

Hall Ticket Number:

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Code No. : 22968

VASAVI COLLEGE OF ENGINEERING (AUTONOMOUS), HYDERABAD

Accredited by NAAC with A++ Grade

M.Tech. (C.S.E.) II-Semester Main Examinations, August-2023

Data Mining

Time: 3 hours

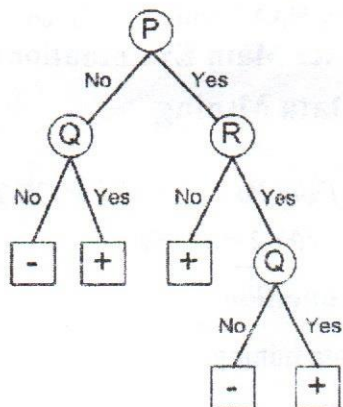
Max. Marks: 60

Note: Answer all questions from Part-A and any FIVE from Part-B

Part-A (10×2 = 20 Marks)

Q. No.	Stem of the question	M	L	CO	PO																																																							
1.	Define Slice and Dice operation in data mining.	2	1	1	1,2																																																							
2.	What are major issues of data mining?	2	1	1	1,2																																																							
3.	What is Decision tree?	2	1	2	1,2																																																							
4.	What is Rule-based classification in data mining?	2	1	2	1,2																																																							
5.	What are Bayesian network give an example?	2	1	3	1,2																																																							
6.	What is the use of support vectors in SVM?	2	1	3	1,2																																																							
7.	Write application of ANN?	2	1	4	1,2																																																							
8.	What is difference between supervised and unsupervised learning?	2	2	4	1,2																																																							
9.	What is Dendrogram Give an example?	2	1	5	1,2																																																							
10.	Explain density based clustering methods.	2	2	5	1,2																																																							
Part-B (5×8 = 40 Marks)																																																												
11. a)	Compute the cosine similarity for the following document vectors D1= (2,2,0,0,3,0,1) D2= (0,1,2,0,1,2,1)	4	3	1	1,2																																																							
b)	Compute the Distance between the two data objects X(22,1,24,10,46) and Y(12,2,24,23,46) using Manhattan, Euclidean distance	4	3	1	1,2																																																							
12. a)	Estimate the conditional probabilities for P(A +), P(B +), P(C +), P(A -), P(B -) and P(C -).	4	3	2	1,2																																																							
<table border="1"> <thead> <tr> <th>Record</th> <th>A</th> <th>B</th> <th>C</th> <th>Class</th> </tr> </thead> <tbody> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>+</td></tr> <tr><td>2</td><td>0</td><td>0</td><td>1</td><td>-</td></tr> <tr><td>3</td><td>0</td><td>1</td><td>1</td><td>-</td></tr> <tr><td>4</td><td>0</td><td>1</td><td>1</td><td>-</td></tr> <tr><td>5</td><td>0</td><td>0</td><td>1</td><td>+</td></tr> <tr><td>6</td><td>1</td><td>0</td><td>1</td><td>+</td></tr> <tr><td>7</td><td>1</td><td>0</td><td>1</td><td>-</td></tr> <tr><td>8</td><td>1</td><td>0</td><td>1</td><td>-</td></tr> <tr><td>9</td><td>1</td><td>1</td><td>1</td><td>+</td></tr> <tr><td>10</td><td>1</td><td>0</td><td>1</td><td>+</td></tr> </tbody> </table>						Record	A	B	C	Class	1	0	0	0	+	2	0	0	1	-	3	0	1	1	-	4	0	1	1	-	5	0	0	1	+	6	1	0	1	+	7	1	0	1	-	8	1	0	1	-	9	1	1	1	+	10	1	0	1	+
Record	A	B	C	Class																																																								
1	0	0	0	+																																																								
2	0	0	1	-																																																								
3	0	1	1	-																																																								
4	0	1	1	-																																																								
5	0	0	1	+																																																								
6	1	0	1	+																																																								
7	1	0	1	-																																																								
8	1	0	1	-																																																								
9	1	1	1	+																																																								
10	1	0	1	+																																																								

b)



4 3 2 1,2

Using above Decision Tree extract the classification rules and list it down

13. a) Explain the Back Propagation technique.

4 2 3 1,2

b) Explain the steps in Naïve bayes classification. Compute the class label for X= {age=31-40, student=yes, credit rating=fair, income=medium} by using Naive Bayesian classifier for the

4 4 3 1,2

RID	age	Income	Student	Credit rating	Buys computer
1	<=30	High	No	Fair	No
2	<=30	High	No	Excellent	No
3	31-40	High	No	Fair	Yes
4	>40	Medium	No	Fair	Yes
5	>40	Low	Yes	Fair	Yes
6	>40	Low	Yes	Excellent	No
7	31-40	Low	Yes	Excellent	Yes
8	<=30	Medium	No	Fair	No
9	<=30	Low	Yes	Fair	Yes
10	>40	Medium	Yes	Fair	Yes
11	<=30	Medium	Yes	Excellent	Yes
12	31-40	Medium	No	Excellent	Yes
13	31-40	High	Yes	Fair	Yes
14	>40	Medium	No	Excellent	No

following training data

14. a) Explain Association Algorithm In Data Mining? 4 2 4 1,2
- b) Apply the Apriori algorithm on the grocery store example with support threshold $s=33.34\%$ and confidence threshold $c=60\%$. Show the candidate and frequent item sets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence 4 3 4 1,2

Transaction ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

15. a) The Nearest Neighbor clustering algorithm and Euclidean distance to cluster if the threshold t is 4. 4 3 5 1,2
 $A1=(2,10)$, $A2=(2,5)$, $A3=(8,4)$, $A4=(5,8)$, $A5=(7,5)$, $A6=(6,4)$, $A7=(1,2)$, $A8=(4,9)$.
- b) Explain K-means algorithm with an example 4 2 5 1,2
16. a) Why we need to perform data preprocessing? Explain data preprocessing techniques. 4 2 1 1,2
- b) Consider the following dataset for a binary class problem calculate the information gain in the Gini index When splitting on attributes A and B which attribute would the decision tree induction algorithm choose? 4 3 2 1,2,3

A	B	Class Label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

17. Answer any *two* of the following:

a) Consider the following data set to predict class with k nearest neighbor

4 3 3 1,2

Name	Acid Durability	Strength	Class
Type-1	7	7	F
Type-2	7	4	F
Type-3	3	4	T
Type-4	1	4	T

For Acid Durability=3 and strength =7, class=?

b) Construct the Fp tree from the below transaction database.

4 3 4 1,2

TID	items bought
100	{f, a, c, d, g, i, m, p}
200	{a, b, c, f, l, m, o}
300	{b, f, h, j, o}
400	{b, c, k, s, p}
500	{a, f, c, e, l, p, m, n}

c) Describe the properties of Clustering Algorithm?

4 2 5 1,2

M : Marks; L: Bloom's Taxonomy Level; CO; Course Outcome; PO: Programme Outcome

i)	Blooms Taxonomy Level – 1	20%
ii)	Blooms Taxonomy Level – 2	30%
iii)	Blooms Taxonomy Level – 3 & 4	50%
