

## DATA MINING AND KNOWLEDGE DISCOVERY

VI SEMESTER: CSE & CS & CSIT																										
Course Code	Category	Hours / Week			Credits	Maximum Marks																				
ACIC01	Core	L	T	P	C	CIA	SEE	Total																		
		3	1	0	4	30	70	100																		
<b>Contact Classes: 45</b>		<b>Tutorial Classes: 15</b>		<b>Practical Classes: Nil</b>		<b>Total Classes:60</b>																				
<b>Prerequisite: Database Management Systems</b>																										
<p><b>I. COURSE OVERVIEW:</b>            Data mining refers to extracting or mining knowledge from large amounts of data. It emphasizes various techniques and algorithms used to explore, analyze and leverage data and turn it into valuable and actionable information. It includes data warehousing and data mining functionalities such as association mining, classification, clustering and outlier analysis. The techniques are used to tackle data centric applications in various domains such as financial analysis, telecommunication industry, intrusion detection, and complex data mining applications in stream, web, text, spatial and other scientific applications.</p> <p><b>II. COURSE OBJECTIVES:</b>  <b>The students will try to learn:</b>            I. The scope and essentiality of data warehousing and mining.            II. The analyses of data, choosing relevant models and algorithms for respective applications.            III. The process and mining of complex data types such as streams, spatial, web and multimedia.            IV. The research perspectives towards advances in data mining.</p> <p><b>III. COURSE OUTCOMES:</b>  <b>After successful completion of the course, students should be able to:</b></p> <table border="0"> <tr> <td>CO 1</td> <td>Relate knowledge discovery in databases (KDD) process with the help of data warehouse fundamentals and data mining functionalities</td> <td>Understand</td> </tr> <tr> <td>CO 2</td> <td>Select appropriate preprocessing techniques on real time data for usage of data mining algorithms</td> <td>Apply</td> </tr> <tr> <td>CO 3</td> <td>Apply Apriori and FP growth methods on transaction data for frequent pattern mining</td> <td>Apply</td> </tr> <tr> <td>CO 4</td> <td>Choose classification or clustering algorithm for building a classification or prediction model.</td> <td>Apply</td> </tr> <tr> <td>CO 5</td> <td>Infer complex data models with respect to multimedia, streams, spatial and web mining</td> <td>Understand</td> </tr> <tr> <td>CO 6</td> <td>Examine data mining algorithms for solving real world problems</td> <td>Analyze</td> </tr> </table> <p><b>IV. COURSE SYLLABUS:</b></p> <p><b>MODULE – I: KNOWLEDGE DISCOVERY (09)</b>            Data Mining definition, knowledge discovery in data (KDD), kinds of data can be mined, kinds of patterns / data mining functionalities, technologies, applications, issues in data mining. data objects and attribute types, basic statistical descriptions of data, data visualization, measuring data similarity and dissimilarity.</p> <p><b>MODULE – II: DATA PREPROCESSING (08)</b>            Data Preprocessing: Data quality, major tasks in data preprocessing, data cleaning, data integration and transformation, data reduction, data discretization.</p> <p><b>MODULE – III: DATA WAREHOUSING AND ONLINE ANALYTICAL PROCESSING (09)</b>            Data warehouse concepts, differences between operational database systems and data warehouses, a multitiered architecture; Data Warehouse Models: Enterprise warehouse, data mart, and virtual warehouse, extraction, transformation, and loading, metadata repository, a multidimensional data model; Schemas for Multidimensional Data Models: Stars, snowflakes, and fact constellations, dimensions, measures, OLAP operations, a star query model for querying multidimensional databases.</p>									CO 1	Relate knowledge discovery in databases (KDD) process with the help of data warehouse fundamentals and data mining functionalities	Understand	CO 2	Select appropriate preprocessing techniques on real time data for usage of data mining algorithms	Apply	CO 3	Apply Apriori and FP growth methods on transaction data for frequent pattern mining	Apply	CO 4	Choose classification or clustering algorithm for building a classification or prediction model.	Apply	CO 5	Infer complex data models with respect to multimedia, streams, spatial and web mining	Understand	CO 6	Examine data mining algorithms for solving real world problems	Analyze
CO 1	Relate knowledge discovery in databases (KDD) process with the help of data warehouse fundamentals and data mining functionalities	Understand																								
CO 2	Select appropriate preprocessing techniques on real time data for usage of data mining algorithms	Apply																								
CO 3	Apply Apriori and FP growth methods on transaction data for frequent pattern mining	Apply																								
CO 4	Choose classification or clustering algorithm for building a classification or prediction model.	Apply																								
CO 5	Infer complex data models with respect to multimedia, streams, spatial and web mining	Understand																								
CO 6	Examine data mining algorithms for solving real world problems	Analyze																								

Business Analysis framework for data warehouse design, data warehouse design process, data warehouse implementation, indexing OLAP data, OLAP server architectures, data generalization by attribute, oriented induction

#### **MODULE – IV: MINING FREQUENT PATTERNS AND CLASSIFICATION (10)**

Market basket analysis, frequent itemsets, closed itemsets, and association rules, frequent itemset mining methods; Apriori algorithm, generating association rules from frequent itemsets, improving the efficiency of Apriori Pattern-Growth Approach. Classification: Basic concepts, decision tree induction, Bayesian belief networks, classification by back propagation, support vector machines, classification using frequent patterns, lazy learners, other classification methods, model evaluation and selection, techniques to improve classification accuracy.

#### **MODULE –V: CLUSTERING AND RESEARCH FRONTIERS (09)**

Cluster Analysis, Partitioning methods, hierarchical methods, density-based methods, grid based methods, evaluation of clustering.

Mining Complex Types of Data: Mining Sequence Data: Time-series, symbolic sequences, and biological sequences, mining graphs and networks.

#### **V. TEXT BOOKS:**

1. Jiawei Han, MichelineKamber, “Data Mining, Concepts and Techniques”, Morgan Kaufmann Publishers, Elsevier, 3<sup>rd</sup> Edition, 2012.
2. Alex Berson, Stephen J.Smith, “Data warehousing Data mining and OLAP”, Tata McGraw-Hill, 2<sup>nd</sup> Edition, 2007.

#### **VI. REFERENCE BOOKS:**

1. Arum K Pujari, “Data Mining Techniques”, Universities Press, 3<sup>rd</sup> Edition, 2005.
2. PualrajPonnaiah, “Data Warehousing Fundamentals”, Wiley, Student Edition, 2004.
3. Ralph Kimball, “The Data Warehouse Life Cycle Toolkit”, Wiley, Student Edition, 2006.
4. VikramPudi, P Radha Krishna, “Data Mining”, Oxford University, 1<sup>st</sup> Edition, 2007.

#### **VII. WEB REFERENCES:**

1. <http://www.anderson.ucla.edu>
2. <https://www.smartzworld.com>
3. <http://iiscs.wssu.edu>