

**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; Feb. - 2021****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																									
I : PART - A		10																												
I a.	Define Data Mining.	2	L1	CO1	PO1																									
b.	List two advantages of Apriori.	2	L1	CO2	PO1																									
c.	Define split algorithm.	2	L1	CO3	PO1																									
d.	List clustering analysis methods.	2	L1	CO4	PO1																									
e.	List any two characteristics of OLAP codd's.	2	L1	CO5	PO1																									
II : PART - B		90																												
UNIT - I		18																												
1 a.	Explain the various data mining process in detail.	9	L2	CO1	PO1																									
b.	Construct a model for data mining process referring to Chapman description.	9	L4	CO1	PO1																									
c.	Identify various sources of errors in data.	9	L3	CO1	PO2																									
UNIT - II		18																												
2 a.	Discuss the first part frequent itemsets parameters of Apriori algorithm.	9	L2	CO2	PO1																									
b.	Summarize the required reasons to improve the efficiency of Apriori algorithm.	9	L2	CO2	PO2																									
c.	Explain dynamic itemset counting with necessary example.	9	L2	CO2	PO1																									
UNIT - III		18																												
3 a.	Evaluate the given frequency table by applying Naive Baye's algorithm. Predict that if a fruit has the following properties and examine which type of fruit it is. Given Fruit = { Yellow, Sweet, Long }																													
	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Fruit</th> <th>Yellow</th> <th>Sweet</th> <th>Long</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Banana</td> <td>350</td> <td>450</td> <td>0</td> <td>650</td> </tr> <tr> <td>Mango</td> <td>400</td> <td>300</td> <td>350</td> <td>400</td> </tr> <tr> <td>Others</td> <td>50</td> <td>100</td> <td>50</td> <td>150</td> </tr> <tr> <td>Total</td> <td>800</td> <td>850</td> <td>400</td> <td>1200</td> </tr> </tbody> </table>	Fruit	Yellow	Sweet	Long	Total	Banana	350	450	0	650	Mango	400	300	350	400	Others	50	100	50	150	Total	800	850	400	1200	9	L4	CO3	PO2
Fruit	Yellow	Sweet	Long	Total																										
Banana	350	450	0	650																										
Mango	400	300	350	400																										
Others	50	100	50	150																										
Total	800	850	400	1200																										

- b. Outline the various steps of tree induction algorithm. 9 L2 CO3 PO1
- c. List and brief the evaluation criteria for classification methods. 9 L4 CO3 PO2

UNIT - IV

18

- 4 a. Categorize the taxonomy of cluster analysis method. 9 L4 CO4 PO3
- b. Solve the given sample data by dividing into two clusters using *k*-means algorithm.

Height (<i>H</i>)	Weight (<i>W</i>)
185	72
170	56
168	60
179	68
182	72
188	77
180	71
180	70
183	84
180	88
180	67
177	76

9 L3 CO4 PO2

- c. Explain the *k*-means method in detail. 9 L2 CO4 PO1

UNIT - V

18

- 5 a. Construct a ODS design system and justify how data warehouse process carry out. 9 L6 CO5 PO3
- b. Explain the characteristics of OLAP systems. 9 L2 CO5 PO2
- c. Classify the different multidimensional view and data cube for university structure. 9 L2 CO5 PO1

* * *