

Reg No.: \_\_\_\_\_

Name: \_\_\_\_\_

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**

B.Tech Degree S8 (R,S)(FT/WP) Examinations April 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  | Differentiate between Datamining and Data Warehousing.   | (3)   |
| 2  | What is data mart?   | (3)   |
| 3  | What do you mean by Data Preprocessing? Why is data preprocessing essential in data mining.                | (3)   |
| 4  | Perform data smoothing by bin means on 3 equi-width bins.<br>Data: [9, 8 ,16 , 21,15,21,24,30,26,27,34,30] | (3)   |
| 5  | What do you mean by Market Basket analysis and how it can help in a supermarket?                           | (3)   |
| 6  | What is Clustering?What are the requirements for a good clustering algorithm?                              | (3)   |
| 7  | What is the purpose of Apriori Algorithm?  | (3)   |
| 8  | Define frequent sets, confidence, support and association rule.  | (3)   |
| 9  | Define text mining   | (3)   |
| 10 | Describe any two text indexing techniques  | (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) Illustrate the various stages in Knowledge discovery process with a diagram. (7)
- b) Write about various functionalities of data mining. Describe different issues in data mining. (7)

## Module II

- 13 a) Explain Integration of datamining system with a data warehouse. (7)  
 b) What do you mean by cleaning of data? Explain various methods of data cleaning in detail. (7)

OR

- 14 a) Explain data reduction methods (7)  
 a) Data cube aggregation  
 b) Dimensionality Reduction  
 b) How to handle missing values in datamining process? (7)

## Module III

- 15 a) classifying 100 people (which includes 40 pregnant women and the remaining 60 are not pregnant women) as pregnant or not pregnant. Out of 40 pregnant women 30 pregnant women are classified correctly and the remaining 10 pregnant women are classified as not pregnant by data mining algorithm. On the other hand, out of 60 people in the not pregnant category, 55 are classified as not pregnant and the remaining 5 are classified as pregnant. Construct the confusion matrix and calculate the precision and recall scores. (7)  
 b) What is a Decision Tree? Explain the steps in creating decision tree using **ID3 algorithm** for the dataset given below, find the first splitting attribute for the decision tree by using the **ID3 algorithm**. (7)

ID	Fever	Cough	Breathing issues	Infected
1	NO	NO	NO	NO
2	YES	YES	YES	YES
3	YES	YES	NO	NO
4	YES	NO	YES	YES
5	YES	YES	YES	YES
6	NO	YES	NO	NO
7	YES	NO	YES	YES
8	YES	NO	YES	YES
9	NO	YES	YES	YES
10	YES	YES	NO	YES
11	NO	YES	NO	NO
12	NO	YES	YES	YES
13	NO	YES	YES	NO
14	YES	YES	NO	NO

OR

- 16 a) Explain DBSCAN hierarchical clustering. Write its advantages. (7)

- b) Discuss PAM(K-Medoid) algorithm for clustering with an example (7)

#### Module IV

- 17 a) A database has five transactions. Let min sup = 60% and min conf = 80%. (7)

TID and Items:

- T100 = {M, O, N, K, E, Y}
- T200 = {D, O, N, K, E, Y}
- T300 = {M, A, K, E}
- T400 = {M, U, C, K, Y}
- T500 = {C, O, O, K, I, E}

Find all frequent itemsets of a single item using the Apriori algorithm.

- b) Explain dynamic item set counting technique in association rule mining? (7)

OR

- 18 a) Explain types of association rules in data mining. (7)

- b) Generate the frequent pattern from the following data set using FP growth, (7)  
where minimum support = 2 and minimum confidence threshold as 70%

Transaction ID	List of items ids
T100	11,12,15
T200	12,14
T300	12,13
T400	11,12,14
T500	11,13
T600	12,13
T700	11,13
T800	11,12,13,15
T900	11,12,13

#### Module V

- 19 a) Explain briefly about Web mining. (7)

- b) .Discuss the applications of web usage mining. (7)

OR

- 20 a) Discuss the basic measures for text retrieval. (7)

- b) Explain web structure mining with a suitable algorithm (7)

\*\*\*\*