CONTINUOUS INTERNAL EVALUATION- 3

Dept: CSE	Sem / Div:6A&B	Sub: Data Mining and Data Ware housing	S Code:18CS641
Date:05/08/2021	Time:3.00-4.30 PM	Max Marks: 50	Elective: Yes
Note: Answer any 2 full questions, choosing one full question from each part.			

-

QN		Questions	Marks	RBT	COs
		PART A			
1	a	Explain how to build decision tree using Hunt's Algorithm.	10	L2	CO4
	b	Define classification. With a neat figure explain the general approach for solving classification problem.	10	L2	CO4
	c	Explain K-Nearest Neighbor classification algorithm.	5	L2	CO4
		OR			
2	a	Consider a training data set that contains 100 positive examples and 400 negative examples. For each of the following candidate rules: R1:A→(Covers 4 positive and 1 negative examples) R2:B→(Covers 30 positive and 10 negative examples) R3:C→(Covers 100 positive and 90 negative examples) Determine which is the best and worst candidate rule according to i)FOIL's information gain ii)The likelihood ratio statistics iii)Laplace measure.	10	L3	CO4
	b	What is rule based classifier? Explain sequential covering algorithm in rule based classifier.	10	L2	CO4
	c	Write a note on Bayesian classifier.	5	L2	CO4
		PART B			
3	a	Explain K means algorithm and what are its limitation.	10	L2	CO4
	b	Explain Agglomerative Hierarchical Clustering with different proximity between clusters.	10	L2	CO4
	c	What is cluster analysis? Explain the different types of clustering techniques with example.	5	L2	CO4
		OR			
4	a	Discuss DBSCAN algorithm and estimate the time and space complexity.	10	L2	CO4
	b	What is graph based clustering? Explain in detail the steps used in graph based clustering.	10	L2	CO4
	c	Explain different types of clusters with neat diagram.	5	L2	CO4

Thapas.

Prepared by: Prof. Thapaswini P S

HOD-CSE