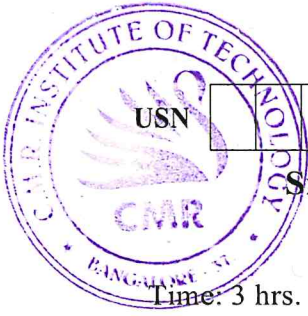


CBCS SCHEME



USN

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

15CS651

Sixth Semester B.E. Degree Examination, Aug./Sept.2020 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. What is a Data Warehouse? Discuss various usage and trends in data warehousing. (08 Marks)
- b. Explain in detail the three – tier Data warehouse architecture. (08 Marks)

OR

- 2 a. Discuss the concept of star, snowflake and galaxy schemas for multidimensional databases. (08 Marks)
- b. Summarize the various OLAP operations in the multidimensional data model. (08 Marks)

Module-2

- 3 a. Identify the different indexing method used for OLAP data with brief explanation. (08 Marks)
- b. Differentiate ROLAP , MOLAP and HOLAP servers. (08 Marks)

OR

- 4 a. What is Data Mining? Explain various data mining tasks, with suitable examples. (08 Marks)
- b. Explain different steps involved in preprocessing steps. Write any 3 challenges faced in Data Mining. (08 Marks)

Module-3

- 5 a. What is Apriori Algorithm? How it is used to find frequent item sets? Explain. (08 Marks)
- b. Illustrate the advantages of using closed frequent itemsets, with an example. Show the relationships among frequent, maximal frequent and closed frequent itemsets. (08 Marks)

OR

- 6 a. Explain FP – growth algorithm for discovering frequent itemsets. (08 Marks)
- b. Briefly explain the Objective Measures of Interestingness for evaluating association patterns. (08 Marks)

Module-4

- 7 a. Define Classification. With a neat figure, explain the general approach for solving classification model. (08 Marks)

CMRIT LIBRARY
BANGALORE - 560 037

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

- b. Consider the following data set for a binary.

Instance	a ₁	a ₂	a ₃	Target class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	-
4	F	F	4.0	+
5	F	T	7.0	-
6	F	T	3.0	-
7	F	F	8.0	-
8	T	F	7.0	+
9	F	T	5.0	-

Classification problem :

- What is the entropy of this collection of training examples, with respect to the positive class?
- What are the information gains of a₁ and a₂ relative to these training examples?
- For a₃, which is a continuous attribute, compute the information gain for every possible split. (08 Marks)

OR

- What is a rule – based classifier? Explain sequential covering algorithm in rule – based classifier. (08 Marks)
 - Write an algorithm for K – Nearest Neighbor (KNN) classification. List the characteristics of Nearest Neighbor classifiers. (08 Marks)

Module-5

- What is Cluster analysis? Discuss the different types of clusters with examples. (08 Marks)
 - Describe K – means clustering algorithm. What are its limitations? (08 Marks)

OR

- Discuss DBSCAN algorithm for clustering. (08 Marks)
 - Explain Agglomerative Hierarchical Clustering Algorithm, with different proximity between clusters. (08 Marks)

CMRIT LIBRARY
BANGALORE - 560 037