

DATA MINING AND DATA WAREHOUSING
[As per Choice Based Credit System (CBCS) scheme]
(Effective from the academic year 2016 -2017)
SEMESTER – VI

Subject Code	15CS651	IA Marks	20
Number of Lecture Hours/Week	3	Exam Marks	80
Total Number of Lecture Hours	40	Exam Hours	03

CREDITS – 03

Course objectives: This course will enable students to

- Define multi-dimensional data models.
- Explain rules related to association, classification and clustering analysis.
- Compare and contrast between different classification and clustering algorithms

Module – 1

Teaching Hours

Data Warehousing & modeling: Basic Concepts: Difference between Operational Database systems and Data warehouse, Data Warehousing: A multitier Architecture, Data warehouse models: Enterprise warehouse ,Data mart and virtual warehouse, Extraction, Transformation and loading, Metadata Repository, Data warehouse design and usage: Business Analysis framework, Data warehouse design process and usage for information processing, Online analytical processing to multidimensional data mining. Data Cube: A multidimensional data model, Stars, Snowflakes and Fact constellations: Schemas for multidimensional Data models, Dimensions: The role of concept Hierarchies, Measures: Their Categorization and computation, Typical OLAP Operations.

8 Hours

Module – 2

Data warehouse implementation& Data mining: Efficient Data Cube computation: An overview, Indexing OLAP Data: Bitmap index and join index, Efficient processing of OLAP Queries, OLAP server Architecture ROLAP versus MOLAP Versus HOLAP. : Introduction: What is data mining, Challenges, Data Mining Tasks, Data: Types of Data, Data Quality, Data Preprocessing, Measures of Similarity and Dissimilarity,

8 Hours

Module – 3

Association Analysis: Association Analysis: Problem Definition, Frequent Item set Generation, Rule generation. Alternative Methods for Generating Frequent Item sets, FP-Growth Algorithm, Evaluation of Association Patterns.

8 Hours

Module – 4

Classification : Basics: General approach to solve classification problem, Decision Trees Induction, Model Over fitting, Evaluating the performance of a classifier, Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbor Classifiers, Bayesian Classifiers.

8 Hours

Module – 5

Clustering Analysis: Overview, K-Means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Density-Based Clustering, Graph-Based Clustering, Scalable Clustering Algorithms.

8 Hours

Course outcomes: The students should be able to:

- Identify data mining problems and implement the data warehouse

- Write association rules for a given data pattern.
- Choose between classification and clustering solution.

Question paper pattern:

The question paper will have TEN questions.

There will be TWO questions from each module.

Each question will have questions covering all the topics under a module.

The students will have to answer FIVE full questions, selecting ONE full question from each module.

Text Books:

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar: Introduction to Data Mining, Pearson, First impression, 2014.
2. Jiawei Han, Micheline Kamber, Jian Pei: Data Mining - Concepts and Techniques, 3rd Edition, Morgan Kaufmann Publisher, 2012.

Reference Books:

1. Sam Anahory, Dennis Murray: Data Warehousing in the Real World, Pearson, Tenth Impression, 2012.
2. Michael J. Berry, Gordon S. Linoff: Mastering Data Mining, Wiley Edition, second edition, 2012.