

GUJARAT TECHNOLOGICAL UNIVERSITY

COMPUTER ENGINEERING (SOFTWARE ENGINEERING) (02)

DATA MINING AND DATA WAREHOUSING

SUBJECT CODE: 2720209

SEMESTER: II

Type of course: Core

Prerequisite: DBMS

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)			
					ESE	OEP	PA	RP		
3	2#	2	5	70	30	20	10	10	10	150

Learning Objectives:

Student will be enabled to learn basics of Data Warehouse fundamentals & Information Search techniques using data mining and will be able to differentiate Database Management (DBMS) and large size data warehousing methods and their needs in business systems. Student will also learn various Data Mining search techniques and need of Data mart over DBMS, database and Data Warehouse. Student will also learn various OLAP technology will make student capable of handling huge information. Will be enabled to learn Data Mining tools such as XML Miner and Weka and Data Mining algorithms.

Content:

Sr. No.	Content	Total Hrs	% Weightage
1	Introduction to Data Mining Importance of Data Mining, Data Mining Functionalities, Classification of Data mining systems, Data mining Architecture, Major Issues in Data Mining, Applications of Data Mining, Social impacts of data mining.	8	20
2	Data Pre-processing & Data Mining primitives Data Pre-processing, Data cleaning, Data Integration and Transformation, Data reduction, Discretization and Concept Hierarchy Generation. Data Mining primitives, Languages and System Architectures, Concept Description: characterization and Comparison, Analytical Characterization, Mining Class Comparison.	12	20
3	Association Rules & Mining Association Rule Mining, Mining of Single dimensional Boolean association rules, Multilevel association rules and Multidimensional association rules, Correlation analysis, Constraint based association Mining.	6	10
4	Classification and Predication:	8	20

	Basic issues regarding classification and predication, Classification by Decision Tree, Bayesian classification, classification by back propagation, Associative classification, Prediction, Classifier accuracy.		
5	Cluster Analysis Cluster Analysis, basic issues, clustering using partitioning methods, Hierarchical methods, Density based methods, Grid based methods and model based methods, Algorithms for outlier analysis.	6	15
6	Mining complex Types of data: Multidimensional analysis and descriptive mining of complex data objects, Introduction to spatial mining, multimedia mining, temporal mining, text mining and web mining with related algorithms.	10	15

Reference Books:

1. Data Mining concepts and Techniques by Jiawei Han, Micheline Kamber –Elsevier.
2. Data Mining by Arun K. Pujari – University Press.
3. Mordern Data Warehousing, Data Mining and Visualization by George M.Marakas –Pearson.
4. Data Mining by Vikram Puri And P.RadhaKrishana –Oxfrod Press.
5. Data Warehousing by Reema Theraja –Oxford Press

Course Outcome:

After learning the course the students should be able to:

Student will be enabled to learn basics of Data Warehous fundamentals and will be able to differentiate Database Management (DBMS) and large size datawarehousing methods and thir needs in usiness systems. Student will also learn various Datamining search teqchniques and need of Data mart over DBMS , database and Datawarehouse. Student will also learn various OLAP technology will make stduent capabale of handing huge information. Will be enabled to learn Datamining tools such as XML Minier and Weka and Data Minining algoorthims

List of Experiments and Open Ended Problems:

- (1) Software Forensic tools such as XML Miner & WeKA
- (2) Implemenation of data mining alg such as Clustering, Classification and Association Rules
- (3) OLAP in standard SQL Server
- (4) MS Excell & Datamining plugin

Major Equipment:

- (1) WeKA
- (2) XMLMinier

Review Presentation (RP): The concerned faculty member shall provide the list of peer reviewed Journals and Tier-I and Tier-II Conferences relating to the subject (or relating to the area of thesis for seminar) to the students in the beginning of the semester. The same list will be uploaded on GTU website during the first two weeks of the start of the semester. Every student or a group of students shall critically study 2 papers, integrate the details and make presentation in the last two weeks of the semester. The GTU marks entry portal will allow entry of marks only after uploading of the best 3 presentations. A unique id number will be

generated only after uploading the presentations. Thereafter the entry of marks will be allowed. The best 3 presentations of each college will be uploaded on GTU website