

Reg. No.:

Name:



W6972

University of Kerala

Third Semester FYUGP Degree Examination, November 2025

Discipline Specific Core Course

COMPUTER SCIENCE

UK3DSCCSC209 - Data Mining Concepts and Techniques

Academic Level: 200-299

2024 Admission

Time: 1 Hour 30 Minutes(90 Mins.)

Max. Marks: 42

Part A. 6 Marks.Time:6 Minutes.(Cognitive Level:Remember(RE)/Understand(UN)) Objective Type. 1 Mark Each.Answer all questions

Qn No.	Question	CL	CO
1	The number of parts in an association rule is -----.	RE	2
2	Expand the term KNN	RE	3
3	List any two differences between classification and prediction.	UN	3
4	Define support?	UN	4
5	List any one application of data mining.	UN	1
6	Define data warehouse?	UN	2

Part B.8 Marks.Time:24 Minutes.(Cognitive Level:Understand(UN)/Apply(AP))Short Answer. 2 marks each.Answer all questions

Qn No.	Question	CL	CO
7	Compare Classification and Prediction.	UN	3
8	Give any four areas in which data mining is used?	UN	1
9	Illustrate the support and confidence with an example.	AP	4
10	Apply the concept of data warehousing in government administration.	AP	2

Part C. 28 Marks.Time:60 Minutes (Cognitive Level:Apply(AP)/Analyse(AN)/Evaluate(EV)/Create(CR)) Long Answer.7 marks each.Answer all 4 Questions choosing among options * within each question

Qn No.	Question	CL	CO
11	A) Illustrate data warehouse and its importance OR B) Compare and analyze the working principles of KNN and Naïve Bayes in handling high-dimensional datasets.	AP	2, 3
12	A) A company maintains the following dataset of customers:	AN	1, 1

Qn No.	Question	CL	CO																																
	<table><tr><th>ID</th><th>AGE</th><th>INCOME</th><th>PURCHASE</th></tr><tr><td>1</td><td>25</td><td>30000</td><td>Yes</td></tr><tr><td>2</td><td>28</td><td>35000</td><td>No</td></tr><tr><td>3</td><td>35</td><td>NULL</td><td>Yes</td></tr><tr><td>4</td><td>NULL</td><td>50000</td><td>Yes</td></tr><tr><td>5</td><td>45</td><td>60000</td><td>No</td></tr><tr><td>6</td><td>38</td><td>55000</td><td>Yes</td></tr><tr><td>7</td><td>52</td><td>NULL</td><td>No</td></tr></table> a. Identify and categorize different types of data quality problems present in this data set. b. Propose suitable data cleaning techniques for each issue identified. c. Analyze how the chosen preprocessing techniques can impact the accuracy of future data mining process. OR B) Analyze the steps involved in a typical data mining process.	ID	AGE	INCOME	PURCHASE	1	25	30000	Yes	2	28	35000	No	3	35	NULL	Yes	4	NULL	50000	Yes	5	45	60000	No	6	38	55000	Yes	7	52	NULL	No		
ID	AGE	INCOME	PURCHASE																																
1	25	30000	Yes																																
2	28	35000	No																																
3	35	NULL	Yes																																
4	NULL	50000	Yes																																
5	45	60000	No																																
6	38	55000	Yes																																
7	52	NULL	No																																
13	A) Design a method to compare multiple classification models on the same data. OR B) Explain data preprocessing in detail?	EV	3, 1																																
14	A) Construct a decision tree using the Information Gain method for the following dataset. Show all calculations and the final tree structure. <table><tr><th>Weather</th><th>Temperature</th><th>Play</th></tr><tr><td>Sunny</td><td>Hot</td><td>No</td></tr><tr><td>Sunny</td><td>Cool</td><td>Yes</td></tr><tr><td>Overcast</td><td>Hot</td><td>Yes</td></tr><tr><td>Rainy</td><td>Cool</td><td>Yes</td></tr><tr><td>Rainy</td><td>Hot</td><td>No</td></tr></table>	Weather	Temperature	Play	Sunny	Hot	No	Sunny	Cool	Yes	Overcast	Hot	Yes	Rainy	Cool	Yes	Rainy	Hot	No	CR	3, 4														
Weather	Temperature	Play																																	
Sunny	Hot	No																																	
Sunny	Cool	Yes																																	
Overcast	Hot	Yes																																	
Rainy	Cool	Yes																																	
Rainy	Hot	No																																	

Qn No.	Question	CL	CO
	OR B) Create your own example to show how Apriori finds frequent patterns.		