INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
UNA [HP]
An Institute of National Importance under MoE
Saloh, Una (HP) – 177 209
Website: www.iiitu.ac.in

AY 2022-23
School of Computing
CURRICULUM: IIITUGCSE20
Cycle Test – II
15, Nov.'22
(2:00 PM – 3:00 PM)

Degree	B. Tech.	**Branch**	CSE/IT/ECE
Semester	V		
Subject Code & Name	CSSE11: Machine Learning		
Time: 60 Minutes	**Answer All Questions**		**Maximum: 20 Marks**

(K1: Remembering, K2: Understanding, K3: Applying, K4: Analyzing, K5: Evaluating)

CO3	Experiment with decision trees and learning rules and analyze the performance of algorithms.
CO4	Identify the real-world problems and apply classification and clustering models on the datasets.

Sl. No.	Question	Marks
1.a	How to deal with an overfitted decision tree in machine learning?	(1)
1.b	Explain any four issues in decision tree learning process.	(2)
1.c	How to select which attribute is the best classifier to test at each node in the tree in C4.5 algorithm? Write the formula to compute that metric.	(2)
2.a	Explain some cases where k-Means clustering fails to give good results.	(1)
2.b	What are outliers? How do they affect clustering? Explain with a suitable example.	(2)
2.c	For given these three points: {P1(1, 2), P2(2, 3) and P3(3, 4)}, find the pair for which the Manhattan distance is maximum.	(2)
3.a	A preview of COVID-19 infection dataset S is shown below:	(1)

ID	Fever	Cough	Breathing Issues	Infected
1.	No	No	No	No
2.	Yes	Yes	Yes	Yes
3.	Yes	Yes	No	No
4.	Yes	No	Yes	Yes
5.	Yes	Yes	Yes	Yes
6.	No	Yes	No	No
7.	Yes	No	Yes	Yes
8.	Yes	No	Yes	Yes
9.	No	Yes	Yes	Yes
10.	Yes	Yes	No	Yes
11.	No	Yes	No	No
12.	No	Yes	Yes	Yes
13.	No	Yes	Yes	No
14.	Yes	Yes	No	No

The columns used to make decision nodes, viz., 'Breathing Issues', 'Cough' and 'Fever' are called feature columns and the column used for leaf nodes, i.e., 'Infected' is called the target column.

Calculate the Entropy of Dataset S.

(log values can be kept intact in calculations.)

3.b	Considering COVID-19 infection dataset S in question 3., Calculate the Information Gain for Fever. (log values can be kept intact in calculations.)	(2)
3.c	How does ID3 algorithm select the best feature at each step while building a Decision tree?	(2)
4.a	Which clustering technique requires a merging approach?	(1)
4.b	List any four differences between partitioning and hierarchical clustering.	(2)
4.c	Illustrate PAM (Partitioning Around Medoids) approach of K-medoids Clustering with a suitable example.	(2)