

University Roll No. 63
Master of Computer Applications
MCAC203: Data Mining
Unique Paper Code: 223421207
Semester II
May 2025
Year of Admission: 2024

Time: Three Hours

Max. Marks: 70

Instructions: All questions are compulsory. Show all the intermediate steps to get full credits.

[10]

1.

Consider the given dataset. Perform the Agglomerative clustering algorithm using
 a. Complete-link and
 b. Centroid distance

Also, show the step-wise construction of the dendrogram for both clustering algorithms.

Data Points	X	Y
A	7	3
B	6	9
C	4	5
D	8	3
E	3	7
F	6	5

2. Consider the dataset given below. Apply k -Nearest-Neighbour classification ($k = 3$) to [6]
 classify the tuple <Height = 172 cm, Weight = 68 kg, Age = 31 years>.

Height (cm)	Weight (kg)	Age (yrs)	Class
160	55	25	Fit
170	65	30	Fit
180	85	35	Overweight
155	48	22	Fit
165	78	45	Overweight
175	60	28	Fit
182	90	38	Overweight
158	50	23	Fit
168	70	32	Overweight

3. Consider the dataset given below, using Naïve Bayes classifier, predict the class label for the instance with attributes: Color = "Green", Size = "Medium", Shape = "Round", Texture = "Smooth". [6]

Color	Size	Shape	Texture	Class
Red	Small	Round	Smooth	Fruit
Green	Medium	Round	Smooth	Fruit
Yellow	Large	Long	Rough	Vegetable
Green	Large	Long	Rough	Vegetable
Red	Small	Long	Rough	Vegetable
Yellow	Medium	Round	Smooth	Fruit
Red	Medium	Long	Rough	Vegetable
Green	Small	Round	Smooth	Fruit

4. Consider the dataset comprising four features, namely F1, F2, F3, and F4. Perform dimensionality reduction using Principal Component Analysis (PCA) to transform to 1D space. [9]

F1	F2	F3	F4
12	10	14	12
8	10	6	6

5. Given the following dataset, apply the FP-Growth algorithm to find the frequent itemset using minimum support count = 3. Show all the intermediate steps, including conditional FP-Trees construction.

[12]

TID	Items
101	Bread, Butter, Milk
102	Bread, Butter, Coffee, Eggs, Milk, Sugar
103	Bread, Coffee, Eggs, Sugar
104	Bread, Coffee, Eggs, Milk, Sugar
105	Coffee, Eggs, Milk, Sugar
106	Bread, Butter, Coffee, Milk, Sugar
107	Coffee, Eggs, Sugar

6. Build a decision tree using C4.5 algorithm on the dataset given below, and illustrate the process, including split selection and termination conditions.

[9]

Outlook	Temperature	Humidity	PlayTennis
Sunny	Hot	High	No
Sunny	Hot	High	No
Overcast	Hot	High	Yes
Rain	Mild	High	Yes
Rain	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes
Sunny	Mild	High	No
Sunny	Cool	Normal	Yes

7. Perform DBSCAN clustering on the given dataset with parameters $\epsilon = 2$ and MinPts = 4. Also, identify the outliers resulting from the clustering, if any.

[9]

Data Points	X	Y
P1	10	8
P2	9	9
P3	8	9
P4	7	8
P5	5	5
P6	6	5
P7	4	5
P8	4	6
P9	1	1
P10	7	6

8. Perform clustering on the given dataset using the BIRCH algorithm with $T < 1.5$ and MaxBranch=3.

[9]

Points	X	Y
P1	3	4
P2	2	6
P3	4	5
P4	4	7
P5	3	8
P6	6	2
P7	7	2
P8	7	4
P9	8	4
P10	7	9