U.S.N | | | | | | | | | | |

P.E.S. College of Engineering, Mandya - 571 401

(An Autonomous Institution affiliated to VTU, Belagavi)

Second Semester, Master of Computer Applications (MCA)

Semester End Examination; October - 2022

Data Warehousing and Data Mining

Time: 3 hrs *Max. Marks: 100*

Course Outcomes

The Students will be able to:

CO1: Explain the basic concept of data warehousing and OLAP implementation.
CO2: Explain the basic concepts, techniques and application of data mining.
CO3: Illustrate the association rules to extract appropriate pattern in massive data.
CO4: Design and deploy appropriate classification techniques.
CO5: Illustrate the clustering techniques and outlier analysis in detail for better organization and retrieval of data.

Note: I) *Answer any* ***FIVE*** *full questions, selecting* ***ONE*** *full question from each unit.*
II) *Any* ***THREE*** *units will have internal choice and remaining* ***TWO*** *unit questions are compulsory.*
III) *Each unit carries 20 marks.*

Q. No.	Questions	Marks	BLs	COs	POs
	UNIT - I	**20**			
1 a.	Define data warehouse. Explain the framework of data warehouse with an example.	10	L2	CO1	PO2
b.	Distinguish between the following: i) OLTP and OLAP with an example for each ii) MOLAP and ROLAP	10	L1,2	CO1	PO1,2
	UNIT - II	**20**			
2 a.	Discuss with an example the proximity measures of; i) Jaccard coefficient ii) Cosine similarity	10	L6	CO2	PO2,3
b.	Justify the need for pre-processing techniques and explain any two pre-processing techniques.	10	L5	CO2	PO2,3
	UNIT - III	**20**			
3 a.	Discuss alternative methods for generating frequent item sets.	5	L6	CO3	PO2
b.	Define the following terms: i) Support ii) Confidence iii) Item sets iv) Minimum support threshold v) Minimum support count threshold	5	L2	CO3	PO2
c.	Illustrate Apriori algorithm to generate frequent item set with example.	10	L2	CO3	PO2
	OR				
3 d.	Discuss how to improve the efficiency of Apriori.(Any two techniques)	10	L6	CO3	PO1,2,3
e.	Discuss how frequent item set are generation. Explain with	10	L2	CO3	PO1,2,3

UNIT - IV **20**

4 a.	Discuss with algorithm how decision tree can be used for classification?	10	L2	CO4	PO2
b.	Define and explain Baye's Theorem.	5	L1,2	CO4	PO1,2
c.	Consider the following data set to predict the class label using Bayesian classification.	5	L5	CO4	PO1,2

RID	Age	Income	Student	Credit_rating	Class (Buys Computer)
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle age	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	No
6	Senior	Low	Yes	Excellent	No
7	Middle age	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes

Classify X = (Age = Youth, Income = Medium, Student = Yes, Credit_rating = Fair)

OR

4 d.	Define Rule based classifier. Discuss how a Rule based classifier works?	10	L1,2	CO4	PO2
e.	Discuss the characteristics of decision tree classifier.	10	L2	CO4	PO2

UNIT - V **20**

5 a.	What is Cluster analysis? Discuss the desired feature of cluster analysis method.	10	L2	CO5	PO 2
b.	Discuss the overview of basic clustering method.	10	L2	CO5	PO2

OR

5 d.	Differentiate between Agglomerative and Divisive hierarchical clustering.	10	L2	CO5	PO2
e.	Discuss the different computing distance measures that are used in hierarchical clustering.	5	L2	CO5	PO2
f.	Discuss K-means clustering algorithm.	5	L2	CO5	PO2

* * * *