16CS305 DATA MINING TECHNIQUES

Hours Per Week:

L	T	P	C
3	-	2	4

Total Hours:

L	T	P	WA/RA	SSH/HSH	CS	SA	S	BS
45	-	30	5	40	3	8	5	2

Course Description and Objectives:

This course introduces the concepts and techniques of knowledge discovery and data mining. It also focuses on issues relating to the feasibility, usefulness, effectiveness and scalability of techniques for the discovery of patterns hidden in *large data sets*. The objective of this course is to apply the techniques of clustering, classification, association finding, feature selection and visualization of real world data.

Course Outcomes:

The student will be able to:

- describe various kinds of data sets used for analysis.
- illustrate various methods used for data cleaning.
- apply data mining techniques for crucial decision making.
- identify the necessity of correlation analysis for association mining.
- examine various clustering techniques in data mining.

SKILLS:

- ✓ *Perform in depth analysis of information requirements for solving problems.*
- ✓ *Manage large databases.*
- ✓ *Deploy knowledge in decision support systems or intelligent systems, both in academic and in industrial environments.*
- ✓ *Design data warehouse for an organization.*

UNIT - 1   **L-9**

INTRODUCTION: Basic Data Mining Tasks, Data Mining Issues, Data Mining Metrics, Data Mining from a Database perspective.

UNIT - 2   **L-9**

PREPROCESSING TECHNIQUES: Data Cleaning, Integration, Transformation, Reduction, Similarity Measures.

UNIT - 3   **L-9**

ASSOCIATION RULES: Basic algorithms, Parallel and distributed algorithms, Incremental rules, Advanced association rule techniques, Measuring the quality of rules.

UNIT - 4   **L-9**

CLASSIFICATION: Statistical-Based algorithms, Distance-Based algorithms, Decision Tree-Based algorithms, Neural Network-Based algorithms, Rule-Based algorithms, Combining Techniques.

UNIT - 5   **L-9**

CLUSTERING: Similarity and distance measures, Hierarchical algorithms, Partitional algorithms, Clustering large databases, Clustering with categorical attributes.

LABORATORY EXPERIMENTS

LIST OF EXPERIMENTS: Total Hours: 30

1. Creation of multidimensional data model using PL/SQL queries.

2. Implement various schemas using multidimensional data model using PL/SQL.

3. Implement various OLAP operations such as SLICE, DICE, ROLLUP, DRILLDOWN and PIVOT.

4. Demonstration of data preprocessing on dataset.

5. Demonstration of association mining rules on a dataset.

    a. Apriori Algorithm

    b. FP Growth Algorithm

6. Demonstration of classification rule process on a dataset.

    a. PART

    b. One R

7. Demonstration of naïve Bayesian classification algorithm.

8. Demonstration of decision tree on a dataset.

    a. J48

9. Demonstration of tree pruning on a dataset.

10. Demonstration on clustering rules using various data mining clustering algorithms.

TEXT BOOK:

1.   J. Han and M. Kamber, "Data Mining: Concepts and Techniques", 3rd edition, Morgan Kaufman, 2012.

REFERENCE BOOKS:

1.   M. H. Dunham, "Data Mining: Introductory and Advanced Topics", 1st edition, Pearson Education, 2001.

2.   I. H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques", 2nd edition, Morgan Kaufmann, 2000.

3.   D. Hand, H. Mannila and P. Smyth, "Principles of Data Mining", 2nd edition, Prentice-Hall. 2001.