## CS702PC: DATA MINING (PC)

#### IV Year B.Tech. CSE I - Sem

# L T P C 2 0 0 2

#### **Pre-Requisites:**

- A course on "Database Management Systems"
- Knowledge of probability and statistics

## **Course Objectives:**

- It presents methods for mining frequent patterns, associations, and correlations.
- It then describes methods for data classification and prediction, and data-clustering approaches.
- It covers mining various types of data stores such as spatial, textual, multimedia, streams.

## **Course Outcomes:**

- Ability to understand the types of the data to be mined and present a general classification of tasks and primitives to integrate a data mining system.
- Apply preprocessing methods for any given raw data.
- Extract interesting patterns from large amounts of data.
- Discover the role played by data mining in various fields.
- Choose and employ suitable data mining algorithms to build analytical applications
- Evaluate the accuracy of supervised and unsupervised models and algorithms.

## UNIT - I

**Data Mining:** Data–Types of Data–, Data Mining Functionalities– Interestingness Patterns– Classification of Data Mining systems– Data mining Task primitives –Integration of Data mining system with a Data warehouse–Major issues in Data Mining–Data Preprocessing.

## UNIT - II

**Association Rule Mining:** Mining Frequent Patterns–Associations and correlations – Mining Methods– Mining Various kinds of Association Rules– Correlation Analysis– Constraint based Association mining. Graph Pattern Mining, SPM.

## UNIT - III

**Classification:** Classification and Prediction – Basic concepts–Decision tree induction–Bayesian classification, Rule–based classification, Lazy learner.

## UNIT - IV

**Clustering and Applications:** Cluster analysis–Types of Data in Cluster Analysis–Categorization of Major Clustering Methods– Partitioning Methods, Hierarchical Methods– Density–Based Methods, Grid–Based Methods, Outlier Analysis.

## UNIT - V

Advanced Concepts: Basic concepts in Mining data streams–Mining Time–series data—Mining sequence patterns in Transactional databases– Mining Object– Spatial– Multimedia–Text and Web data – Spatial Data mining– Multimedia Data mining–Text Mining– Mining the World Wide Web.

## **TEXT BOOKS:**

- 1. Data Mining Concepts and Techniques Jiawei Han & Micheline Kamber, 3<sup>rd</sup> Edition Elsevier.
- 2. Data Mining Introductory and Advanced topics Margaret H Dunham, PEA.

## **REFERENCE BOOK:**

1. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques (Second Edition), Morgan Kaufmann, 2005.