Reg No.:_____   Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Fifth Semester B.Tech Degree Regular and Supplementary Examination December 2022 (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	Explain the different components of data science.	3
2	Identify major research challenges in data science.	3
3	Define binning and explain different smoothing techniques.	3
4	Explain the term covariance with an example.	3
5	Compare eager classification versus lazy classification with an examples.	3
6	Differentiate linearly separable and linearly Inseparable data. How to handle each in SVM?	3
7	Discuss the term frequent item sets and association rule mining.	3
8	Discuss the requirements of clustering in data mining.	3
9	Explain TPR, FPR, and ROC.	3
10	Differentiate Holdout method and Cross-Validation.	3

PART B

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

Module -1

11	a)	Discuss the emerging trends in data science.	6
	b)	Differentiate data science and data analytics.	8
12	a)	Explain the term data and explain different types with an examples.	7
	b)	Describe different steps involved in data science process.	7

Module -2

13	a)	Describe the procedures in data reduction strategy using PCA.	8
	b)	Explain the issues related to data integration.	6
14	a)	Explain data visualization techniques in detail.	8
	b)	Describe different data reduction strategies.	6

Module -3

15 a) Describe the steps of classification processes using the back propagation   8
        algorithm.

    b) Explain the classification processes using Bayesian belief networks with an   6
        example.

16 a) Discuss the algorithm for rule extraction using sequential covering and decision   8
        tree with an example.

    b) Discuss different attribute selection methods used for decision tree.   6

Module -4

17 a) Explain different types of hierarchical clustering methods with an example.   8

    b) Explain PAM, k-medoids partitioning algorithm with an example.   6

18 a) Find the frequent item sets, generate the association rules and list all strong   10
        associations using the Apriori algorithm if minimum support is 2 and minimum
        confidence is 50%.

TID	ITEMSETS
T1	A, B
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

    b) Explain the steps of K means clustering algorithm with an example.   4

Module -5

19 a) Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy   6
        and 1000 are actually sick. For the sick people, a test was positive for 620 and
        negative for 380. For the healthy people, the same test was positive for 180 and
        negative for 8820. Construct confusion matrix for the data and compute the
        precision, sensitivity, specificity, and accuracy for the data.

    b) Explain the concepts of bootstrap aggregation with help of an algorithm.   8

20 a) Explain different methods for improving the model performance.   8

    b) Explain different performance evaluation parameters.   6

*****