

GUJARAT TECHNOLOGICAL UNIVERSITY

DATA MINING AND BUSINESS INTELLIGENCE

SUBJECT CODE: 2170715

B.E. 7th SEMESTER

Type of course: 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)	
				PA	ALA	ESE	OEP			
3	0	2	5	70	20	10	20	10	20	150

Content:

1.	Overview and concepts Data Warehousing and Business Intelligence	05 Hours	12%
	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.		
2.	The Architecture of BI and DW	07 Hours	16%
	BI and DW architectures and its types - Relation between BI and DW - OLAP (Online analytical processing) definitions - Difference between OLAP and OLTP - Dimensional analysis - What are cubes? Drill-down and roll-up - slice and dice or rotation - OLAP models - ROLAP versus MOLAP - defining schemas: Stars, snowflakes and fact constellations		
3.	Introduction to data mining (DM)	04 Hours	08%
	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		
4.	Data Pre-processing	07 Hours	16%
	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.		
5.	Concept Description and Association Rule Mining	07 Hours	16%
	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			
6.	Classification and Prediction	07 Hours	16%
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			
7.	Data Mining for Business Intelligence Applications	04 Hours	08%
Data mining for business Applications like Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc., Data Analytics Life Cycle: Introduction to Big data Business Analytics - State of the practice in analytics role of data scientists Key roles for successful analytic project - Main phases of life cycle - Developing core deliverables for stakeholders.			
8.	Advance topics	04 Hours	08%
Introduction and basic concepts of following topics. Clustering, Spatial mining, web mining, text mining, 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.			

Suggested Specification table with Marks (Theory):

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level

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.

Course Outcome:

After learning the course the students should be able to:

1. Students will be able to use mining tool.
2. Students are able to perform various data warehouse related exercise.

List of Experiments:

Assignments based on above course content will be given to the students at the end of each chapter. Each assignment contains minimum 5 questions.

Quizzes and Surprise tests will be conducted for testing the knowledge of students for particular topic.

Design based Problems (DP)/Open Ended Problem:

ACTIVE LEARNING ASSIGNMENTS: Preparation of power-point slides, which include videos, animations, pictures, graphics for better understanding theory and practical work – The faculty will allocate chapters/ parts of chapters to groups of students so that the entire syllabus to be covered. The power-point slides should be put up on the web-site of the College/ Institute, along with the names of the students of the group, the name of the faculty, Department and College on the first slide. The best three works should submit to GTU.