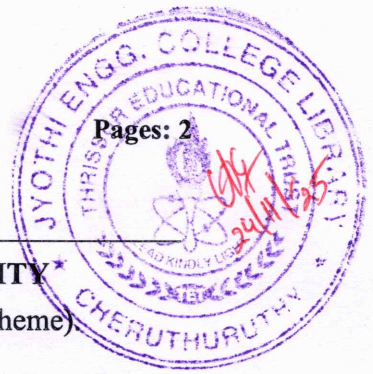Reg No.:_____   Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S3 (R) Examination November 2025 (2024 Scheme)

Course Code: PBADT304

Course Name: INTRODUCTION TO DATA SCIENCE

Max. Marks: 40   Duration: 2 hours 30 minutes

PART A

(Answer all questions. Each question carries 2 marks)   CO   Marks

		CO	Marks
1	Identify the different domains where data science plays an active role?	CO1	(2)
2	Discuss the emerging trends in data science.	CO1	(2)
3	Define null hypothesis and alternate hypothesis with example.	CO2	(2)
4	Different types of sampling used in statistics.	CO2	(2)
5	What is the meaning of the area under the ROC curve (AUC)?	CO3	(2)
6	Briefly explain preprocessing technique in data mining.	CO3	(2)
7	Merits and demerits of decision tree classification.	CO4	(2)
8	Explain briefly about linear regression.	CO4	(2)

PART B

(Answer any one full question from each module, each question carries 6 marks)

Module -1

9	Differentiate Artificial Intelligence, Machine Learning and Deep Learning.	CO1	6
10	List and briefly explain various tools and skills required for data science?	CO1	6

Module -2

11	Perform singular value decomposition on matrix $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$.	CO2	6
12	A university claims that the average number of hours its students spend studying per week is 15 hours. A random sample of 30 students is selected, and the sample mean is found to be 17 hours, with a sample standard deviation of 4 hours. Based on the calculated t-value and critical value, decide whether to reject the null hypothesis.	CO2	6

Module -3

| 13 | A university uses a binary classifier to predict whether students will pass (P) or fail (F) a final exam. On a test set of 200 students the results were: (a) 120 students actually passed, 80 actually failed. (b) The classifier predicted "Pass" for 110 students, and "Fail" for 90 students. (c) Among those predicted "Pass", 100 students truly passed. (d)Among those predicted "Fail", 70 students truly failed. Construct the confusion matrix (with actual classes as rows, predicted classes as columns). Compute accuracy, precision (for predicting "Pass"), and recall (for "Pass"). | CO3 | 6 |

| 14 | 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 with a bin depth of 3. Illustrate your steps. | CO3 | 6 |

Module -4

| 15 | | CO4 | 6 |

Outlook	Play Tennis
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	No

Compute entropy and information gain of the above dataset

| 16 | Explain wrappers and filter | CO4 | 6 |

***