Objective: To introduce the fundamental principles, algorithms and applications of intelligent data processing and analysis and to provide an in depth understanding of various concepts and popular techniques used in the field of data mining. The subject code of Data Mining which is set as elective I by IOE is CT725. **Data mining can be chosen by Computer (BCT) and Electronics & Communication (BEX) students** for elective in Fourth Year – First Part. It will have a total of 80 marks of final exam and 20 marks as internal marking. The chapter along with their credit hours and marking scheme are as below.

1. **Introduction (2 hours)**
    1. Data Mining Origin
    2. Data Mining & Data Warehousing basics
2. **Data Pre-Processing (6 hours )**
    1. Data Types and Attributes
    2. Data Pre-processing
    3. OLAP & Multidimensional Data Analysis
    4. Various Similarity Measures
3. **Classification (12 hours)**
    1. Basics and Algorithms
    2. Decision Tree Classifier [humanoriented]
    3. Rule Based Classifier
    4. Nearest Neighbor Classifier
    5. Bayesian Classifier
    6. Artificial Neural Network Classifier
    7. Issues : Overfitting, Validation, Model Comparison
4. **Association Analysis (10 hours)**
    1. Basics and Algorithms
    2. Frequent Itemset Pattern & *Apriori* Principle
    3. FP-Growth, FP-Tree
    4. Handling Categorical Attributes
    5. Sequential, Subgraph, and Infrequent Patterns
5. **Cluster Analysis (9 hours)**
    1. Basics and Algorithms
    2. K-means Clustering [Wikipedia]
    3. Hierarchical Clustering
    4. DBSCAN Clustering
    5. Issues : Evaluation, Scalability, Comparison
6. **Anomaly / Fraud Detection (3 hours)**
7. **Advanced Applications (3 hours)**
    1. Mining Object and Multimedia
    2. Web-mining
    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.

**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.

**Reference Materials:**


- 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.