2018 Scheme

| 1 | N | T | | | | | | |

Department of Computer Science and Engineering

Course Title with code	Data Mining, 18CS54	Maximum Marks	30 Marks
Date and Time	19/11/2021, 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 and PSO
1.a	Suppose that you are employed as a data mining consultant for an Over The Top (OTT) media service company. Describe how data mining can help the company by giving specific examples of how techniques, such as clustering, classification and association rule mining can be applied.	6	1	3	1,2/1
1.b	Discuss whether each of the following activities is a data mining task. (1) Order the Universities based on the student enrollment rate. (2) Grouping of similar traffic frames from a video. (3) Determine the maximum salary of academic staff in every department at NMIT. (4) Detect the key frames from a rugby sport video footage.	6	1	1	1,2/1
1.c	Discuss at least 3 challenges that motivated the development of data mining.	3	1	1	1,2/1
2. a	Distinguish between Qualitative and Quantitative attribute types. Classify the following attributes as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). **Hint:** Age in years. Answer: Discrete, quantitative, ratio i) Military rank. ii) Number of patients in a hospital. iii) Temperature in Fahrenheit iv) Eye color	5	2	2	1,2/1

2. b	Consider the following data (in increasing order) for the attribute age: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70, answer the following: (a) Use min-max normalization to transform the value 45 for age onto the range [0.0,1.0]. (b) Use z-score normalization to transform the value 45 for age, where the standard deviation of age is 12.94 years. (c) Use normalization by decimal scaling to transform the value 45 for age.	5	2	3	1,2,3/1
2. c	Outline the key principle for effective sampling and the different types of sampling approaches used by data miners.	5	2	2	1/1
3. a	In real-world data, tuples with *missing values* for some attributes are a common occurrence. Describe various methods for handling this problem.	5	2	2	1,2/1

3. b	Compute the following for the vectors $X=(1,1,1,0,1,0)$ and $Y=(1,0,1,0,1,0)$
	    i. Jaccard similarity
	   ii. Cosine similarity
	  iii. Euclidean distance
		5	2	3	1,2,3/1

3. c	Suppose that your local bank has a data mining system. The bank has been studying your debit card usage patterns. Noticing that you make many transactions at home renovation stores, the bank decides to contact you, offering information regarding their special loans for home improvements. (a) Discuss how this may conflict with your right to privacy. (b) Describe another situation in which you feel that data mining can infringe on your privacy. (c) Describe a privacy-preserving data mining method that may allow the bank to perform customer pattern analysis without infringing on its customers' right to privacy.	5	3	1	1,2,3/1

Faculty Signature	Course Coordinator/Mentor Signature	HoD Signature Dr. Thippeswamy M N