

Uka Tarsadia University



M.Sc. (C.A.)

Data Mining Models and Methods (040020304)

3rd Semester

EFFECTIVE FROM JUNE-2012

Uka Tarsadia University
M.Sc.(C. A.) (3rd Semester) Syllabus, June 2012

Prerequisites: RDBMS and basic of probability

Aim and Objective: To understand the need for data mining and learn different data mining models and techniques to solve several business problems.

Paper No.: 040020304

Subject: Data Mining Models and Methods

Total: 48 hrs.

[Lecture: 4, Tutorial:0, Practical: 0]

- 1. Data Mining: Introduction [07 hrs.]**
 - 1.1. An Overview; What is Data Mining; Data Mining – on What Kind of Data
 - 1.2. Data Mining Functionalities – What Kind of Patterns Can be Mined; Concept/Class Description: Characterization & Discrimination; Mining Frequent Patterns, Associations, and Correlations; Classification & Prediction; Cluster Analysis; Outlier Analysis
 - 1.3. Interesting Patterns
 - 1.4. Classification of Data Mining Systems
 - 1.5. Data Mining Task Primitives
 - 1.6. Integration of a Data Mining System with a Database or Data Warehouse System
 - 1.7. Major Issues in Data Mining

- 2. Pre-processing [08 hrs.]**
 - 2.1. The need for Pre-processing, Descriptive Data Summarization
 - 2.2. Data Cleaning
 - 2.3. Data Integration & Transformation
 - 2.4. Data Cube Aggregation; Attribute Subset Selection
 - 2.5. Basic of Dimensionality Reduction
 - 2.6. Numerosity Reduction: Regression & Log-linear Models, Histograms, Clustering, Sampling
 - 2.7. Data Discretization & Concept Hierarchy Generation

- 3. Introduction to Data Warehousing, OLAP and Multi-dimensional data model [10 hrs.]**
 - 3.1. Data Warehouse

- 3.2. Differences between Operational Database Systems and Data Warehouses
- 3.3. OLTP and OLAP and their Differences
- 3.4. Multi-dimensional Data Model: Star, Snowflakes, and Fact Constellations Schemas for Multi-dimensional Databases
- 3.5. Concept Hierarchies
- 3.6. OLAP Operations in Multi-dimensional Data Model: Roll-up, Drill-down, Slice & Dice, Pivot (Rotate)
- 3.7. The Design and construction of Data Warehouse
- 3.8. Three-Tier Data Warehouse Architecture
- 3.9. Metadata Repository
- 3.10. Type of OLAP Servers, ROLAP versus MOLAP versus HOLAP
- 3.11. Indexing OLAP Data, Efficient Processing of OLAP Queries

4. Mining Frequent Patterns, Associations, and Correlations [06 hrs.]

- 4.1. Basic Concepts: Market Basket Analysis; Frequent Itemsets, Closed Itemsets, and Association Rules; Frequent Pattern Mining
- 4.2. Apriori Algorithm: Finding Frequent Itemsets Using Candidate Generation; Generating Association Rules from Frequent Itemsets; Improving the Efficiency of Apriori
- 4.3. Association Mining to Correlation Analysis

5. Classification & Prediction [10 hrs.]

- 5.1. Introduction
- 5.2. Issues regarding classification and Prediction
- 5.3. Classification by Decision Tree Induction- Attribute Selection Measures; Tree Pruning; Scalability and Decision Tree Induction
- 5.4. Bayesian Classification- Bayes' Theorem, Naïve Bayesian Classification; Bayesian Belief Networks
- 5.5. Rule-based Classification
- 5.6. Prediction

6. Cluster Analysis [07 hrs.]

- 6.1. Introduction to Cluster Analysis
- 6.2. Types of Data in Cluster Analysis
- 6.3. Categorization of major Clustering Methods
- 6.4. Partitioning Methods; K-Means Method
- 6.5. Outlier Analysis

MODES OF TRANSCATION (i.e. Delivery)

Various methods of teaching could be employed depending on the objectives of the content taught.

- Lecture method** along with various appropriate audio-visual aids for all the units.
- Apart of lecture method, assignment activity** can be designed for unit 5 and 6.

Teachers Activities/Practicum:

The following activities may be carried out by the teachers.

1. Demonstration of Weka tool.
2. Demonstration of clustering on a relevant dataset.
3. Demonstration of classification on a relevant dataset.

Student Activities/Practicum:

The following activities may be carried out by the students.

1. Study of BIRCH Method for Cluster Analysis and its application.
2. Study the importance of text mining.

[Weightage to be given in Continuous Internal Evaluation]

Text Books:

1. Han, J. and Kamber, M. Data Mining: Concepts & Techniques, Morgan Kaufmann Publishers

Reference Books:

1. Pudi, V. and Radhakrishnan, P. Data Mining, Oxford University Press
2. Larose, D., Data Mining Methods & Models, Wiley-India
3. Inmon, W. Building the Data Warehouse, Wiley Dreamtech India Pvt. Ltd.
4. Soumendra, M. Data Warehousing: Design, Development and Best Practices, Tata McGraw Hill
5. Berry, M. and Linoff, G., Data Mining Techniques, Wiley-India