

**P.E.S. College of Engineering, Mandya - 571 401***(An Autonomous Institution affiliated to VTU, Belagavi)***Fifth Semester, B.E. - Information Science and Engineering****Semester End Examination; February / March - 2022****Data Mining**

Time: 3 hrs

Max. Marks: 100

**Course Outcomes***The Students will be able to:**CO1: Apply different pre-processing techniques for data cleaning.**CO2: Evaluate performance of algorithms for Association Rules.**CO3: Apply the different classification techniques.**CO4: Analyze different clustering algorithms.**CO5: Understand different data models used in data warehouse.***Note: I) PART - A is compulsory. Two marks for each question.****II) PART - B: Answer any Two sub questions (from a, b, c) for Maximum of 18 marks from each unit.**

Q. No.	Questions	Marks	BLs	COs	POs												
<b>I : PART - A</b>		<b>10</b>															
I a.	Discuss, why data mining has become important?	2	L1	CO1	PO1												
b.	Define support and confidence.	2	L1	CO2	PO1												
c.	Define information gain and entropy.	2	L1	CO3	PO1												
d.	Differentiate monothetic and polythetic divisive methods.	2	L2	CO4	PO1												
e.	Define ODS and ETL.	2	L2	CO5	PO1												
<b>II : PART - B</b>		<b>90</b>															
<b>UNIT - I</b>		<b>18</b>															
1 a.	Define data mining. Explain the Cross Industry Standard Process (CRISP-DM) for data mining with a neat diagram.	9	L2	CO1	PO1												
b.	Explain the different methods for handling missing data.	9	L2	CO1	PO1												
c.	Why do we need data pre-processing? Explain the different forms of data pre-processing.	9	L2	CO1	PO1												
<b>UNIT - II</b>		<b>18</b>															
2 a.	Derive the frequent item set for the below transactional data, given the minimum support as 50% and minimum confidence as 60%.																
	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>TID</th> <th>Items</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Bread, milk</td> </tr> <tr> <td>2</td> <td>Bread, diaper, biscuit, eggs</td> </tr> <tr> <td>3</td> <td>Milk, diaper, biscuit, coke</td> </tr> <tr> <td>4</td> <td>Bread, milk, diaper, biscuit</td> </tr> <tr> <td>5</td> <td>Bread , milk, diaper, coke</td> </tr> </tbody> </table>	TID	Items	1	Bread, milk	2	Bread, diaper, biscuit, eggs	3	Milk, diaper, biscuit, coke	4	Bread, milk, diaper, biscuit	5	Bread , milk, diaper, coke	9	L4	CO2	PO1
TID	Items																
1	Bread, milk																
2	Bread, diaper, biscuit, eggs																
3	Milk, diaper, biscuit, coke																
4	Bread, milk, diaper, biscuit																
5	Bread , milk, diaper, coke																
b.	What is dynamic itemset counting? Explain the steps to generate FP-trees.	9	L2	CO2	PO1												
c.	Discuss the issues that need to be considered in Apriori algorithm.	9	L3	CO2	PO1												

**UNIT - III****18**

- 3 a. Define decision tree. Construct the decision tree to classify bank loan applications with Risk as class label.

Owens home?	Married	Gender	Employed	Credit rating	Risk
Yes	Yes	Male	Yes	A	B
No	No	Female	Yes	A	A
Yes	Yes	Female	Yes	B	C
Yes	No	Male	No	B	B
No	Yes	Female	Yes	B	C
No	No	Female	Yes	B	A
No	No	Male	No	B	B
Yes	No	Female	Yes	A	A
No	Yes	Female	Yes	A	C
yes	yes	Female	yes	A	C

9 L4 CO3 PO1

- b. Define Baye's theorem. Discuss the methods for estimating the accuracy of classification methods.
- c. Explain over fitting and pruning and discuss the evaluation criteria for classification.

9 L2 CO3 PO1

9 L2 CO3 PO1

**UNIT - IV****18**

- 4 a. What is cluster analysis? Explain desired features of cluster analysis.
- b. Discuss the methods for computing distances between clusters.
- c. Explain types of cluster analysis methods.

9 L2 CO4 PO1

9 L2 CO4 PO1

9 L2 CO4 PO1

**UNIT - V****18**

- 5 a. Discuss the data warehouse implementation steps in detail.
- b. Discuss the data cube operations with an example and define ROLAP.
- c. List the OLAP software and discuss the guidelines for OLAP implementation.

9 L2 CO5 PO1

9 L2 CO5 PO1

9 L2 CO5 PO1

\* \* \* \*