Data Mining and Data Warehousing
EG3105CT

Year: III   **Total: 6 hours /week**
Part: I   **Lecture: 3 hours/week**
Tutorial: … hour/week
Practical: … hours/week
Lab: 3 hours/week

Course description:
This course studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. The course will cover all these issues and will illustrate the whole process by examples.

Course objectives:
After completion of this course students will be able to:
1. Explain the concept of Data preprocessing, Data Mining and Data Warehousing
2. Understand Data preprocessing Techniques.
3. Discuss multi-dimensional data representation and OLAP operations
4. Understand the concept and use of clustering, classification, and association rule mining algorithms.
5. Discuss on advanced concept and trends of Data Mining and Data Warehousing.

Course Contents:
Theory

Unit 1. Introduction to Data Mining   **[5 Hrs.]**
1.1.   Basic concepts of Data Mining
1.2.   Use and benefits of Data Mining
1.3.   Application of data mining
1.4.   Knowledge Discovery Process (KDD)
1.5.   Data Mining Functionalities
1.6.   Data Mining System Architecture

Unit 2. Data Preprocessing   **[5 Hrs.]**
2.1.   Data Objects and attribute types
2.2.   Statistical description of data
2.3.   Data Preprocessing Concepts
2.4.   Data Preprocessing
2.4.1. Data Cleaning
2.4.2. Data Integration
2.4.3. Data Reduction
2.4.4. Data Transformation

Unit 3. Data Warehousing and Online Analytical Processing (OLAP)   **[8 Hrs.]**
3.1.   Basic concepts of data warehousing
3.2.   Use and benefits of data warehousing
3.3.   Application of data warehousing
3.4.   Characteristics of Data Warehouse
3.5.   Operational Database Vs. Data Warehouse
3.6.   Data Warehouse Architecture
3.7.   Data Warehouse Models: Enterprise Warehouse and Data marts

3.8. Multi-dimensional Data, Data Cube
3.9. Data Warehouse Schemas
    3.9.1. Star Schema
    3.9.2. Snowflake Schema
    3.9.3. Fact Constellation Schema
3.10. Fact tables and Dimensions Tables
3.11. OLAP Operations in Multidimensional data models: Roll-up, Drill-Down, Slice and Dice, and Pivot (Rotate) Operations
3.12. Types of OLAP Servers: ROLAP, MOLAP, HOLAP

Unit 4. Mining Frequent Pattern and Associations   **[7 Hrs.]**
4.1. Frequent patterns, Market basket analysis, Frequent Item sets, Support and Confidence, Association Rules
4.2. Finding Frequent Itemset (Apriori Algorithm)
4.3. Limitation and improving Apriori Algorithm

Unit 5. Classification   **[8 Hrs.]**
5.1. Concept of Classification, Learning and Testing of Classification
5.2. Decision Tree Induction
5.3. Bayesian Classification
5.4. Rule Based Classification
5.5. Linear Regression

Unit 6. Clustering   **[8 Hrs.]**
6.1. Concept and Definition of Clustering
6.2. Clustering Methods and General Characteristics
    6.2.1. Partitioning Method (k-Means, k-Medoids)
    6.2.2. Hierarchical Method (Agglomerative, Divisive)

Unit 7. Data Mining Trends and Applications   **[4 Hrs.]**
7.1. Multimedia Data Mining
7.2. Text mining
7.3. Web Mining
    7.3.1. Web Content Mining
    7.3.2. Web Usage Mining
    7.3.3. Web Structure Mining
7.4. Data Mining support in SQL Server
7.5. Data Mining in Oracle
7.6. Data Mining Standards
7.7. Importance of data mining in Marketing, E- commerce and CRM
7.8. Aspects of Security and Privacy in Data Mining

Practical: **[45 Hrs.]**
1. Design data warehouse by using SQL Server or Oracle.
2. Implement OLAP operations
3. Implement clustering algorithms K-means and K- medoid by using Weka
4. Implement classification algorithms Naïve-Bayes and decision trees by using Weka
5. Implement regression algorithms by using Weka
6. Implement association mining algorithms by using Weka

Final written exam evaluation scheme			
Unit	Title	Hours	Marks Distribution*
1	Introduction to Data Mining	5	8
2	Data Preprocessing	5	8
3	Data Warehousing and Online Analytical Processing (OLAP)	8	16
4	Mining Frequent Pattern and Associations	7	12
5	Classification	8	12
6	Clustering	8	16
7	Data Mining Trends and Applications	4	8
	Total	**45**	**80**

* There may be minor deviation in marks distribution.

References:
1. Jiawei Han, Micheline Kamber, Jian Pei; *Data Mining: Concepts and Techniques*, Morgan Kaufman Publication, 3rd Edition
2. Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, *Introduction to DataMining*, Pearson Publication, First Edition
3. Charu C. Agrawal, *Data Mining: The Textbook*, Springer Nature Publication, First Edition
4. Sam Anahory, Dennis Murray, *Data warehousing In the Real World*, Pearson Education.
5. Alex Berson and Stephen J. Smith, *Data Warehousing, Data Mining & OLAP*, Tata McGraw Hill, 1st Edition.