

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2024

B.Tech-VI Semester (ECE)

COURSE CODE(CREDITS): 19B1WEC636

MAX. MARKS: 15

COURSE NAME: Machine Learning for Data Analysis

COURSE INSTRUCTORS: Dr. Alok Kumar

MAX. TIME: 1 Hour

**Note:** (a) All questions are compulsory.

(b) Marks are indicated against each question in square brackets.

(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q.1 Consider two datasets X and Y as follows:

X: [12, 15, 18, 21, 24], and Y: [20, 21, 22, 23, 24]

Calculate the coefficient of variation for each dataset and discuss the differences in variability represented by these values. [CO1] [2]

Q.2 Discuss the importance of handling missing values during the data preprocessing stage. Describe various strategies for dealing with missing data, including deletion, imputation, and prediction-based methods. Provide examples illustrating when each strategy would be appropriate and discuss the potential implications of each approach on the analysis results.

[CO1] [3]

Q.3 Explain the concept of a decision tree in machine learning, detailing its construction, splitting criteria. For the given dataset, construct a decision tree using the Gini impurity criterion. [CO2, CO3] [4]

Feature 1	Feature 2	Class
1	0	Positive
0	1	Negative
1	1	Positive
0	0	Negative
1	0	Negative
1	1	Positive
0	1	Positive
0	1	Positive
1	0	Negative
0	0	Positive

Q.4 What is the Gini index, and how does it differ from entropy as a measure of impurity? In what scenarios might one measure be preferred over the other? [CO2] [3]

Q.5 Differentiate between Box plot and Scatter plot. What difference does the availability of class labels make in scatter plots? Draw the Box plot for the following sets of numbers representing percentages achieved on a test by a group of 10 students: 64, 63, 42, 70, 64, 73, 81, 77, 81, and 76. [CO1] [3]

UNIT TEST-1 EXAMINATION- Feb-2024