# DATA WAREHOUSING AND DATA MINING
## (Common to CSE & IT)

**Course Code: 22CT1111**

**L  T  P  C**
**3  0  0  3**

**Pre-requisites:** Database Management Systems
**COURSE OUTCOMES:**
At the end of the Course, the Student will be able to:
**CO1:** Explain different data types and data preprocessing techniques.(L2)
**CO2:** Apply data warehouse models for any given data sets. (L3)
**CO3:** Identify associations and correlations on any given data sets.(L4)
**CO4:** Apply classification techniques on any given data sets.(L3)
**CO5:** Identify clusters using clustering techniques on any given data sets.(L4)

**UNIT-I**                                                                                   **(12 LECTURES)**
**INTRODUCTION:**
Data Mining, Motivating Challenges, The origins of Data Mining, Data Mining Tasks
**DATA:**
Types of Data, Data Quality, Data Preprocessing: Aggregation, Sampling, Dimensionality Reduction, Feature Subset Selection, Feature Creation, Discretization and Binarization, Variable Transformation
**MEASURES OF SIMILARITY AND DISSIMILARITY:**
Basics, Similarity and Dissimilarity between simple attributes, Dissimilarities between data objects, similarities between data objects, Examples of proximity measures, Issues in proximity calculation, selecting the right proximity measure. (Text Book-2)
**Learning Outcomes:** At the end of the module the student will be able to
1. Summarize the basics of data mining (L2)
2. Apply various data preprocessing techniques on the given data set(L3)
3. Illustrate various measures of similarity and dissimilarity between data objects(L2)

**UNIT-II**                                                                                   **(8 LECTURES)**
**DATA WAREHOUSE AND OLAP TECHNOLOGY:**
Data Warehouse: Basic concepts, Data Warehouse Modeling: Data Cube and OLAP, Data Warehouse Implementation.
**Learning Outcomes:** At the end of the module the student will be able to
1. Describe the basic concepts of Data Warehousing (L2)
2. Model data cubes for the given data (L3)
3. Apply OLAP operations on data cubes (L3)

## UNIT-III                                                    (10 LECTURES)
**MINING FREQUENT PATTERNS, ASSOCIATION AND CORRELATIONS:**
Basic Concepts, Efficient and Scalable Frequent Item set Mining Methods. Mining various kinds of association rules.
**Learning Outcomes:** At the end of the module the udent will be able to
1. Describe the basic concepts of associations and correlations (L2)
2. Infer association rules using Candidate set generation and without candidate set generation algorithms (L2)
3. Apply various kinds of association rules (L3)


## UNIT-IV                                                     (10 LECTURES)
**CLASSIFICATION:**
BASIC CONCEPTS- Classification, Prediction, Issues regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification-Baye's Theorem, Naive Bayesian Classification, Rule-Based Classification-Using IF-THEN Rules for Classification, Rule Extraction from a Decision Tree, Classification by Back propagation.
**Learning Outcomes:** At the end of the module the student will be able to
1. Distinguish classification and prediction tasks. (L4)
2. Model a decision tree for any given dataset. (L3)
3. Infer rules using decision tree for any given dataset. (L2)
4. Apply Bayesian classification algorithm to classify any given dataset. (L3)
5. Apply backpropagation algorithm for any given neural network. (L3)


## UNIT-V                                                      (10 LECTURES)
**CLUSTER ANALYSIS:**
BASIC CONCEPTS- Cluster analysis, A Categorization of Major Clustering Methods, Partitioning Methods (k- Means, k-Medoids), Hierarchical Methods: Agglomerative Vs Divisive (BIRCH), Density Based Methods (DBSCAN) Grid- Based Methods (STING), Model-Based Clustering Methods (Expectation Maximization)
**EVALUATION OF CLUSTERING:**
Assessing Clustering Tendency, Determining the Number of Clusters, Measuring Clustering Quality, Extrinsic Methods, Intrinsic Methods
**Learning Outcomes:** At the end of the module the student will be able to
1. Describe the basics of clustering (L2)
2. Summarize different clustering methods (L2)
3. Apply different clustering algorithms on any given data sets (L3)
4. Make use of different cluster evaluation techniques to assess the quality of clusters (L3)

**TEXT BOOKS:**

1. Jlawei Han &Micheline Kamber, Jian Pei ,"*Data Mining, Concepts and Techniques*", 3$^{rd}$ Edition, Morgan Kaufmann Publishers, 2012.
2. Pang- Ning Tan, Michael Steinbach, Vipin Kumar, "*Introduction to Data Mining*", 2$^{nd}$ Edition, Pearson Education, 2012. [For UNIT-1]

**REFERENCES:**

1. Arun K Pujari, "*Data Mining Techniques*", 4$^{th}$ Edition, Universities Press, 2016.
2. Sam Aanhory & Dennis Murray, "*Data Warehousing in the Real World*", 4$^{th}$ impression, 2009.
3. Margaret H Dunham, "*Data Mining Introductory and advanced topics*", 6$^{th}$ Edition, Pearson Education, 2009.

**WEB REFERENCES:**

1. https://swayam.gov.in/nd2_cec20_cs12/preview
2. http://www.saedsayad.com/data_mining.htm