

<b>DATA MINING AND DATA WAREHOUSING</b> <b>[As per Choice Based Credit System (CBCS) scheme]</b> <b>(Effective from the academic year 2017 - 2018)</b> <b>SEMESTER – VI</b>			
Subject Code	17CS651	IA Marks	40
Number of Lecture Hours/Week	3	Exam Marks	60
Total Number of Lecture Hours	40	Exam Hours	03
<b>CREDITS – 03</b>			
<b>Module – 1</b>			<b>Teaching Hours</b>
<b>Data Warehousing&amp;modeling:</b> 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.			<b>8 Hours</b>
<b>Module – 2</b>			
<b>Data warehouse implementation&amp; Data mining:</b> 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,			<b>8 Hours</b>
<b>Module – 3</b>			
<b>Association Analysis:</b> Association Analysis: Problem Definition, Frequent Item set Generation, Rule generation. Alternative Methods for Generating Frequent Item sets, FP-Growth Algorithm, Evaluation of Association Patterns.			<b>8 Hours</b>
<b>Module – 4</b>			
<b>Classification :</b> Decision Trees Induction,Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbor Classifiers,Bayesian Classifiers.			<b>8 Hours</b>
<b>Module – 5</b>			
<b>Clustering Analysis:</b> Overview, K-Means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Density-Based Clustering, Graph-Based Clustering, Scalable Clustering Algorithms.			<b>8 Hours</b>
<b>Course outcomes:</b> The students should be able to:			
<ul style="list-style-type: none"> <li>• Understands data mining problems and implement the data warehouse</li> <li>• Demonstrate the association rules for a given data pattern.</li> <li>• Discuss between classification and clustering solution.</li> </ul>			
<b>Question paper pattern:</b> 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.			
<b>Text Books:</b>			
1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar: Introduction to Data Mining,			

Pearson, First impression,2014.

- 
2. Jiawei Han, MichelineKamber, Jian Pei: Data Mining -Concepts and Techniques, 3<sup>rd</sup> 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.