

A

1100ADT301112403



Reg No.: \_\_\_\_\_

Name: \_\_\_\_\_

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**

Fifth Semester B.Tech Degree (R, S) Examination November 2024 (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  | List the real-life applications of data science.                           | 3 |
| 2  | Differentiate Artificial Intelligence, Machine Learning and Deep Learning. | 3 |
| 3  | Illustrate simple linear regression.                                       | 3 |
| 4  | What do you mean by data cleaning?   | 3 |
| 5  | Write a short note on Maximal Margin Hyperplanes.                          | 3 |
| 6  | Is case-based reasoning a lazy classification method? Justify your answer. | 3 |
| 7  | Explain Market Basket Analysis.  | 3 |
| 8  | Discuss the requirements of clustering in data mining.                     | 3 |
| 9  | What is meant by bootstrap sampling?                                       | 3 |
| 10 | Explain the concept of Cross-Validation.                                   | 3 |

**PART B**

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

**Module -1**

- |    |   |   |
|----|---|---|
| 11 | a) List and briefly explain various tools and skills required for data science. | 7 |
|    | b) Differentiate data science and data analytics.                               | 7 |
| 12 | a) Identify the different domains where data science plays an active role.      | 7 |
|    | b) Explain in detail the various steps in the Data Science process.             | 7 |

**Module -2**

- |    |   |   |
|----|---|---|
| 13 | a) Define data and explain various types of data with examples for each.                        | 7 |
|    | b) What is Discretization? Explain entropy-based discretization in detail with examples.        | 7 |
| 14 | a) Explain various data reduction methods.  | 7 |
|    | b) What is data visualization, and what are the different techniques used for visualizing data? | 7 |

**Module -3**

- 15 a) What is a multi-layer feed-forward neural network? Explain Backpropagation in detail. 8
- b) Discuss in detail about Rule-based Classification 6
- 16 a) Describe the concept of Support Vector Machine Classification in detail. 7
- b) Explain the decision tree induction algorithm for classifying data tuples with suitable examples. 7

**Module -4**

- 17 a) Differentiate Agglomerative and Divisive Hierarchical Clustering. 6
- b) Find the frequent item sets and generate the association rules using the Apriori algorithm if minimum support is two and minimum confidence is 70%. 8

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

- 18 a) Explain the DBSCAN algorithm. 7
- b) Explain how the k-means algorithm works with the help of an example. 7

**Module -5**

- 19 a) Consider a database that holds 200 documents related to a medical condition. Out of these, 80 documents are relevant to a particular diagnosis. A search retrieved 100 documents, and 65 of them were relevant. Build the confusion matrix and determine the precision and recall for this search. 7
- b) What are TPR and FPR related to the ROC curve? Explain the ROC curve in detail. 7
- 20 a) Discuss different performance evaluation parameters. 6
- b) Explain the Ensemble learning methods with the algorithms. 8

\*\*\*