INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad - 500 043

MODEL QUESTION PAPER

B.Tech VI Semester End Examinations (Regular), May – 2020
Regulation: IARE-R16

DATA WAREHOUSING AND DATA MINING
(CSE/IT)

Time:3hours   **Max. Marks:70**

Answer ONE Question from each
MODULE All Questions Carry Equal
Marks
All parts of the question must be answered in one place only

MODULE – I

1. a) Explain mapping data warehouse with multiprocessor architecture with the concept of
        parallelism and data partitioning?   [7M]
    b) Discuss the motivation behind OLAP Mining?
        Explain the   performance problems with star schema?   [7M]

2. a) Describe various schemes used for the design of multidimensional data model?   [7M]
        briefly analyze different categories of measures?
    b) Describe concept hierarchy generation for Categorical data? Explain Data Warehouse   [7M]
        implementation of Indexing OLAP Data

MODULE – II

3. a) What are the major challenges of mining a huge amount of data (such as billions of   [7M]
        tuples) in comparisons with mining a small amount of data (such a few hundred tuples
        data set)?
    b) Describe Data discretization and concept hierarchy generation? State why Concept
        hierarchies are useful in data mining?   [7M]

4. a) Discuss Data mining functionalities and give examples of each using real-life
        database with which you are familiar?   [7M]

b) Explain how the evolution of database technology led to data mining? Write notes [7M] on various performance issues that are encountered in Data Mining?

MODULE – III

5. a) Find all Frequent item sets for the given training set using APRIORI and FP-Growth respectively. Compare the efficiency of the two mining process

TID	Items _brought
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}

[7M]

b) What is association rule Mining problem? State APRIORI principle?

6. a) What is the difference between mining frequent item sets with candidate generation and without candidate generation? Explain.

b) Explain difference between partitions based APRIORI and APRIORI algorithm? [7M]

MODULE –IV

7. a) Why tree pruning useful in decision tree induction? What is a drawback of using a separate set of tuples to evaluate pruning? [7M]

b) Why naive Bayesian classification is called naïve? Briefly outline the major ideas of naive Bayesian classification?

8. a) Discuss the methods that are commonly used to evaluate the performance of a Classifier? [7M]
   b) What is bootstrapping? State why it may improve the accuracy of decision tree induction. [7M]

MODULE – V

9. a) Many clustering algorithms handle either only numerical data or categorical data but not both? Analyze why is this case? Briefly explain EM clustering algorithm efficiency? [7M]

    b) Why is outlier mining important? Briefly describe different approaches behind outlier detection? [7M]

10. a) What is the goal of clustering? How does partitioning around medoids algorithm achieve this goal? [7M]

    b) Illustrate the strength and weakness of k-means in comparison with k-medoids algorithm? [7M]

CLOURSE OBJECTIVES:

I	Identifying necessity of Data Mining and Data Warehousing for the society.
II	Familiar with the process of data analysis, identifying the problems, and choosing the relevant models and algorithms to apply.
III	Develop skill in selecting the appropriate data mining algorithm for solving practical problems.
IV	Develop ability to design various algorithms based on data mining tools.
V	Create further interest in research and design of new Data Mining techniques and concepts.

CLOURSE LEARNING OUTCLOMES:

CLO Code	CLO's	At the end of the course, the student will have the ability to:
AIT006.01	CLO 1	Learn data warehouse principles and find the differences between relational Databases and data warehouse.
AIT006.02	CLO 2	Explore on data warehouse architecture and its components.
AIT006.03	CLO 3	Learn Data warehouse schemas.
AIT006.04	CLO 4	Differentiate different OLAP Architectures.
AIT006.05	CLO 5	Understand Data Mining concepts and knowledge discovery process.
AIT006.06	CLO 6	Explore on Data preprocessing techniques.
AIT006.07	CLO 7	Apply task related attribute selection and transformation techniques.
AIT006.08	CLO 8	Understand the Association rule mining problem.
AIT006.09	CLO 9	Illustrate the concept of Apriori algorithm for finding frequent items and generating association rules. Association rules.
AIT006.10	CLO 10	Illustrate the concept of FP-growth algorithm and different representations of frequent item sets.
AIT006.11	CLO 11	Understand the classification problem and prediction.
AIT006.12	CLO 12	Explore on decision tree construction and attribute selection.
AIT006.13	CLO 13	Understand the classification problem and Bayesian classification.
AIT006.14	CLO 14	Illustrate the rule based and back propagation classification algorithms.
AIT006.15	CLO 15	Understand the Cluster and Analysis.
AIT006.16	CLO 16	Understand the Types of data and categorization of major clustering methods.
AIT006.17	CLO 17	Explore on partition algorithms for clustering.
AIT006.18	CLO18	Explore on different hierarchical based methods, different density based methods, grid based and Model based methods.
AIT006.19	CLO19	Understand the outlier Analysis.
AIT006.20	CLO20	Understand mining complex data types.

MAPPING OF SEMESTER END EXAMINATION TO COURSE LEARNING OUTCOMES

SEE QUESTION No			COURSE LEARNING OUTCOMES	BLOOM TAXONOMY LEVELS
1	a	AIT006.01	Learn data warehouse principles and find the differences between relational Databases and data warehouse.	Understand
	b	AIT006.01	Learn data warehouse principles and find the differences between relational Databases and data warehouse.	Understand
2	a	AIT006.04	Differentiate different OLAP Architectures.	Remember
	b	AIT006.03	Learn Data warehouse schemas.	Understand
3	a	AIT006.06	Explore on Data preprocessing techniques.	Remember
	b	AIT006.08	Understand the Association rule mining problem.	Understand
4	a	AIT006.08	Understand the Association rule mining problem.	Remember
	b	AIT006.07	Apply task related attribute selection and transformation techniques.	Remember
5	a	AIT006.13	Understand the classification problem and Bayesian classification.	Remember
	b	AIT006.10	Illustrate the concept of FP-growth algorithm and different representations of frequent item sets.	Understand
6	a	AIT006.10	Illustrate the concept of FP-growth algorithm and different representations of frequent item sets.	Remember
	b	AIT006.13	Understand the classification problem and Bayesian classification.	Understand
7	a	AIT006.18.	Explore on different hierarchical based methods, different density based methods, grid based and Model based methods.	Remember
	b	AIT006.15	Understand the Cluster and Analysis.	Understand
8	a	AIT006.15	Understand the Cluster and Analysis.	Remember
	b	AIT006.18	Explore on different hierarchical based methods, different density based methods, grid based and Model based methods.	Understand
9	a	AIT006.20	Understand mining complex data types.	Remember
	b	AIT006.20	Understand mining complex data types.	Understand
10	a	AIT006.20	Understand mining complex data types.	Remember
	b	AIT006.20	Understand mining complex data types.	Understand

Signature of Course Coordinator   **HOD, CSE**