

## DATA WAREHOUSING AND DATA MINING

| <b>VI Semester: CSE / IT</b>  |  |                             |   |                               |         |                         |     |                    |
|---|--|-----------------------------|---|-------------------------------|---------|-------------------------|-----|--------------------|
| Course Code   | Category   | Hours / Week                |   |                               | Credits | Maximum Marks           |     |                    |
|   |  | L                           | T | P                             |         | C                       | CIA | SEE                |
| ACSB14  | Core   | 2                           | 1 | -                             | 3       | 30                      | 70  | 100                |
| <b>Contact Classes: 30</b>  |  | <b>Tutorial Classes: 15</b> |   | <b>Practical Classes: Nil</b> |         | <b>Total Classes:45</b> |     |                    |
| <b>I. COURSE OVERVIEW:</b>  |  |                             |   |                               |         |                         |     |                    |
| <p>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 analytical processing, descriptive analysis, 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> |  |                             |   |                               |         |                         |     |                    |
| <b>II. OBJECTIVES:</b>  |  |                             |   |                               |         |                         |     |                    |
| <b>The course should enable the students to:</b>  |  |                             |   |                               |         |                         |     |                    |
| <ul style="list-style-type: none"> <li>I The scope and essentiality of data warehousing and mining.</li> <li>II The analysis of data, choosing relevant models and algorithms for respective applications.</li> <li>III The process and mining of complex data types such as streams, spatial, web and multimedia</li> <li>IV The research perspectives towards advances in data mining</li> </ul>  |  |                             |   |                               |         |                         |     |                    |
| <b>III. COURSE OUTCOMES:</b>  |  |                             |   |                               |         |                         |     |                    |
| <b>After successful completion of the course, students should be able to:</b>   |  |                             |   |                               |         |                         |     |                    |
| 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 Apriority 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                     |   |                               |         |                         |     |                    |
| <b>IV. SYLLABUS:</b>  |  |                             |   |                               |         |                         |     |                    |
| <b>MODULE-I</b>   | <b>DATA WAREHOUSING</b>  |                             |   |                               |         |                         |     | <b>Classes: 08</b> |
| <p>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 Multi dimensional data model, Data warehouse architecture- A three tier Data warehouse architecture, types of OLAP servers, Data warehouse Implementation, Data warehouse models.</p>  |  |                             |   |                               |         |                         |     |                    |
| <b>MODULE-II</b>  | <b>DATA MINING</b>   |                             |   |                               |         |                         |     | <b>Classes: 10</b> |
| <p>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 objects. Data Preprocessing: Data cleaning, Data integration and transformation, Data reduction, Data discretization and Concept hierarchy.</p>  |  |                             |   |                               |         |                         |     |                    |

|  |                                      |                    |
|--|--------------------------------------|--------------------|
| <b>MODULE-III</b>  | <b>ASSOCIATION RULE MINING</b>       | <b>Classes: 10</b> |
| <p>Association Rules: Problem Definition, Frequent item set generation, The APRIORI Principle, support and confidence measures, association rule generation; APRIORI algorithm.</p> <p>FP-Growth Algorithm, Compact Representation of Frequent item Set-Maximal Frequent item set, closed frequent item set.</p>   |                                      |                    |
| <b>MODULE-IV</b>   | <b>CLASSIFICATION AND PREDICTION</b> | <b>Classes: 10</b> |
| <p>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.</p>   |                                      |                    |
| <b>MODULE-V</b>  | <b>CLUSTERING</b>                    | <b>Classes: 07</b> |
| <p>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.</p> <p>Mining Complex Types of Data: Multidimensional 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.</p>                                    |                                      |                    |
| <b>Text Books:</b>   |                                      |                    |
| <ol style="list-style-type: none"> <li>1. Jiawei Han, Micheline Kamber, "Data Mining-Concepts and Techniques", Morgan Kaufmann Publishers, Elsevier, 2<sup>nd</sup> Edition, 2006.</li> <li>2. Alex Berson, Stephen J. Smith, "Data Warehousing Data Mining and OLAP", Tata McGraw-Hill, 2<sup>nd</sup> Edition, 2007.</li> </ol>  |                                      |                    |
| <b>Reference Books:</b>  |                                      |                    |
| <ol style="list-style-type: none"> <li>1. Arun K Pujari, "Data Mining Techniques", Universities Press, 3<sup>rd</sup> Edition, 2005</li> <li>2. Pualraj Ponnaiah, "Data Warehousing Fundamentals", Wiley, Student Edition, 2004.</li> <li>3. Ralph Kimball, "The Data Warehouse Life Cycle Toolkit", Wiley, Student Edition, 2006.</li> <li>4. Vikram Pudi, P Radha Krishna, "Data Mining", Oxford University, 1<sup>st</sup> Edition, 2007.</li> </ol>  |                                      |                    |
| <b>Web References:</b>   |                                      |                    |
| <ol style="list-style-type: none"> <li>1. <a href="http://www.anderson.ucla.edu">http://www.anderson.ucla.edu</a></li> <li>2. <a href="https://www.smartzworld.com">https://www.smartzworld.com</a></li> <li>3. <a href="http://iiscs.wssu.edu">http://iiscs.wssu.edu</a></li> </ol>   |                                      |                    |
| <b>E-Text Books:</b>   |                                      |                    |
| <ol style="list-style-type: none"> <li>1. <a href="https://www.cisco.com/application/pdf/en/us/guest/products/ps2011/c2001/ccmigration_09186a00802342cf.pdf">https://www.cisco.com/application/pdf/en/us/guest/products/ps2011/c2001/ccmigration_09186a00802342cf.pdf</a><a href="https://www.jntubook.com">https://www.jntubook.com</a></li> <li>2. <a href="http://ftp.utcluj.ro/pub/users/cemil/dwdm/dwdm_Intro/0_5311707.pdf">http://ftp.utcluj.ro/pub/users/cemil/dwdm/dwdm_Intro/0_5311707.pdf</a>.</li> </ol> |                                      |                    |
| <b>MOOC Course</b>   |                                      |                    |
| <a href="https://3ca1513rbm.wordpress.com">https://3ca1513rbm.wordpress.com</a>  |                                      |                    |