

**NIRMA UNIVERSITY**  
**Institute of Technology**  
**B. Tech. Computer Science and Engineering**  
**Semester – VI**  
**Department Elective-III**

| L | T | P | C |
|---|---|---|---|
| 2 | 0 | 2 | 3 |

|                     |             |
|---------------------|-------------|
| <b>Course Code</b>  | 2CSDE71     |
| <b>Course Title</b> | Data Mining |

**Course Outcomes:**

At the end of the course, students will be able to –

1. interpret data types and infer various data pre-processing techniques for the mining process
2. apply basic principles and algorithms used in practical data mining
3. Choose and evaluate data mining algorithms for various real life problems.

**Syllabus:**

**Teaching  
Hours:30**  
02

**Unit I**

**Introduction:** Motivation and importance, different kinds of data, data mining functionalities, classification of data mining systems, major issues in data mining

**Unit II**

**Data Pre-processing:** Data summarization, data cleaning, data integration and transformation, data reduction, data discretization and concept hierarchy generation, feature extraction , feature transformation, feature selection, introduction to Dimensionality Reduction, CUR decomposition

**Unit III**

**Mining Frequent Patterns, Associations and Correlations:** Basic concept, efficient and scalable frequent item-set mining methods, mining various kind of association rules, from association mining to correlation analysis, Advanced Association Rule Techniques, Measuring the Quality of Rules.

**Unit IV**

**Classification and Prediction:** Classification vs. prediction, Issues regarding classification and prediction, Statistical-Based Algorithms, Distance-Based Algorithms, Decision Tree-Based Algorithms, Neural Network-Based Algorithms, Rule-Based Algorithms, Combining Techniques, accuracy and error measures, evaluation of the accuracy of a classifier or predictor.

**Unit V**

**Advanced Cluster Analysis:** Types of data in cluster analysis, overview of major clustering methods, probabilistic model based clustering, clustering high dimensional data, clustering Graph and Network data.



## Unit VI

03

**Case studies:** Applications of Distributed and parallel Data Mining. Advanced Techniques: Web Mining, Spatial Database Mining, Temporal Mining, And Multimedia Mining.

### Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

### Laboratory Work:

Laboratory work will be based on the above syllabus with minimum 10 experiments to be incorporated.

### Suggested Readings<sup>^</sup>:

1. Jiawei Han and Micheline Kamber, Data mining: Concepts and Techniques, Morgan Kaufmann Publishers.
2. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann
3. Hand, Mannila, and Smyth., Principles of Data Mining, MIT Press
4. Berry and Linoff, Mastering Data Mining, Wiley
5. Delmater and Hancock, Data Mining Explained, Digital Press

L=Lecture, T=Tutorial, P=Practical, C=Credit

---

<sup>^</sup>this is not an exhaustive list