2018 Scheme

1	N	T							

NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

An Autonomous Institution Approved by UGC/AICTE/Govt. of Karnataka
Accredited by NBA (Tier – I) and NAAC 'A+' Grade
Affiliated to Visveswaraya Technological University, Belagavi
Post Box No. 6429, Yelahanka, Bengaluru – 560 064, Karnataka, India

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MID SEMESTER EXAMINATION-II

Course Title with code	Data Mining, 18CS54	**Maximum Marks**	**30 Marks**
Date and Time	30/12/2021, 9.30am to 10.30am	**No. of Hours**	1.0
Course Instructor(s)	Dr. Vijaya Shetty S, Dr. Sujata Joshi, Dr. Vani V		

Instructions to Students

1. Answer any **two full questions.**
2. Any missing data may assume suitably.

Q. No	Question	MAX MARKS	CO	BL	PO/ PSO
1. a	Illustrate the working of a decision tree with a suitable example.	6	3	2	1/2
1. b	Consider the data set shown in Table 1b to find **coverage** and **accuracy** of the rule: **(Give Birth = yes) ∧ (Blood Type = warm) → Mammals** Table 1b Dataset of vertebrates Name | Blood Type | Give Birth | Can Fly | Live in Water | Class human | warm | yes | no | no | mammals python | cold | no | no | no | reptiles salmon | cold | no | no | yes | fishes whale | warm | yes | no | yes | mammals frog | cold | no | no | sometimes | amphibians komodo | cold | no | no | no | reptiles bat | warm | yes | yes | no | mammals pigeon | warm | no | yes | no | birds cat | warm | yes | no | no | mammals leopard shark | cold | yes | no | yes | fishes turtle | cold | no | no | sometimes | reptiles penguin | warm | no | no | sometimes | birds porcupine | warm | yes | no | no | mammals eel | cold | no | no | yes | fishes salamander | cold | no | no | sometimes | amphibians gila monster | cold | no | no | no | reptiles platypus | warm | no | no | no | mammals owl | warm | no | yes | no | birds dolphin | warm | yes | no | yes | mammals eagle | warm | no | yes | no | birds	4	3	3	1/2
1. c	Apply KNN algorithm to the data set of Person given in Table 1c and find if the person in the query example is Underweight or Normal after standardizing the input attributes between 0 and 1 range. Given: K=3, Distance Measure: Euclidean **170** | **57** | **?** Table 1c: Dataset of Person Height(CM) | Weight(KG) | Class 167 | 51 | Underweight 182 | 62 | Normal 176 | 69 | Normal 173 | 64 | Normal 172 | 65 | Normal 174 | 56 | Underweight 169 | 58 | Normal	5	3	3	3/2
2. a	Consider the 5 transactions and 6 items shown in Table 2a. With the help of **Apriori algorithm find the association rules** with **50% support and 75% confidence** Table 2a: Transactions TID | Items Bought 1 | A,B,C,D 2 | A,B,D 3 | A,E,F 4 | A,D,E	8	4	3	1,2, 3/2

	5 B,D,E				
2. b	Write the pseudocode for the frequent itemset generation part of the Apriori algorithm.	4	4	2	1,2, 3/2
2. c	Table 2c Market basket Dataset Transaction ID / Items Bought 1 {Milk, Beer, Diapers} 2 {Bread, Butter, Milk} 3 {Milk, Diapers, Cookies} 4 {Bread, Butter, Cookies} 5 {Beer, Cookies, Diapers} 6 {Milk, Diapers, Bread, Butter} 7 {Bread, Butter, Diapers} 8 {Beer, Diapers} 9 {Milk, Diapers, Bread, Butter} 10 {Beer, Cookies} Consider the market basket transactions shown in Table 2c to answer (a) and (b) (a) What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)? (b) What is the maximum size of frequent itemset that can be extracted (assuming minsup > 0)?	3	4	3	1,2, 3/2
3. a	Consider the dataset given in Table 3a Table 3a Weather Data Day / Outlook / Temperature / Humidity / Wind / Play Tennis D1 Sunny Hot High Weak No D2 Sunny Hot High Strong No D3 Overcast Hot High Weak Yes D4 Rain Mild High Weak Yes D5 Rain Cool Normal Weak Yes D6 Rain Cool Normal Strong No D7 Overcast Cool Normal Strong Yes D8 Sunny Mild High Weak No D9 Sunny Cool Normal Weak Yes D10 Rain Mild Normal Weak Yes D11 Sunny Mild Normal Strong Yes D12 Overcast Mild High Strong Yes D13 Overcast Hot Normal Weak Yes D14 Rain Mild High Strong No i. Compute the information gains on the attributes "Temperature" and "Wind" relative to these training examples. ii. What is the best split among "Temperature" and "Wind" according to the information gain iii. Which attribute among these would the decision tree algorithm choose as the first split?	6	3	3	3/2
3. b	Consider a training set that contains 100 positive examples and 400 negative examples. For each of the following candidate rules, R_1: $A \longrightarrow +$ (covers 4 positive and 1 negative examples), R_2: $B \longrightarrow +$ (covers 30 positive and 10 negative examples), determine which is the best and worst candidate rule according to: (a) FOIL's information gain. (b) The Laplace measure.	5	3	3	3/2
3. c	Illustrate with an example, F_{k-1} x F_1 Method of candidate generation and pruning.	4	4	2	1/1

Faculty Signature	Course Co-Ordinator/Mentor Signature	HoD Signature