A 1100ADT301112404 Pages: 2

Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S5 (S,FE) Examination May 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		List and explain various tools and skills required for data scientists.	3
2		What is Data Science? Why Data Science is required?	3
3		Is regression a supervised learning technique? Justify your answer.	3
4		Define categorical attributes. What are the different types?	3
5		Discuss about support vectors.	3
6		Differentiate between classification and regression.	3
7		What is meant by support and confidence in Apriori algorithm?	3
8		Illustrate clustering? List out clustering methods.	3
9		Explain the concept of bagging and boosting.	3
10		Explain k-fold cross-validation.	3

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

Module -1

11	a)	Explain in detail the various steps in the Data Science process	7
	b)	Identify the different domains where data science plays an active role.	7
12	a)	What is data? Describe the various types of data, providing examples for each.	7
	b)	Discuss about the emerging trends in data science.	7

Module -2

13	a)	Explain the pre-processing techniques available in data mining.	7
	b)	What is data visualization, and what are the different techniques used for visualizing data?	7
14	a)	Discuss various data compression techniques in detail	6
	b)	Describe in detail about different data cleaning methods.	8

Module -3

15 a) Illustrate the concept of the Bayesian belief network. Explain in detail. 7

b) Explain the KNN classification algorithm. 7

16 a) Describe the data classification process with a neat diagram. How does the Naïve Bayesian classification work? Explain 7

b) Define Bayesian Belief Network. What is the role of conditional probability tables (CPTs) in Bayesian Belief Networks? 7

Module -4

17 a) Find the frequent item sets and generate the association rules using the Apriori algorithm if minimum support is 2 and minimum confidence is 60%. 7

TID	List of 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

b) Explain the partition method of clustering with appropriate algorithms. 7

18 a) What is the Apriori algorithm? Give the steps used in the Apriori algorithm to find the most frequent item sets. 7

b) Discuss about density-based clustering algorithm. 7

Module -5

19 a) A database has 300 entries, and 120 are considered relevant to a particular case study. After running a search, 100 records were retrieved, and 75 were relevant. Create the confusion matrix for the search results and compute the precision and recall scores. 7

b) Explain the different methods for improving the model performance. 7

20 a) Describe different methods for evaluating model performance. 7

b) Explain the concept of Random Forest. 7

***