USN

Sixth Semester B.E. Degree Examination, Jan./Feb. 2021
Data Mining and Data Warehousing

Time: 3 hrs.   Max. Marks: 80

Note: *Answer any FIVE full questions, choosing ONE full question from each module.*

Module-1

1	a.	Discuss the star and snowflake schema in detail with suitable example.	(05 Marks)
	b.	Compare Enterprise warehouse, Data mart and Virtual warehouse.	(06 Marks)
	c.	Distinguish between OLTP and OLAP.	(05 Marks)

OR

2	a.	Differentiate between Designing and Data warehouse and OLAP cube.	(05 Marks)
	b.	Define ETL and Data cleaning. Describe the data cleaning steps.	(06 Marks)
	c.	Compare between the MOLAP and HOLAP.	(05 Marks)

Module-2

3	a.	Explain with example : i) Euclidean distance and Coefficient ii) Jacord coefficient iii) Cosine similarity.	(06 Marks)
	b.	Describe the Feature Subset selection.	(04 Marks)
	c.	What are various methods for data reduction? Explain with examples.	(06 Marks)

OR

4	a.	Explain the stages involved in knowledge discovery process with a neat schematic diagram.	(06 Marks)
	b.	Discuss about data mining. Describe the challenges that motivated development of data mining.	(05 Marks)
	c.	Explain Sampling. Compare Linear sampling Vs Stratified sampling.	(05 Marks)

Module-3

5	a.	Explain FP – Growth algorithm for discovering frequent item sets. What are its limitations?	(08 Marks)
	b.	Define Apriori principle. Explain frequent item set generation algorithm.	(08 Marks)

OR

6	a.	What is Association Analysis? Explain in brief with an example.	(04 Marks)
	b.	Define the following with an example to each : i) Support of a rule ii) Confidence of a rule.	(04 Marks)
	c.	Consider the following transaction data set :	

Tid	1	2	3	4	5	6	7	8	9	10
Items	{a,b}	{b,c,d}	{a,c,d,e}	{a,d,e}	{a,b,c}	{a,b,c,d}	{a}	{a,b,c}	{a,b,d}	{b,c,e}

Construct the FP tree. Show the trees separately after reading each transaction.   (08 Marks)

Module-4

7	a.	What is Classification? Write decision tree induction algorithm.	(08 Marks)
	b.	What are Bayesian classifiers? Explain Bayes theorem for classification.	(08 Marks)

OR

8	a.	What are Rule based classifiers? Explain with an example.	(04 Marks)
	b.	Differentiate between classification and prediction.	(04 Marks)
	c.	What is the Nearest Neighbor classifiers? Write and explain nearest neighbor algorithm.	(08 Marks)

Module-5

9	a.	Explain Clustering Analysis methods briefly.	(04 Marks)
	b.	What are the features of Cluster analysis?	(04 Marks)
	c.	Write and explain the basic K – means algorithm of clustering.	(08 Marks)

OR

10	a.	What is Hierarchical Clustering method? Explain the algorithm for computing distance between clusters.	(08 Marks)
	b.	Explain DBScan algorithm with its time and space complexity.	(08 Marks)

* * * * *