

## **CAP446:DATA WAREHOUSING AND DATA MINING**

**Course Outcomes:** Through this course students should be able to

CO1 :: Understand the various concepts of data warehousing like metadata, data mart, summary table, fact data and dimension data.

CO2 :: Analyze the various methods to extract knowledge using data mining techniques

CO3 :: Apply different data mining methodologies with information systems.

CO4 :: Evaluate the current trends in data mining such as web mining, spatial-temporal mining.

### **Unit I**

**Data warehousing and online analytical processing** : Basic concepts, Data warehouse modeling: data cube and OLAP, Data warehouse design and usage, Data warehouse implementation

### **Unit II**

**Introduction to data mining** : Basic concepts of data mining, Different types of data repositories, data mining functionalities, Concept of interesting patterns, Data mining tasks, Current trends, Major issues and ethics in data mining

### **Unit III**

**Data Preprocessing** : Data cleaning, Data integration and transformation, Data reduction, Discretization and concept hierarchy generation

### **Unit IV**

**Association and correlation analysis** : Basic concepts of frequent pattern and association rule, Frequent itemset generation with Apriori algorithm and FP Growth algorithm, Rule generation, Applications of association rules

### **Unit V**

**Clustering algorithms and cluster analysis** : Measures of similarity, K means partitioning method, k medoids method, CLARANS method, Agglomerative and divisive clustering hierarchical method, BIRCH method, Density based methods, Cluster evaluation, Outlier detection and analysis

### **Unit VI**

**Classification** : Basic concepts of binary classification, Bayes theorem and Naive Bayes classifier, Association based classification, Rule based classifiers, Nearest neighbour classifiers, Decision Trees, Random Forest, Model overfitting, Cross validation

### **Text Books:**

1. DATA MINING: CONCEPTS AND TECHNIQUES by JAWEI HAN, MICHELINE KAMBER AND JIAN PE, MORGAN KAUFMANN

### **References:**

1. INTRODUCTION TO DATA MINING by PANG-NING TAN , MICHAEL STEINBACH , VIPIN KUMAR, PEARSON