GUJARAT TECHNOLOGICAL UNIVERSITY

INFORMATION TECHNOLOGY (23)

DATA MINING

SUBJECT CODE: 2712309

SEMESTER: I

Type of course: Major Elective I

Prerequisite: Database fundamentals . Basic mathematical calculation

Rationale: Data mining is widely used in different diverted areas such as Financial Data Analysis, Retail Industry, Telecommunication Industry, Biological Data Analysis, Other Scientific Applications ,Intrusion Detection. So it is important to choose appropriate data mining system. This course enables to understand all data mining concepts so that proper data analysis can be performed.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks						Total Marks
L	T	P	C	Theory Marks		Practical Marks				
				ESE (E)	PA (M)	PA (V)		PA (I)		
						ESE	OEP	PA	RP	
3	2	2	5	70	30	20	10	20	0	150

Content:

Sr. No.	Topics	Teaching Hrs.	Module Weightage
1	**Overview and concepts Data Warehousing and Business Intelligence** 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	8	13.33
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	6	10
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	8	13.33
4	**Concept Description and Association Rule Mining** What is concept description? - Data Generalization and summarization-based characterization - Attribute relevance - class comparisons	8	13.33

	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		
5	**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	10	16.67
6	**Data Mining for Business Intelligence Applications** Data mining for business Applications like Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc	6	10
7	**Advance topics** Introduction and basic concepts of following topics. Multirelational Data Mining, Clustering, Spatial mining, web mining, text mining, Ensemble Classifier (Multiple Classifier, Bagging, Boosting, Stacking), Incremental learning	4	6.67

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:

1) Develop a supermarket application to implement defining subject area, design of fact dimension table, data mart. Include sales, order, inventory fact tables.
2) Develop an application to implement OLAP, roll up, drill down, slice and dice operation.
3) Example transaction with 4 items and 5 transactions.

Transaction_Id	milk	butter	bread	oats
1	1	0	0	1
2	1	1	1	0
3	1	0	1	0
4	1	1	1	1
5	0	1	1	1

Develop an application to extract association mining rule.

4) Develop an application for implementing Naïve Bayes classifier.
5) Implement Decision tree algorithm.

6) Implement K-means algorithm.
7) Implement Apriori algorithm.
8) Use WEKA tool and show how classification and clustering can be done.
9) Use DTREG DM tool and show how prediction can be made.
10) Use DB Miner and show how data mining paradigm can be applied.
11) Various doses of drugs were injected in 3 animals and change in blood pressure was observed. Take suitable data. Apply nonlinear regression and implement the necessary methods.

Major Equipments: Desktop,Laptop

Open Ended Problems:

1) How data mining techniques can be refined for text mining? For example Automatic processing of messages, emails, etc. Find a suitable technique for text mining to aid in the automatic classification of texts. For example, it is possible to "filter" out automatically most undesirable "junk email" based on certain terms or words that are not likely to appear in legitimate messages, but instead find out undesirable electronic mail and can automatically be discarded. Such automatic systems for classifying electronic messages can also be useful in applications where messages need to be automatically forwarded to the most appropriate department for example, email messages with complaints or petitions to a municipal authority are automatically routed to the appropriate departments.
2) How an automatic assessment system can be defined for short answers and essay questions? Which data mining technique is most appropriate?

List of Open Source Software/learning website: Weka