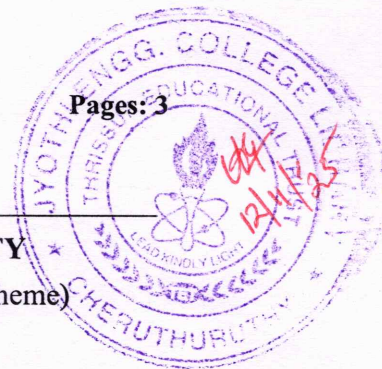


Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
 B.Tech Degree S5 (R, S) Examination November 2025 (2019 Scheme)



Course Code: ADT301
Course Name: FOUNDATIONS OF DATA SCIENCE

Max. Marks: 100

Duration: 3 Hours

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

Marks

- | | | |
|----|--|---|
| 1 | Summarize the essential traits of a data scientist. | 3 |
| 2 | What does the term "data science" encompass, and what is the significance of its role? | 3 |
| 3 | Define binning and list different binning methods. | 3 |
| 4 | Illustrate the major tasks in data preprocessing. | 3 |
| 5 | Differentiate between supervised and unsupervised learning. Give examples for each. | |
| 6 | Discuss the concept of Case-based Reasoning. | 3 |
| 7 | Define the terms frequent item set and association rule mining. | 3 |
| 8 | Define the following terms:
a. Core point
b. Border point
c. Noise | 3 |
| 9 | How can one use a ROC curve to assess whether a classifier is effective? | 3 |
| 10 | Illustrate the concept of k-fold cross-validation. | 3 |

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

- | | | |
|----|--|---|
| 11 | a) Explain about the privacy implications associated with data science. | 6 |
| | b) Outline any two fields of data science where data science is widely used. | 8 |
| 12 | a) Enumerate the different tools and skills needed for a data scientist. | 8 |
| | b) Differentiate AL, ML and DL. | 6 |

Module -2

- 13 a) Explain the methods used to handle missing values during the data preprocessing step. 6
- b) Given the following data for the attribute age: 8, 16, 9, 15, 21, 21, 24, 30, 26, 27, 30, 34. Apply binning techniques to achieve data smoothing, utilizing a bin depth of 3. Demonstrate the procedural steps involved in this process. 8
- 14 a) Apply the following methods to normalize the data: 1000, 2000, 3000, 5000, 9000. 8
- min-max normalization
 - z-score normalization
- b) Discuss in detail different data reduction strategies. 6

Module -3

- 15 a) Explain ID3 algorithm for decision tree induction with an example. 7
- b) Outline Naïve Bayes classification using the below data set and classify the given data: X= (Color = Yellow, Type= SUV, Origin= Domestic, Stolen=?) 7

Example No	Colour	Type	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

- 16 a) Explain the concept of backpropagation algorithm with an example. 7
- b) Illustrate the major steps of KNN with an example. 7

Module -4

- 17 a) Differentiate between AGNES and DIANA with an example. 4
- b) Consider the following dataset and find frequent itemsets and generate all strong association rules for them. Let minimum support count be 2 and minimum confidence be 70%. 10

TID	Items
T1	I1,I2,I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

- 18 a) Describe the concept of DBSCAN algorithm with an example. 7
- b) Identify the uses of Apriori algorithm. Give the steps used in the Apriori algorithm to find the most frequent itemsets. Suppose the data containing frequent itemset X is {11, 12, 15}. Apply the association rules that can be generated from X, if the nonempty subsets of X are {11, 12}, {11, 15}, {12, 15}, {11}, {12}, and {15} and minimum confidence threshold is 70%? Output the strong association rules. 7

Module -5

- 19 a) Explain bootstrap model with the help of an example. 6
- b) Assume the following: A database contains 80 records on a particular topic, of which 55 are relevant to a certain investigation. A search was conducted on that topic, and 50 records were retrieved. Of the 50 records retrieved, 40 were relevant. Construct the confusion matrix for the search and calculate the precision and recall scores for the search. 8
- 20 a) Provide a concise description of the methods used to estimate a classifier's performance. 6
- b) Explain Random Forest algorithm with an example. 8
