GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Bachelor of Vocation (B.Voc), 2nd Semester
Branch: Software Development
Subject Name: Concepts of Data Mining
Subject Code: 1120202

With effective from academic year 2018-19

Type of course: Compulsory

Prerequisite: NA

Rationale: Data Mining studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. It is currently regarded as the key element of a more general process called Knowledge Discovery that deals with extracting useful knowledge from raw data. The knowledge discovery process includes data selection, cleaning, coding, using different statistical and machine learning techniques, and visualization of the generated structures.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks						Total Marks
				Theory Marks			Practical Marks			
L	T	P	C	ESE (E)	PA (M)		PA(V)		PA (I)	
					PA	ALA	ESE	OEP		
03	00	00	03	50	00	00	00	00	00	50

L- Lectures; P- Practical; OJT- On Job Training; C- Credit; ESE- End Semester Examination; PA- Progressive Assessment

Contents:

Sr. No.	Practical / Hands on Exercise	Hrs.	Module Weightage
1	**Introduction to Data warehousing:-** Needs for developing data Warehouse, Datawarehouse systems and its Components, Design of Data Warehouse, Dimension and Measures, Data Marts:-Dependent Data Marts, Independents Data Marts & Distributed Data Marts, Conceptual Modeling of Data Warehouses: -Star Schema, Snow flake Schema, Fact Constellations, Multidimensional Data Model & Aggregates.	8	20
2	**OLAP:-** Characteristics of OLAP System, Motivation for using OLAP, Multidimensional View and Data Cube, Data Cube Implementations, Data Cube Operations, Guidelines for OLAP Implementation, Difference between OLAP & OLTP, OLAP Servers: -ROLAP, MOLAP, HOLAP Queries	6	10
3	**Introduction to Data Mining:-** Knowledge Discovery, Data Mining Functionalities, Data Mining System categorization and its Issues. Data Processing:-Data Cleaning, Data Integration and Transformation. Data Reduction, Data Mining Statistics. Guidelines for Successful Data Mining.	10	25
4	**Association Rule Mining:-** Introduction, Basic, The Task and a Naïve Algorithm, Apriori Algorithms, Improving the efficiency of the Apriori Algorithm, Apriori - Tid, Direct Hasing and Pruning (DHP), Dynamic Item set Counting (DIC), Mining Frequent Patterns without Candidate Generation (FP-Growth), Performance Evaluation of Algorithms	8	20
5	**Classification:-** Introduction, Decision Tree, The Tree Induction Algorithm, Split	10	25

	Algorithms Basedon Information Theory, Split Algorithm Based on the Gini Index, Over fitting and Pruning, Decision Trees Rules, Naïve Bayes Method. Cluster Analysis: -Introduction, Desired Features of Cluster Analysis, Types of Cluster Analysis Methods: -Partitioned Methods, Hierarchical Methods, Density-Based Methods, Dealing with Large Databases. Quality and Validity of Cluster Analysis Methods.		
	Total	**42**	

Reference Books:

1. Data Mining Concepts and Techniques, J. Han, M. Kamber, Morgan Kaufmann
2. Data Mining and Warehousing, Ikvinderpal Singh, Khanna Publishing House

Course Outcomes:

Sr. No.	CO Statement	Marks % Weightage
CO-1	Understand the need for developing Data Warehouse	20
CO-2	Study various OLAP Operations	10
CO-3	Understand the concept of Data Mining and Data Preprocessing	25
CO-4	Apply association rule mining on different data set	20
CO-5	Apply various classification and clustering algorithms on different data sets	25

Laboratory work: NA

List of Open Source Software/learning website:

Students must refer to following sites to enhance their learning ability.
1) www.coursera.org
2) NPTEL tutorials