SEAT NO. AI-22001

NED UNIVERSITY OF ENGINEERING & TECHNOLOGY

THIRD YEAR (Artificial Intelligence)

FALL SEMESTER EXAMINATIONS 2024

Time : 3 Hours | **Batch 2022** | **Dated : 06-DEC-24** | **Max Marks : 60**

Data Mining - CT-377

Note: Attempt all questions. Answer all parts of a question in a row. State your assumptions clearly. Draw diagram where necessary.

Q1. [CLO 1: 24 Marks]

a) What is overfitting in supervised learning? ***Show*** the methods to resolve overfitting condition in Decision tree classifier. **[12]**

b) ***Relate*** the zero probability problem in Naïve Bayes classifier and how to overcome this problem. Show the mathematical equation and give example. **[12]**

Q2. [CLO 2: 12 Marks]

Based on the confusion matrix of a Machine learning model on a sentiment analysis problem. ***Apply*** using the following metrics to evaluate the model's performance. **[12]**

i. Accuracy
ii. Support
iii. Weighted Precision
iv. Weighted Recall
v. F-Measure

		Predicted				
		Very negative	Negative	Neutral	Positive	Very Positive
Actual	Very negative	45	3	2	1	4
	Negative	5	40	3	6	1
	Neutral	2	4	35	3	6
	Positive	1	5	4	38	2
	Very Positive	3	2	5	4	41

Q3. [CLO 3: 24 Marks]

a) ***Suppose*** the 6 points below and cluster them using hierarchical agglomerative clustering, Euclidean distance for the distance measure between objects and **complete link** for the distance between clusters. Draw dendrogram to illustrate the clusters. Show all iterations. **[12]**

A(1, 1), B(1, 2), C(5, 4), D(7, 5), E(7, 7)

b) The given data presents a training set of Class labelled tuples randomly selected from all the elections customers' database. The **buy computer** is the class variable in the given dataset. ***Solve*** using **Gini Index** to find the attribute that will be chosen to split the root of the tree. Show all your calculations. **[12]**

S. #	Age	Income	Student	Credit Rating	Buys Computer
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle-Aged	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle-Aged	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle-Aged	Medium	No	Excellent	Yes
13	Middle-Aged	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No