



GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering

Subject Code: 3163209

DATA MINING & BUSINESS INTELLIGENCE

6th SEMESTER

Type of course: Undergraduate (Elective)

Prerequisite: NA

Rationale: NA

Teaching and Examination Scheme:

Teaching Scheme			Credits C	Examination Marks				Total Marks
L	T	P		Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Content:

Sr. No.	Content	Total Hrs.	% Weightage
1	Overview and concepts Data Warehousing (DW) and Business Intelligence (BI): Why reporting and Analysing data, Raw data to valuable information- Lifecycle of Data - What is Business Intelligence - BI and DW in today's perspective - What is data warehousing - The building Blocks: Defining Features - Data warehouses and data marts - Overview of the components - Metadata in the data warehouse - Need for data warehousing - Basic elements of data warehousing - trends in data warehousing, Relation between BI and DW - OLAP (Online analytical processing) definitions - Difference between OLAP and OLTP - Dimensional analysis	6	15
2	Introduction to data mining (DM): Motivation for Data Mining - Data Mining-Definition and Functionalities – Classification of DM Systems - DM task primitives - Integration of a Data Mining system with a Database or a Data Warehouse - Issues in DM – KDD Process	3	08
3	Data Pre-processing: Why to pre-process data? - Data cleaning: Missing Values, Noisy Data - Data Integration and transformation - Data Reduction: Data cube aggregation, Dimensionality reduction - Data Compression - Numerosity Reduction - Data Mining Primitives - Languages and System Architectures: Task relevant data - Kind of Knowledge to be mined - Discretization and Concept Hierarchy.	6	15
4	Concept Description and Association Rule Mining: What is concept description? - Data Generalization and summarization-based characterization - Attribute relevance - class comparisons Association Rule Mining: Market basket analysis - basic concepts - Finding frequent item sets: Apriori algorithm - generating rules – Improved Apriori algorithm – Incremental ARM – Associative Classification – Rule Mining	8	20



GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering

Subject Code: 3163209

5	<p>Classification and Prediction: What is classification and prediction? – Issues regarding Classification and prediction: Classification methods: Decision tree, Bayesian Classification, Rule based, CART, Neural Network Prediction methods: Linear and nonlinear regression, Logistic Regression Introduction of tools such as DB Miner / WEKA / DTREG DM Tools</p>	8	20
6	<p>Data Mining for Business Intelligence Applications: Data mining for business Applications like Fraud Detection, Clickstream, Market Segmentation, Retailing, Telecommunications, Banking & Finance and CRM etc</p>	3	07
7	<p>Advance topics: Introduction and basic concepts of - Clustering, Spatial mining, Spatial mining, Temporal Mining, Web mining, Text mining.</p> <p>Big Data: Introduction to big data: distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. Introduction to Hadoop architecture: Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands , Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering – Monitoring & Maintenance.</p>	8	15

Suggested Specification table with Marks (Theory):

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
10	20	15	15	5	5

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom’s Taxonomy)

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Reference Books:

1. J. Han, M. Kamber, “Data Mining Concepts and Techniques”, Morgan Kaufmann
2. M. Kantardzic, “Data mining: Concepts, models, methods and algorithms, John Wiley & Sons Inc.
3. Paulraj Ponnian, “Data Warehousing Fundamentals”, John Willey.
4. M. Dunham, “Data Mining: Introductory and Advanced Topics”, Pearson Education.
5. G. Shmueli, N.R. Patel, P.C. Bruce, “Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner”, Wiley India
6. Ning Tan, Vipin Kumar, Michael Steinbach Pang, “Introduction to Data Mining”, Pearson Education



GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering

Subject Code: 3163209

Course Outcome: After learning the course the students will be able

Sr. No.	CO Statement	Marks % Weightage
CO-1	Understand why the data warehousing is important in addition to database systems.	10
CO-2	Perform the preprocessing of data and apply mining techniques on it.	20
CO-3	Identify the association rules, classification, and clusters in large data sets.	25
CO-4	Solve real world problems in business and scientific information using data mining.	15
CO-5	Use data analysis tools for scientific applications.	15
CO-6	Implement various supervised machine learning algorithms.	15

List of Experiments:

Laboratory work will be based on the above syllabus with minimum 10 experiments to be incorporated.