GUJARAT TECHNOLOGICAL UNIVERSITY

SUBJECT NAME: Data Mining
SUBJECT CODE: 3715507
Semester I

Type of course: NA

Prerequisite:
1. Bigdata

Rationale: NA

Teaching and Examination Scheme:

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

L- Lectures; T- Tutorial/Teacher Guided Student Activity; P- Practical; C- Credit; ESE- End Semester Examination; PA- Progressive Assessment;

Content:

Sr. No.	**Content**	**Total Hrs**	**% Weightage**
1	Introduction to Data mining principles- Data mining and knowledge discovery, The need for data mining, Overview of Data warehousing and mining, Advantages and challenges, Data mining applications in various application areas, Datawarehousing, Datamarts and OLAP - Data warehouse architectures, Datawarehouse design, Steps in Datawarehousing (ETL), OLAP Vs OLTP, Data marts, design and performance considerations	10	20
2	Data mining Overview, Data mining process, Understanding different types of data and preprocessing considerations, Classification and Prediction - Naive bayes Classification, back propagation based classification based, tree based, Support Vector machines, Associative classification, Prediction - Regression Trees	10	25
3	Clustering and association Rule mining, k-means clustering, EM technique, Hierarchical Clustering, Density based methods, Grid based methods, Model based methods, Cluster Analysis and Outlier Analysis. Association Rule mining - Mining Frequent patterns, MIning Various associations and rules, Correlation Analysis, Apriori (Market basket Analysis), Constraint based association rule mining, Data mining Applications in Text mining, Stream mining and Fraud Detection.Working with weka, Exploring different data sets, Mining trends, Case studies	10	25

Reference Books:

1. Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications, Volume V , John Wang (ed) ,IGI Global ,ISBN:9781599049519
2. Practical Data Mining , Monte F. Hancock, Auerbach Publications, ISBN:9781439868362
3. Data Mining and Data Warehousing ,Bharat Bhushan Agarwal and Sumit Prakash Tayal, Laxmi Publications ,ISBN:9788131806586
4. Handbook of Statistical Analysis and Data Mining Applications, by Robert Nisbet, John Elder and Gary Miner ,Academic Press ISBN 9780123747655

Course Outcome:

After learning the course the students should be able to:

1. Because the development of the data mining systems requires highly skilled programmers / problem –solvers.
2. Data mining is one of the highest paid professions in Information Technology
3. The requirements of this profession include ability to Conceptualize the problem at hand, ability to investigate/ research the problem
4. Ability to select methods and techniques appropriate for the task, and the ability to develop the methods and tools for the given task
5. This is turn requires understanding of data mining problems and techniques excellent programming skills and virtually permanent self-development
6. In the light of this, the course objectives are organized to help student understand the fundamental process, concepts and techniques of data mining and develop an appreciation for the inherent complexity of the data mining task advance relevant programming skills and research skills through the investigation of data mining literature