DATA MINING

Course Code: 19CS1105 **L T P C**
3 0 0 3

Course Outcomes: At the end of the Course the student shall be able to:

CO1:Summarize the concepts of data mining and data preprocessing.

CO2: Illustrate different techniques of Association rule mining.

CO3:Apply classification methods on the given data.

CO4: Choose an appropriate algorithm to form clusters.

CO5:Describe the trends of data mining.

UNIT – I
(12Lectures)

INTRODUCTION: Data Mining, kinds of data can be mined, kinds of patterns to be mined,technologies used, Major issues in Data Mining.

DATA PREPROCESSING: An Overview- Data Cleaning: missing values, noisy data- Data Integration: Entity Identification, Redundancy and Correlation Analysis-Data Reduction: Overview of Data Reduction Strategies, Wavelet Transforms Principal Components Analysis, Attribute Subset Selection- Data Transformation and Data Discretization: Data Transformation Strategies Overview, Data Transformation by Normalization, Discretization by Binning.(Text book1:Ch-1,3).

Learning outcomes: At the end of unit, student will be able to

1.understand the kinds of data to be mined.(L2)
2.explain the data preprocessing techniques. (L2)
3.describe the data transformation and discretization techniques.(L2)

UNIT – II
(10Lectures)

ASSOCIATION RULE MINING: Mining Frequent Patterns-Apriori algorithm- FP-growth algorithm-Multi-level Association Rules- Mining Multidimensional Associations-Constraint-based Association mining:Metarule-guided mining of Association Rules. (Text book1:Ch-6)

Learning outcomes: At the end of unit, student will be able to

1.understand the frequent pattern mining methods.(L2)
2.outline the concepts of Multilevel and Multidimensional Space.(L2)
3.explain constraint based association rules. (L2)

UNIT – III **(12 Lectures)**

CLASSIFICATION: Basic concepts-Decision tree induction-Attribute selection measures–Bayesian classification:Bayes Theorem-Naive Bayesian Classification- Rule– based classification:using IF-THEN Rules-Rule Extraction from a Decision tree-Lazy learners-k-Nearest Neighbor Classifiers. (Text book1,Ch-8,9)

Learning outcomes: At the end of unit, student will be able to

1. understand the basic classification methods.(L2)
2. apply the Bayesian classification method on the given data.(L3)
3. make use of rule-based classification methods for given data.(L3)

UNIT – IV (10 Lectures)

CLUSTERING AND APPLICATIONS: Cluster analysis–Requirements–Overview of Clustering Methods– Partitioning Methods-Classical Partitioning Methods:k- means and k-Medoids–Hierarchical Methods:Agglomerative and Divisive Hierarchical Clustering.

DENSITY BASED METHODS:DBSCAN-OPTICS.(Text book1,Ch-10)

Learning outcomes: At the end of unit, student will be able to

1. outline the requirements for the clustering analysis. (L2)
2. apply clustering methods for data using partition methods and hierarchical methods .(L3)
3. describe density based clustering methods .(L2)

UNIT – V (8 Lectures)

DATA MINING TRENDS: Mining other kinds of data, Other Methodologies of Data Mining: Statistical Data Mining,Views on Data Mining Foundations, Visual and Audio Data Mining.

DATA MINING APPLICATIONS: Data Mining for Financial Data Analysis, Data Mining for Retail and Telecommunication Industries, Data Mining in Science and Engineering, Data Mining for Intrusion Detection and Prevention, Data Mining and Recommender Systems.(Text book1:Ch-13)

Learning outcomes: At the end of unit, student will be able to

1.understand the mining concepts of semi or unstructured data.(L2)
2.outline the different Methodologies to handle complex data.(L2)
3.explain Data Mining Applications.(L2)

TEXT BOOKS:

1. Huawei Han &Micheline Kamber, *Data Mining Concepts and Techniques*, 3rd Edition, Elsevier publishers, 2012.

REFERENCES:

1. Pang-Ning Tan, Micheal Steinbach, Vipin Kumar, *Introduction to data mining*,2nd Edition,Pearson publishers,2018.
2. Margaret H Dunham, *Data Mining Introductory and Advanced topics* ,Pearson publishers,1st Edition,2020.