Parul University - Faculty of Engineering and Technology

Department of Computer Science & Engineering

SYLLABUS FOR 6th Sem BTech PROGRAMME

Data Mining and Business Intelligence (203105453)

Type of Course: BTech

Prerequisite: Database Management System

Rationale: This course helps the students to understand the overall architecture of a data warehouse and methods for data gathering and data pre-processing using OLAP tools. The different data mining models and techniques will be discussed in this course. Data mining and data warehousing applications in bioinformatics will also be explored.

Teaching and Examination Scheme:

Teaching Scheme			Credit	Examination Scheme					Total
				External		Internal			
Lect Hrs/ Week	Tut Hrs/ Week	Lab Hrs/ Week		T	P	T	CE	P	
3	0	0	3	60	-	20	20	-	100

Lect - Lecture, **Tut** - Tutorial, **Lab** - Lab, **T** - Theory, **P** - Practical, **CE** - CE, **T** - Theory, **P** - Practical

Contents:

Sr.	Topic	Weightage	Teaching Hrs.
1	**Introduction to data mining (DM)**: Importance of Data Mining, Data Mining-Definition and Functionalities, Classification of Data mining systems, Data mining Architecture, KDD, DM task primitives, Major Issues in Data Mining, Data mining Technologies, Applications of Data Mining.	10%	6
2	**Overview and concepts Data Warehousing and Business Intelligence**: Reporting and Analysis of data, Raw data to valuable information-Lifecycle of Data, Business Intelligence, Relation between BI, DM and DW.	5%	2
3	**Data Warehousing and Online Analytical Processing**: Introduction, Motivation for Data Warehouse, Multitier Architecture, ETL process, Multidimensional Data Model, Schemas for Multidimensional Data Models, Data Warehouse Models - Enterprise Warehouse, Data Mart, and Virtual Warehouse, The Role of Concept Hierarchies, OLAP Server, Typical OLAP Operations, Dimensional analysis.	20%	6
4	**Data Pre-processing**: Data Quality, Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization, Reasons of Missing Values, &Noisy Data, Tasks in Data Preprocessing, Cleaning, Integration, Reduction, Transformation and Discretization, Concept Hierarchy Generation.	15%	6

5	**Mining Frequent Patterns, Associations, and Correlations**: Market Basket Analysis, Association Rules, Multidimensional & Multilevel association rules, Frequent Item sets, Frequent Itemset Mining Methods Apriori algorithm, Improved Apriori algorithm, Pattern-Growth Approach: FP-growth, Generating Association Rules from Frequent Item sets, Interesting Pattern Evaluation Methods, Correlation analysis.	15%	6
6	**Classification**: Classification vs. prediction, Supervised learning, Approach to Classification: Decision Tree Induction, Bayes Classification Methods Rule-Based Classification, Model Evaluation and Selection, Advanced Classification Methods.	15%	5
7	**Clustering**: Unsupervised learning, Cluster Analysis: Partitioning Methods, Hierarchical Methods, Density-Based Methods, Evaluation of Clustering, Outlier Detection	10%	7
8	**Applications**: Introduction to spatial mining, multimedia mining, temporal mining, text mining and web mining with related algorithms.	10%	4

***Continuous Evaluation:**

It consists of Assignments/Seminars/Presentations/Quizzes/Surprise Tests (Summative/MCQ) etc.

Reference Books:

1. Data Mining concepts and Techniques (TextBook)
   Jiawei Han, Micheline Kamber; Elsevier
2. Data Mining Techniques (TextBook)
   Arun K. Pujari; Universities Press
3. Modern Data Warehousing, Data Mining and Visualization
   George M. Marakas; Pearson
4. Data Warehousing
   Reema Theraja; Oxford Press

Course Outcome:

After Learning the course the students shall be able to:

After undergoing the course, Students will be able to understand
1. Design a data mart or data warehouse for any organization
2. Develop skills to write queries using DMQL
3. Extract knowledge using data mining techniques
4. Adapt to new data mining tools.
5. Apply the techniques of clustering, classification, association finding, feature selection and visualization
to real world data

Parul University - Faculty of Engineering and Technology

Department of Computer Science & Engineering

SYLLABUS FOR 5th Sem BTech PROGRAMME

Data Mining and Business Intelligence Laboratory (203105454)

Type of Course: BTech

Prerequisite: Database Management System

Rationale: This course helps the students to understand the overall architecture of a data warehouse and methods for data gathering and data pre-processing using OLAP tools. The different data mining models and techniques will be discussed in this course. Data mining and data warehousing applications in bioinformatics will also be explored.

Teaching and Examination Scheme:

Teaching Scheme			Credit	Examination Scheme					Total
				External		Internal			
Lect Hrs/ Week	Tut Hrs/ Week	Lab Hrs/ Week		T	P	T	CE	P	
0	0	2	1	-	30	-	-	20	50

Lect - Lecture, **Tut** - Tutorial, **Lab** - Lab, **T** - Theory, **P** - Practical, **CE** - CE, **T** - Theory, **P** - Practical

Course Outcome:

After Learning the course the students shall be able to:

After undergoing the course, Students will be able to understand
1. Design a data mart or data warehouse for any organization
2. Develop skills to write queries using DMQL
3. Extract knowledge using data mining techniques
4. Adapt to new data mining tools.
5. Apply the techniques of clustering, classification, association finding, feature selection and visualization
to real world data

List of Practical:

1. **Practical-1**

   Design and Create cube by identifying measures and dimensions for Star Schema, Snowflake schema and fact Constellation Schema.

2. **Practical-2**

   Make an OLAP cube and perform Roll Up and Drill Down operations on it. Show the Apex and Base cuboid for the same. Draw Star-net query model for the cube.

3. **Practical-3**

   Create calculated member using arithmetic operators and member property of dimension member.

4. **Practical-4**

   Design and Create cube by identifying measures and dimensions for Design storage using storage mode MOLAP, ROLAP and HOLAP.

5. **Practical-5**

Perform Pre-processing on a dataset. Apply various Filters and discuss the effect of each filter applied.
a. Handle Missing Values
b. Handle Infrequent Nominal Values
c. Derive an attribute from the existing attribute
d. Sampling
e. Discretization
1)Use Weka Tool 2) Use XL Miner Tool

6. **Practical-6**

Perform different binning techniques to smooth out the noise in the dataset. Make sure that the user should have the choice to apply all the possible techniques. Show the output of different bins. Use histogram to partition the dataset into groups.

7. **Practical-7**

Perform regression on the data set using R programming.

8. **Practical-8**

Perform Association rule mining using WEKA tool.

9. **Practical-9**

Perform classification with WEKA tool.
a. using Decision Tree Classifier
b. using Naïve Bayes Classifier
c. using Multilayer Perceptron

10. **Practical-10**

Perform Clustering using WEKA tool