RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

Computer Science & Information Technology, VIII-Semester

Departmental Elective CSIT-802(A) Data Warehousing & Mining

Objective:
1. Student should understand the value of Historical data and data mining in solving real world problems.
2. Student should become affluent with the basic Supervised and unsupervised learning algorithms commonly used in data mining.
3. Student develops the skill in using data mining for solving real-world problems.

Course Outcomes: After the completion of this course, the students will be able to:
1. Understand the functionality of the various data mining and data warehousing component
2. Appreciate the strengths and limitations of various data mining and data warehousing models
3. Explain the analyzing techniques of various data
4. Describe different methodologies used in data mining and data ware housing.
5. Compare different approaches of data ware housing and data mining with various technologies.

Unit-I
Data Warehousing: Introduction, Delivery Process, Data warehouse Architecture, Data Preprocessing: Data cleaning, Data Integration and transformation, Data reduction. Data warehouse Design: Datawarehouse schema, Partitioning strategy Data warehouse Implementation, Data Marts, Meta Data, Example of a Multidimensional Data model. Introduction to Pattern Warehousing.

Unit-II
OLAP Systems: Basic concepts, OLAP queries, Types of OLAP servers, OLAP operations etc. Data Warehouse Hardware and Operational Design: Security, Backup And Recovery,

Unit-III
Introduction to Data& Data Mining: Data Types, Quality of data, Data Preprocessing, Similarity measures, Summary statistics, Data distributions, Basic data mining tasks, Data Mining V/s knowledge discovery in databases. Issues in Data mining. Introduction to Fuzzy sets and fuzzy logic.

Unit-IV
Supervised Learning: Classification: Statistical-based algorithms, Distance-based algorithms, Decision tree-based algorithms, Neural network-based algorithms, Rule-based algorithms, Probabilistic Classifiers

Unit-V

Clustering & Association Rule mining : Hierarchical algorithms, Partitional algorithms, Clustering large databases – BIRCH, DBSCAN, CURE algorithms. Association rules : Parallel and distributed algorithms such as Apriori and FP growth algorithms.

Recommonded Books:

1. Pang – ningTan , Steinbach & Kumar, *"Introduction to Data Mining"*, Pearson Edu, 2019.

2. Jaiwei Han, Micheline Kamber, *"Data Mining : Concepts and Techniques"*, Morgan Kaufmann Publishers.

3. Margaret H. Dunham, *"Data Mining : Introductory and Advanced topics"*, Pearson Edu.,

4. Anahory& Murray, *"Data Warehousing in the Real World"*, Pearson Edu.

List of Experiments:

1. Create an Employee Table with the help of Data Mining Tool WEKA.

2. Create a Weather Table with the help of Data Mining Tool WEKA.

3. Apply Pre-Processing techniques to the training data set of Weather Table

4. Apply Pre-Processing techniques to the training data set of Employee Table

5. Normalize Weather Table data using Knowledge Flow.

6. Normalize Employee Table data using Knowledge Flow.

7. Finding Association Rules for Buying data.

8. Finding Association Rules for Banking data.

9. Finding Association Rules for Employee data.

10. To Construct Decision Tree for Weather data and classify it.