

## DATA MINING, WAREHOUSE AND VISULIZATION

IV Semester: MBA								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
CMBC59	Elective	L	T	P	C	CIA	SEE	Total
		4	-	-	4	30	70	100
<b>Contact Classes:45</b>		<b>Tutorial Classes: Nil</b>		<b>Practical Classes: Nil</b>		<b>Total Classes: 45</b>		
<p><b>I. COURSE OVERVIEW:</b>            This course helps the learner to practically understand a data mining, warehouse and visualization techniques and methods for data gathering and data pre-processing using OLAP tools. The different data mining models and managing warehouse efficiency, classification problems and logistics regression and data visualization for business.</p> <p><b>II. OBJECTIVES:</b>  <b>The students will try to learn:</b></p> <ol style="list-style-type: none"> <li>I. The concepts of data mining and preprocessing of data.</li> <li>II. Insights on association rule mining and clustering.</li> <li>III. The warehouse management efficiency and it operations, process and risk management.</li> <li>IV. The application of logistic regression and sentiment analytics to solve business problems.</li> <li>V. Data visualization for business and creation of dashboard.</li> </ol> <p><b>III. COURSE OUTCOMES:</b>  <b>After successful completion of the course, students will be able to:</b></p> <p><b>CO 1:</b> Summarizethe kinds of data that can be mined-database data and design data warehouses, transactional data for any organization.</p> <p><b>CO 2:</b> Extractknowledge using data mining techniques and experiment various methods of database types.</p> <p><b>CO 3:</b> Evaluate the overview of data preprocessing and apply in business.</p> <p><b>CO 4:</b> Analyze the case study of Handling Missing Values in Melbourne Housing Price Data.</p> <p><b>CO 5:</b> Discuss the picking methods and measure the warehouse efficiency</p> <p><b>CO 6:</b> Design and develop the cross docking, warehousing operations and order mixing to reduce the risk</p> <p><b>CO 7:</b> Evaluate the logistic regression and apply regression and printing model to test the data forbusiness operation.</p> <p><b>CO 8:</b> Apply different Cost-Based Approach understand the case study to reduce the cost.</p> <p><b>CO 9:</b> Describe the data visualization for businesses and numerical and non-numerical data for innovation in business</p> <p><b>CO 10:</b> Apply the creation dashboards using it tools to monitor the business activity.</p>								
<b>IV. SYLLABUS</b>								
<b>UNIT-I</b>	<b>INTRODUCTION TO DATA MINING, WAREHOUSE AND VISULIZATION</b>						<b>Classes: 08</b>	
Data Mining; Kinds of data that can be mined-Database Data, Data Warehouses, Transactional Data, Other Kinds of Data; Major Issues in Data Mining-Mining Methodology, User Interaction, Efficiency and Scalability, Diversity of Database Types, Data Mining and Society.								
<b>UNIT-II</b>	<b>DATA PRE-PROCESSING</b>						<b>Classes: 10</b>	
Data Preprocessing: An Overview-Reasons to process the data, Major Tasks in Data Preprocessing; Data Cleaning-Missing Values, Noisy Data, Data Cleaning as a Process; Data Reduction-Principal Component Analysis, Histograms, Clustering, Sampling, Data Cube Aggregation; Data Transformation and Data Discretization-Data Transformation by Normalization, Discretization by Binning, Discretization by Histogram Analysis. Case Study: Handling Missing Values in Melbourne Housing Price Data.								

