

THE KAVERY ENGINEERING COLLEGE*(An Autonomous Institution)***MCA DEGREE END SEMESTER THEORY EXAMINATIONS, APRIL-MAY 2025***Third Semester**Master of Computer Applications***MC4020 – Data Mining and Data Warehousing Techniques***Regulations - 2021***Duration : 3 Hrs****Maximum : 100 Marks****PART – A (ANSWER ALL QUESTIONS) (Marks 10*2=20)**

Q.No	Questions	BLT	CO	Marks
1.	Differentiate DBMS and Data Mining.	L2	1	2
2.	List the challenges of Data Mining.	L1	1	2
3.	Identify the two-step process of association rule mining.	L2	2	2
4.	Write the differences between Classification and prediction.	L2	2	2
5.	List the applications of Rule Based classification.	L2	3	2
6.	State the working principle of Support Vector Machine.	L2	3	2
7.	Define the term cluster analysis.	L1	4	2
8.	State the fuzzy cluster.	L1	4	2
9.	Specify the need for data warehouse.	L1	5	2
10.	Draw the data warehouse architecture.	L2	5	2

PART – B (ANSWER ALL QUESTIONS) (Marks 5*16 = 80)

Q.No	Questions	BLT	CO	Marks
11. a	With neat diagram illustrate data mining as a step in the process of knowledge discovery.	L2	1	16
	Or			
b	Give an overview about data preprocessing task and specify the major role of the task.	L2	1	16
12. a	Find the frequent item sets and generate association rules using the below data set. Assume that minimum support threshold ($s = 33.33\%$) and minimum confident threshold ($c = 60\%$).	L3	2	16

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

Or

- b A database has five transactions. Let $\text{min_sup}=60\%$ and $\text{min_conf}=80\%$. L3 2 16

TID	Items_bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all frequency item sets using Apriori and FP-growth, respectively.

13. a The golf data set given in table contains six attributes, namely Day, Outlook, Temperature, Humidity, Wind and Play Golf. L3 3 16

Day	Outlook	Temperature	Humidity	Wind	Play Golf
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	?

Using the above data, apply a decision tree to find whether the player will play the game or not.

Or

- b The marks for nine students on Mid term and End term examination has given in table. L3 3 16

Mid term	82	73	95	66	84	89	51	82	75
End term	76	83	89	76	79	73	62	89	77

Obtain the simple linear regression analysis to predict the score on the end term exam from the midterm exam score.

14. a The table below is a distance matrix for 6 objects. L3 4 16

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

Show the result using Hierarchical clustering method.

Or

- b Perform K-means clustering with initial centroids $m_1=(1.0,1.0)$ and $m_2=(5.0,7.0)$ on the following data. L3 4 16

X	1.0	1.5	3.0	5.0	3.5	4.5	3.5
Y	1.0	2.0	4.0	7.0	5.0	5.0	4.5

15. a With neat diagram illustrate the multidimensional data model. L2 5 16

Or

- b Write an overview about the OLAP different operations. L2 5 16