Data Mining
[CT725]

Lecture : 3 **Year : IV**
Tutorial : 1 **Part : I**
Practical : 1.5

Course Objective:

This course introduces the fundamental principles, algorithms and applications of intelligent data processing and analysis. It will provide an in depth understanding of various concepts and popular techniques used in the field of data mining.

1. Introduction (2 hours)

   1.1. Data Mining Origin
   1.2. Data Mining & Data Warehousing basics

2. Data Preprocessing (6 hours)

   2.1. Data Types and Attributes
   2.2. Data Pre-processing
   2.3. OLAP & Multidimensional Data Analysis
   2.4. Various Similarity Measures

3. Classification (12 hours)

   3.1. Basics and Algorithms
   3.2. Decision Tree Classifier
   3.3. Rule Based Classifier
   3.4. Nearest Neighbor Classifier

3.5. Bayesian Classifier
3.6. Artificial Neural Network Classifier
3.7. Issues : Overfitting, Validation, Model Comparison

4. Association Analysis (10 hours)

4.1. Basics and Algorithms
4.2. Frequent Itemset Pattern & *Apriori* Principle
4.3. FP-Growth, FP-Tree
4.4. Handling Categorical Attributes
4.5. Sequential, Subgraph, and Infrequent Patterns

5. Cluster Analysis (9 hours)

5.1. Basics and Algorithms
5.2. K-means Clustering
5.3. Hierarchical Clustering
5.4. DBSCAN Clustering
5.5. Issues : Evaluation, Scalability, Comparison

6. Anomaly / Fraud Detection (3 hours)

7. Advanced Applications (3 hours)

7.1. Mining Object and Multimedia

7.2. Web-mining

7.3. Time-series data mining

Practical:

Using either MATLAB or any other DataMining tools (such as WEKA), students should practice enough on real-world data intensive problems like IRIS or Wiki dataset.

References:

- Pang-NingTan, Michael Steinbach and Vipin Kumar, Introductionto Data Mining, 2005, Addison-Wesley.
- Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, 2nd Edition, 2006, Morgan Kaufmann.

Evaluation Scheme:

The question will cover all the chapters of the syllabus. The evaluation scheme will be as indicated in the table below:

Chapters	Hours	Marks Distribution*
1	2	4
2	6	10
3	12	20
4	10	18
5	9	16
6	3	6
7	3	6
Total	45	80

*There may be minor variation in marks distribution.