DATA MINING AND DATA WAREHOUSING
(Effective from the academic year 2018 -2019)
SEMESTER – VI

Course Code	**18CS641**	**CIE Marks**	40
Number of Contact Hours/Week	3:0:0	**SEE Marks**	60
Total Number of Contact Hours	40	**Exam Hours**	03

CREDITS –3

Course Learning Objectives: This course (18CS641) 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	**Contact Hours**
Data Warehousing & modeling: Basic Concepts: Data Warehousing: A multitier Architecture, Data warehouse models: Enterprise warehouse, Data mart and virtual warehouse, Extraction, Transformation and loading, 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 **Textbook 2: Ch.4.1,4.2** **RBT: L1, L2, L3**	08
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. **Textbook 2: Ch.4.4** **Textbook 1: Ch.1.1,1.2,1.4, 2.1 to 2.4** **RBT: L1, L2, L3**	08
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. **Textbook 1: Ch 6.1 to 6.7 (Excluding 6.4)** **RBT: L1, L2, L3**	08
Module 4	
Classification : Decision Trees Induction, Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbor Classifiers, Bayesian Classifiers. **Textbook 1: Ch 4.3,4.6,5.1,5.2,5.3** **RBT: L1, L2, L3**	08
Module 5	
Clustering Analysis: Overview, K-Means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Density-Based Clustering, Graph-Based Clustering, Scalable Clustering Algorithms. **Textbook 1: Ch 8.1 to 8.5, 9.3 to 9.5** **RBT: L1, L2, L3**	08

Course Outcomes: The student will 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.
- Each full Question consisting of 20 marks
- There will be 2 full questions (with a maximum of four sub questions) from each module.
- Each full question will have sub questions covering all the topics under a module.
- The students will have to answer 5 full questions, selecting one full question from each module.

Textbooks:

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 edtion,2012.

Dept. Of Information [illegible] & Engineering
Alva's Institute of [illegible] Technology
Mijar, MOODBIDRI-574 225