in document d4.

***