

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S5 (R,S) Examination November 2025 (2019 Scheme).

Course Code: CDT305**Course Name: DATA ANALYTICS**

Max. Marks: 100

Duration: 3 Hours

PART A*(Answer all questions; each question carries 3 marks)*

		Mar ks
1	List and explain any 3 data analytics application.	3
2	How can we compute the dissimilarity between two binary objects ?	3
3	State the need for data transformation.	3
4	Explain the purpose of data reduction in data preprocessing.	3
5	Illustrate a situation where two variables might be associated but not correlated linearly.	3
6	Differentiate between support and confidence.	3
7	Explain the attribute selection method in decision tree algorithm	3
8	Write the advantages of DBSCAN over K-means clustering algorithm ?	3
9	List any three application of Naïve Bayes Text classification.	3
10	What are Stop words? How a stop list is created.	3

PART B*(Answer one full question from each module, each question carries 14 marks)***Module -1**

11	a) Explain the data analytical process model.	8
	b) Distinguish between a one-tailed and two-tailed test in hypothesis testing, and give examples of situations where each would be appropriate.	6
12	a) Define Normal distribution and mention its properties.	7
	b) What are the Proximity measures of nominal attributes ?	7

Module -2

13	a) Harsha has been playing golf on the weekends for the past three years. Recently, she started keeping track of her recorded scores. Her scores for June and July at her favourite 9-hole (par 36) golf course are provided below. 45,49,42,56,41,36,34,38,41,40,42,41,39,38,40,39,36,41. Find the Mean, Median and Mode for the above data. Is this data skewed?	7
	b) Use the methods given below to normalize the following group of data: 1000,2000,3000,5000,9000	7
	i) min-max normalization by setting min=0 and max=1	

ii) z-score normalization

- 14 a) Explain dimensionality reduction in data preprocessing. 7
 b) Describe about data cleaning methods. 7

Module -3

- 15 a) A database has five transactions. Let min sup = 60% and min confidence = 50%. 10
 Find all frequent patterns using FP-growth algorithm.

Tid	Items_bought
T1000	{M,O,N,K,E,Y
T2000	{D,O,N,K,E,Y}
T3000	{M,A,K,E}
T4000	{M,U,C,K,Y}
T5000	{C,O,O,K,I,E}

Find all strong association rules for the above table.

- b) How association rule mining can be viewed in frequent item sets. 4
 16 a) How to find patterns that are rare instead of frequent. 7
 b) Explain how are metarules useful in mining of association rules. 7

Module -4

- 17 a) Explain decision tree induction with the help of an example. 8
 b) How clustering defines a group, and how such groups are identified by k-means 6
 clustering algorithm.
 18 a) Explain DBSCAN algorithm. 10
 b) What are the merits and demerits of Naïve bayes classifier ? 4

Module -5

- 19 a) Write about the construction of inverted table. 9
 b) What is tokenization, and how does it differ from case-folding. 5
 20 a) Compare unigram and bigram language model 5
 b) Explain language model for IR 9
