Reg. No.
----------



## MANIPAL INSTITUTE OF TECHNOLOGY (A constituent unit of MAHE, Manipal 576104)

## I SEM M. Tech (MI) DEGREE END-SEMESTER EXAMINATIONS, NOV/DEC 2023 SUBJECT: FEATURE ENGINEERING AND DATA MINING (BME 5118) (REVISED CREDIT SYSTEM) Tuesday, 5<sup>th</sup> December 2023, 9.30 AM To 12.30 PM

## **TIME: 3 HOURS**

MAX. MARKS: 50

4

3

Instructions to Candidates:							
1. Answer ALL questions.							
2. Draw labeled diagram wherever necessary							

- 1A Considering the application of diagnosis of diseases in healthcare for predicting and 3 description, Explain the Data Mining Process with relevant block.
  1B The numerical data are available in 4 types of attributes, considering relevant application 3
- for each explain the attributes with statistical operations & examples.
- **1C** Interpret the importance of the Architecture of Data mining with the block diagram.
- 2A Analyse the different steps involved in data pre-processing for the example of weather 4 forecasting.
- 2B Justify the use of multidimensional views of data and data-cubes. With a neat diagram, 3explain data-cube implementations
- **2C** Two binary vectors are given below: (04 Marks) X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0) Y = (0, 0, 3, 0, 0, 0, 0, 1, 0, 0, 1) Calculate (i) SMC (ii) Jaccord similarly coefficient and hamming distance.
- **3A** Interpret the necessity of feature engineering and selection process.
- **3B** For the given Confusion-matrix below for 3 classes. Find sensitivity & specificity metrics **4** to estimate predictive accuracy of classification methods.

Predicted class	True class				
	1	2	3		
1	8	1	1		
2	2	9	2		
3	0	0	7		

3C A diagnosis has been predicted and need to be marked for classes, Identify the procedure 3 and draw a neat figure and explain general approach for solving classification-model.

	Reg. No.										
Interpret the different types of dor	main spaces	in im	age s	segm	entat	ion.					3
Interpret the algorithm for k-neared	est neighbour	r clas	sific	ation							4
Briefly illustrate the Data warehow	use architect	ure w	vith C	DLAI	P tecl	hnolo	ogy.				3
Develop the Apriori algorithm for	generating	frequ	ent-i	tem s	set.						3
List and explain four distance mea and find out the distance betwee	asures to con en two obj	npute ects	e the repre	dista: esente	nce b ed b	oetwe y att	een a ribut	pair te val	of p lues	oints	4

- (1,6,2,5,3) & (3,5,2,6,6) by using any 2 of the distance measures
- **5**C Consider the following transaction data-set 'D' shows 9 transactions and list of items 3 using Apriori algorithm frequent-item set minimum support = 2

Tid	T <sub>1</sub>	T <sub>2</sub>	T	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T7	T <sub>8</sub>	To
List of items	I1, I2, I5	I2, L4	$I_2, I_3$	$I_1, I_2, I_4$	$I_1, I_3$	$I_2, I_3$	I1, I3	11, 12, 13, 18	$I_1, I_2, I_3$

**4**A

**4B** 

**4**C

5A

**5B**