DATAWAREHOUSING AND DATA MINING

VI Semester: IT | CSE

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P	C	CIA	SEE	Total
AIT006	Core	3	1	-	4	30	70	100
Contact Classes: 45	Tutorial Classes: 15	Practical Classes: Nil				Total Classes: 60		

OBJECTIVES:

The course should enable the students to:

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.

COURSE LEARNING OUTCOMES (CLOs):

The students should enable to:

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

SYLLABUS

UNIT-I	DATA WAREHOUSING	Classes:09

Introduction to Data warehouse, Differences between OLAP and OLTP, A Multi dimensional data model- Star, Snow flake and Fact constellation schemas, Measures, Concept hierarchy, OLAP Operations in the Multidimensional Data Model, Data warehouse architecture- A three tier Data warehouse architecture, Data warehouse Back-End Tools and Utilities, Metadata Repository, types of OLAP servers, Data warehouse Implementation, Data Warehouse models- Enterprise warehouse

UNIT-II	DATA MINING	Classes:09

Introduction, what is Data Mining, Definition, Knowledge Discovery in Data (KDD), Kinds of data

bases, Data mining functionalities, Classification of data mining systems, Data mining task primitives, Data Preprocessing: Data cleaning, Data integration and transformation, Data reduction, Data discritization and Concept hierarchy.

UNIT-III	ASSOCIATION RULE MINING	Classes:09

Association Rules: Problem Definition, Frequent item set generation, The APRIORI Principle, support and confidence measures, association rule generation; APRIORI algorithm.
FP-Growth Algorithms, Compact Representation of Frequent item Set-Maximal Frequent item set, closed frequent item set.

UNIT-IV	CLASSIFICATION AND PRIDICTION	Classes:09

Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Classification by Back propagation, Classification Based on Concepts from Association Rule Mining, Other Classification Methods, Prediction, Classifier Accuracy.

UNIT-V	CLUSTERING	Classes:09

Types of data, categorization of major clustering methods, K-means partitioning methods, hierarchical methods, density based methods, grid based methods, model based clustering methods, outlier analysis.
Mining Complex Types of Data: Multi dimensional Analysis and Descriptive Mining of Complex, Data Objects, Mining Spatial Databases, Mining Multimedia Databases, Mining Time-Series and Sequence Data, Mining Text Databases, Mining the World Wide Web.

Text Books:

1. Jiawei Han, Michelin Kamber, “Data Mining-Concepts and techniques”, Morgan Kaufmann Publishers, Elsevier, 2nd Edition, 2006
2. Alex Berson, Stephen J.Smith, “Data warehousing Data mining and OLAP”, Tata McGraw- Hill, 2nd Edition, 2007

References:

1. Arum K Pujari, “Data Mining Techniques”, 3rd Edition, Universities Press, 2005
2. PualrajPonnaiah, Wiley, “Data Warehousing Fundamentals”, Student Edition, 2004.
3. Ralph Kimball, Wiley, “The Data warehouse Life Cycle Toolkit”, Student Edition, 2006.

Web References:

https://www.sanfoundry.com/best-reference-books-data-mining-data-warehousing/

E-Text Books:

https://www.wileyindia.com/data-warehousing-data-mining.html