

GANPAT UNIVERSITY									
FACULTY OF ENGINEERING & TECHNOLOGY									
Programme		Bachelor of Technology			Branch/Specc.		Computer Science & Engineering (CBA/CS/BDA/CSE)		
Semester		VI			Version		1.0.0.1		
Effective from Academic Year			2024-25		Effective for the batch Admitted in			June 2022	
Subject code		2CSE60E27	Subject Name		Data Mining & Warehousing				
Teaching scheme					Examination scheme (Marks)				
(Per week)	Lecture (DT)		Practical (Lab.)		Total		CE	SEE	Total
	L	TU	P	TW					
Credit	2	0	1	0	3	Theory	40	60	100
Hours	2	0	2	0	4	Practical	60	40	100
Pre-requisites:									
Basics of Probability & Statistics, Python/JAVA Programming Language, Database Management System									
Learning Outcomes:									
After successful completion of the course, the student will be able to:									
<ul style="list-style-type: none"> <li>Learn the process of data cleaning and pre-processing for mining applications</li> <li>Apply various mining techniques on various types of data</li> <li>Distinguish problems related to classification and clustering</li> <li>Evaluate accuracy of various classification and clustering algorithms</li> </ul>									
Theory syllabus									
<b>Unit</b>	<b>Content</b>								<b>Hrs</b>
<b>1</b>	<b>Data Warehousing</b> Basic Concepts, Data Cube and OLAP, Design and Usage, ETL, Implementation								<b>5</b>
<b>2</b>	<b>Data Mining Basics</b> Importance, Functionalities, Classification, Architecture, Major Issues, Data mining metrics, Applications, Social impacts of data, Data Mining from a Database Perspective								<b>3</b>
<b>3</b>	<b>Data Pre-processing</b> Descriptive Data Summarization, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization								<b>3</b>
<b>4</b>	<b>Association Rules</b> Algorithms, Advanced Association Rule Techniques, Measuring the Quality of Rules								<b>5</b>
<b>5</b>	<b>Classification</b> Basic issues regarding classification and prediction, Decision Tree, Bayesian classification, K Nearest Neighbour Algorithm, Associative classification, Statistical-Based Algorithms, Rule-Based Classification								<b>7</b>
<b>6</b>	<b>Clustering</b> Similarity and Distance Measures, Hierarchical Algorithms, Partitioned Algorithms, Clustering Large Databases, Clustering with Categorical Attributes								<b>7</b>
<b>Self-Study Topics</b>									
Data Mining Applications, Mining Event Sequences, Visual DM, The WEKA data mining Workbench									

Practical Content												
Practical contents will be based on following concepts like Web Scraping, Text Mining, Implementation of Classification, Association Rule based and Clustering Algorithms												
Mooc Course												
Course Name: Data Mining												
Link: <a href="https://onlinecourses.nptel.ac.in/noc21_cs06/">https://onlinecourses.nptel.ac.in/noc21_cs06/</a>												
Text Books												
1	J. Han and M. Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufman											
Reference Books												
1	Alex Berson, Stephen J. Smith, "Data Warehousing, Data Mining, and OLAP", MGH											
Course Outcomes:												
COs	Description											
CO1	Recognize the process of data cleaning and pre-processing for mining applications											
CO2	Apply various frequent pattern mining techniques on various data and understand the importance of Attribute Selection											
CO3	Distinguish problems related to classification or clustering											
CO4	Evaluate accuracy of various algorithms and select appropriate method for their respective data.											
Mapping of CO and PO:												
COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	1	0	1	2	0	0	0	0	3	0	0	1
CO2	2	1	1	2	1	0	0	0	0	0	2	1
CO3	2	1	2	2	0	1	0	0	2	0	0	1
CO4	2	1	1	3	1	0	0	0	0	0	1	1