Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Fifth Semester B.Tech Degree Regular and Supplementary Examination December 2023 (2019 Scheme)

Course Code: ADT 301

Course Name: FOUNDATIONS OF DATA SCIENCE

Max. Marks: 100 Duration: 3 Hours

PART A

(Answer all questions; each question carries 3 marks) Marks

1		What is data science? Identify different domains where data science plays an active role.	3
2		List and explain various tools and skills required for data scientist.	3
3		Explain different data reduction techniques.	3
4		Describe different methods for filling missing value.	3
5		Differentiate between SVM and neural network.	3
6		Summarize case-based reasoning and challenges.	3
7		Explain distance measures in algorithmic methods.	3
8		What are the conditions to be satisfied for an association rule to be strong? Illustrate with an example.	3
9		Explain k-fold cross validation.	3
10		Summarize bootstrap sampling.	3

PART B

(Answer one full question from each module, each question carries 14 marks)

Module -1

11	a)	Differentiate data science and data analytics.	6
	b)	Demonstrate different stages of data science process.	8
12	a)	Differentiate AI, ML, and DL.	7
	b)	Discuss some ethical practices of data science.	7

Module -2

13	a)	Apply the following methods to normalize the data: 200, 300, 400, 600, 1000. a. min-max normalization. b. z-score normalization.	6

b) Explain major tasks in data pre processing. 8

14 a) Define binning. 8

Given the following data for the attribute age: 13, 15, 16, 16, 19,20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Use binning methods to smooth these data. Illustrate the steps.

b) Explain different data visualization techniques. 6

Module -3

15 a) Explain the algorithm for decision tree induction with an example. 8

b) Describe the classification processes using support vector machine with an example. 6

16 a) Explain rule-based classification and different rule extraction techniques with algorithms and examples. 8

b) Outline the Naive Bayesian classification using the below dataset classify given data: X = (Colo r = Red, Type = SUV, Origin = Domestic, Stolen = ?). 6

Example No.	Color	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

Module -4

17 a) Identify the uses of Apriori algorithm. Give the steps used in the Apriori algorithm to find the most frequent item sets. 4

b) Consider the following dataset, find all frequent item sets and generate all strong association rules for them. Let minimum support count be 2 and minimum confidence be 60%. 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) How does the k-means algorithm work? Clearly state the k-means partitioning algorithm with the help of an example. 7

b) Explain DBSCAN algorithm. 7

Module -5

19 a) 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, recall, specificity and accuracy scores for the search. 8

b) Explain the concepts of Random Forest. 6

20 a) Explain the concepts of bagging with help of an algorithm. 8

b) Explain different performance evaluation parameters. 6

***