VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SLOT: F2

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

Programme Name & Branch	**: B.Tech. CSE & CSE with Specialization in Data Science**
Course Code and Course Name	**: BCSE208L, Data Mining**
Faculty Name(s)	**: Dr. Geetha Mary A, Dr. Chandramohan B, Dr. Gunavathi C**
Class Number(s)	**: VL2025260502060, VL2025260502553, VL2025260502551**
Date of Examination	**: 22.3.26**
Exam Duration	**: 90 minutes**

Maximum Marks: 50

General instruction(s): Answer All Questions.

Course Outcomes: **CO3** Discover interesting patterns from large amounts of data using Association Rule Mining. **CO4** Extract useful information from the labeled data using various classifiers and Compile unlabeled data into clusters applying various clustering algorithms.

<table>
<tr><th>Q. No.</th><th>Question</th><th>Module</th><th>Marks</th><th>CO</th><th>BL</th></tr>
<tr><td>1.</td><td>Find all frequent itemsets and strong association rules using Apriori algorithm. Use the minimum support value as 3 and the minimum confidence value as 65%.
<table>
<tr><th>Transaction ID</th><th>Items</th></tr>
<tr><td>T1</td><td>1, 2, 3, 4, 5, 6</td></tr>
<tr><td>T2</td><td>7, 2, 3, 4, 5, 6</td></tr>
<tr><td>T3</td><td>1, 8, 4, 5</td></tr>
<tr><td>T4</td><td>1, 9, 0, 4, 6</td></tr>
<tr><td>T5</td><td>0, 2, 2, 4, 5</td></tr>
</table></td><td>4</td><td>10</td><td>3</td><td>3</td></tr>
<tr><td>2.</td><td>Use the tree-based algorithm to find all frequent patterns for the given dataset with the minimum support value as 2.
<table>
<tr><th>Transaction ID</th><th>Items</th></tr>
<tr><td>001</td><td>A, B, C, D, E</td></tr>
<tr><td>002</td><td>B, C, E</td></tr>
<tr><td>003</td><td>C, E. D</td></tr>
<tr><td>004</td><td>A, C, D</td></tr>
</table></td><td>4</td><td>10</td><td>3</td><td>3</td></tr>
<tr><td>3.</td><td>A dataset contains the information about objects classified into two categories namely Class A and Class B. Each object is described using three categorical attributes. Construct the Decision Tree using information gain as the splitting criterion.
<table>
<tr><th>Size</th><th>Color</th><th>Shape</th><th>Class</th></tr>
<tr><td>Small</td><td>Yellow</td><td>Round</td><td>A</td></tr>
<tr><td>Big</td><td>Yellow</td><td>Round</td><td>A</td></tr>
<tr><td>Big</td><td>Red</td><td>Round</td><td>A</td></tr>
<tr><td>Small</td><td>Black</td><td>Round</td><td>B</td></tr>
<tr><td>Big</td><td>Black</td><td>Cube</td><td>B</td></tr>
<tr><td>Big</td><td>Yellow</td><td>Cube</td><td>B</td></tr>
<tr><td>Big</td><td>Black</td><td>Round</td><td>B</td></tr>
<tr><td>Small</td><td>Yellow</td><td>Cube</td><td>B</td></tr>
</table></td><td>5</td><td>10</td><td>4</td><td>3</td></tr>
</table>

VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SLOT: F2

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

4.	Consider the following dataset which includes demographic information and credit scores of individuals. Apply the Naive Bayes classification algorithm to predict the credit score of a new individual based on the following demographic features: Age>40, Gender=Male, Income=Medium, Education=Graduate, Occupation=Skilled.	5	10	4	3

Age	Gender	Income	Education	Occupation	Credit Score
<40	Male	High	Graduate	Professional	Low
<40	Female	Medium	High School	Managerial	Low
>40	Male	High	Postgraduate	Professional	High
>40	Female	High	Postgraduate	Professional	High
>40	Male	Low	High School	Skilled	Low
<40	Female	Medium	Graduate	Managerial	Medium
<40	Male	Medium	Graduate	Professional	High
>40	Female	High	Postgraduate	Managerial	High
>40	Male	High	High School	Skilled	Low
<40	Female	Medium	Graduate	Professional	Medium
>40	Male	Medium	High School	Skilled	Medium
>40	Female	Low	Graduate	Skilled	High

	A classifier was tested on the below dataset for predicting whether a mushroom is edible or not based on their shape, colour and odour. The actual and predicted labels are shown below.	5	10	4	4

Shape	Colour	Odour	Actual Edible	Predicted Edible
C	B	1	Yes	Yes
D	B	1	Yes	Yes
D	W	2	Yes	Yes
D	B	2	Yes	Yes
C	U	2	No	Yes
C	B	2	No	Yes
C	W	3	No	No
C	D	3	No	No
D	W	3	Yes	No

Give the confusion matrix for the test dataset and calculate the accuracy, precision, recall, FPR and F1-Score. Comment on why/why not accuracy is a reliable performance metric, based on the evidence from the above dataset.