IT_3253: DATA WAREHOUSING AND DATA MINING [3 0 0 3]

Objectives:

- To understand the concepts of pre-processing to ease the mining process
- To learn data warehouse architecture and OLAP operations
- To gain insight into the data mining algorithms
- To study about outliers and detection of anomalies
- To learn basics of web mining

Abstract:

Introduction to Data mining- Technologies, applications, Major issues in data mining, Data objects & attribute types, Statistical descriptions of data, Data visualization, Data pre-processing-data cleaning, data integration, data transformation, data reduction, data discretization, Data warehouse modeling, Data warehouse design and usage, Data warehouse implementation, Association rule mining techniques- Apriori algorithm, Partition algorithm, Pincer Search algorithm, FP Tree Growth algorithm, PC tree algorithm, Dynamic Itemset Counting algorithm, Multilevel association rules, Classification and prediction techniques- Decision Tree Induction, Bayes Classification Methods, Techniques to Improve Classification Accuracy, Clustering techniques- Partitioning Methods, Hierarchical Methods, Density-Based Methods, Outlier detection, Web mining.

Introduction: Data Mining-Introduction, Kinds of data mined, Technologies used, targeted Applications and Major issues in Data mining. **[1 hour]**

Data exploration: Data objects & attribute types, Statistical descriptions of data, Data Visualization, Measuring data similarity & dissimilarity **[3 hours]**

Data pre-processing: Overview, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization. **[6 hours]**

Data warehouse: Data warehouse definition, Data Warehouse Modeling: Data Cube and OLAP, Data Warehouse Design and Usage, Data Warehouse Implementation, Data Generalization by Attribute-Oriented Induction **[5 hours]**

Association rule mining: Market Basket Analysis, Frequent Itemset Mining Methods-Apriori Algorithm, Partition algorithm, Pincer-Search algorithm, Dynamic Itemset Counting algorithm, Frequent pattern tree, PC tree, Mining Frequent Itemsets Using the Vertical Data Format, Pattern Evaluation Methods **[8 hours]**

Classification and prediction: Classification: Basic concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Model Evaluation and Selection, Techniques to Improve Classification Accuracy **[5 hours]**

Clustering: Cluster analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Evaluation of Clustering. **[4 hours]**

Outlier detection: Outliers and Outlier Analysis, Outlier Detection Methods, Clustering-Based Approaches, Classification-Based Approaches **[2 hours]**

Web mining: Overview of information retrieval and Text and Web Page Pre-Processing **[2 hours]**

Outcomes:

The students are able to

- Describe the data warehouse architecture for facilitating querying
- Identify data mining techniques and apply on datasets
- Identify web mining techniques

References:

1. Han J. and Kamber M., *Data Mining: Concepts and Techniques (3e),* Morgan Kaufmann Publishers, 2011.
2. Pujari A. K., *Data Mining Techniques (4e),* Orient Blackswan, 2016.
3. Pang-N. T., Steinbach M., Anuj K., Vipin K., *Introduction to Data Mining, Pearson Education (2e)*, Pearson 2018.
4. Bing L., *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (2e),* Springer, Second Edition, 2011.