Syllabus for Bachelor of Vocation (D.Voc), 6ᵗʰ Semester
Branch: Software Development
Subject Name: Concepts of Data Mining
Subject Code: 1260202

With effective from academic year 2019-20

Teaching Scheme			Credits	Examination Marks				Total Marks
				Theory Marks		Practical Marks		
L	T	P	C	ESE (E)	PA (M)	ESE (V)	PA (I)	
3	-	0	3	50	-	-	-	50

Course Content:

Unit No.	Content	Hrs
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
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
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
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
5	**Classification:-** Introduction, Decision Tree, The Tree Induction Algorithm, Split 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.	10
	Total Hours:	42

Suggested Specification table with Marks (Theory):

> **Distribution of Theory Marks**

GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Bachelor of Vocation (D.Voc), 6th Semester
Branch: Software Development
Subject Name: Concepts of Data Mining
Subject Code: 1260202

With effective from academic year 2019-20

R Level	U Level	A Level	N Level	E Level
5	25	25	10	5

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate and above Levels (Bloom's Taxonomy)

Reference Books:
1. Data Mining Concepts and Techniques, J. Han, M. Kamber, Morgan Kaufmann
2. Data Minin