<b>UNIT-III</b>	<b>MANAGING WAREHOUSE EFFICIENCY</b>	<b>Classes: 09</b>
<p>Order picking - Picking methods-pick path - Measuring Warehouse Efficiency - Warehouse Workforce design and development - cross docking.</p> <p>Warehousing Operations: warehousing operations- inbound process, outbound processes, Functions of Warehouse- break-bulk, cross docking, order mixing, Risk management.</p>		
<b>UNIT-IV</b>	<b>CLASSIFICATION PROBLEMS – I-LOGISTICS REGRESSION</b>	<b>Classes: 10</b>
<p>Overview of Classification; Binary Logistic Regression; Classification-Encoding Categorical Features, Building Logistic Regression Model, Printing Model Summary, Predicting on Test Data; Measuring Accuracies-Creating Confusion Matrix, Receiver Operating Characteristic (ROC) and Area Under the Curve; Finding Optimal Classification Cut-off -Youden’s Index and Cost-Based Approach.</p> <p>Case Study: Predicting Employee Attrition on HR Attrition Dataset.</p>		
<b>UNIT-V</b>	<b>DATA VISUALISATION FOR BUSINESS</b>	<b>Classes: 08</b>
<p>Introduction to Data Visualization for businesses, Visualization of Numerical and Non-Numerical Data Creation of Dashboards using IT Tools, Business Activity Monitoring through Dashboard.</p>		
<b>Text Books:</b>		
<ol style="list-style-type: none"> <li>1. Ikvinderpal “Data Mining &amp; Warehousing”, 1<sup>st</sup> Edition, 2020.</li> <li>2. Max Bramer and Springer “Principles of Data Mining 2020” 8<sup>th</sup> Edition, 2020.</li> <li>3. Szabo, Gungar Polatkan, Oscar Boykin, Chalkiopoulos, “Social Media Data Mining and Analytics”, Wiley, 3<sup>rd</sup> Edition, 2019.</li> <li>4. Pang-Ning Tan, “Introduction to Data Mining, Global Edition”, 4<sup>th</sup> Edition, May 2019.</li> <li>5. Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal, “Data Mining: Practical Machine Learning Tools and Techniques”, Elsevier, 4<sup>th</sup> Edition, 2017.</li> <li>6. Megan Squire, “Mastering Data Mining with Python –Find patterns hidden in your data”, 1<sup>st</sup> Edition. PACKT Publishing, 2016.</li> <li>7. Charu C. Aggarwal, “Data Mining and Warehousing”, 3<sup>rd</sup> Edition, 2016.</li> <li>8. Khushboo Saxena Sandeep Saxena Akash Saxena, “Data Mining and Warehousing”, BPB Publications, 3<sup>rd</sup> Edition, 2015.</li> <li>9. Florin Gorunescu, “Data Mining: Concepts, Models and Techniques”, Vol 12, Springer, 2011.</li> </ol>		
<b>Reference Books:</b>		
<ol style="list-style-type: none"> <li>1. Luis Torgo, “Data Mining with R: Learning with Case Studies”, CRC Press, 2<sup>nd</sup> Edition, 2011.</li> <li>2. Jiawei Han, Jian Pei, Micheline Kamber, “Data Mining: Concepts and Techniques”, Elsevier, 3<sup>rd</sup> Edition, 2010.</li> <li>3. Joseph B. Pigus, “Data Mining with Neural Networks”, 2<sup>nd</sup> Edition, TMH, 2017.</li> <li>4. Robert Layton, “Learning Data Mining with Python”, PACKT Publishing, 2<sup>nd</sup> Edition, 2015.</li> <li>5. Xin-She Yang, “Introduction to Algorithms for Data Mining and ML”, Academic Press, 1<sup>st</sup> Edition, 2019.</li> <li>6. Boris Kovalerchuk, Evgeni Vityaev, “Data Mining in Finance”, Kluwer Academic Publishers, 3<sup>rd</sup> Edition, 2010.</li> <li>7. Dinesh K., “Business Analytics: The Science of Data - Driven Decision Making”, Wiley, 1<sup>st</sup> Edition, 2009.</li> <li>8. Yau, N., “Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics”. John Wiley &amp; Sons.</li> </ol>		
<b>Web References:</b>		
<ol style="list-style-type: none"> <li>1. <a href="https://www.amazon.in/dp/B07YG4QSZR/ref=cm_sw_r_apan_glt_2FGRNQAEBE4AEV5JQRB7?_encoding=UTF8&amp;psc=1">https://www.amazon.in/dp/B07YG4QSZR/ref=cm_sw_r_apan_glt_2FGRNQAEBE4AEV5JQRB7?_encoding=UTF8&amp;psc=1</a></li> </ol>		
<b>E-Text Books:</b>		
<ol style="list-style-type: none"> <li>1. <a href="https://www.amazon.in/dp/B00UVBJSAQ/ref=cm_sw_r_apan_glt_1Y1H0P2MRK1KM0SFG1AJ?_encoding=UTF8&amp;psc=1">https://www.amazon.in/dp/B00UVBJSAQ/ref=cm_sw_r_apan_glt_1Y1H0P2MRK1KM0SFG1AJ?_encoding=UTF8&amp;psc=1</a>.</li> <li>2. <a href="https://www.amazon.in/dp/B075GB7FT5/ref=cm_sw_r_apan_glt_G8RNA0P6W9YSQZ2N0P16?_encoding=UTF8&amp;psc=1">https://www.amazon.in/dp/B075GB7FT5/ref=cm_sw_r_apan_glt_G8RNA0P6W9YSQZ2N0P16?_encoding=UTF8&amp;psc=1</a>.</li> </ol>		