L	T	P	C
4	0	0	4

Course Code: INT303
Semester: VI

DATA WAREHOUSING & DATA MINING

Course Objectives

This course will help the learner to construct a data warehouse for analytical processing and apply data mining techniques to uncover interesting patterns.

UNIT - I   **15 Periods**

Data warehousing And OLAP: Basic concepts - Data Warehouse Modeling - Data cube and OLAP - **Data cube technology:** Data cube computation - Preliminary concepts. **Data Mining:** Introduction - Data Objects and Attribute Types - Basic Statistical Descriptions of Data, Measuring data similarity and dissimilarity. **Data Preprocessing:** Overview - Data cleaning - Data Integration - Data Reduction - Data Transformation and Data Discretization

UNIT - II   **15 Periods**

Mining Frequent Patterns, Associations, and Correlations: Basic Concepts - Frequent Itemset Mining Methods - Pattern evaluation methods, Pattern Mining in Multilevel, Multidimensional Space. **Classification:** Basic concepts - Bayes classification methods - Decision tree induction

UNIT - III   **15 Periods**

Cluster Analysis: Cluster Analysis - Partitioning Methods - Hierarchical Methods - Density Based Methods - Grid based methods - Evaluation of clustering. **Outlier Analysis:** Introduction - Extreme Value Analysis - Probabilistic Models - Clustering for Outlier Detection - Distance-Based Outlier Detection - Density-Based Methods

UNIT - IV   **15 periods**

Applications: Mining Text Data: Introduction - Document Preparation and Similarity Computation - specialized Clustering Methods for Text - Topic Modeling. **Mining Time Series Data:** Introduction -Time Series Preparation and Similarity - Time Series Forecasting - Time Series Motifs. **Mining Graph Data:** Introduction - Matching and Distance Computation in Graphs Transformation-Based Distance Computation - Frequent Substructure Mining in Graphs

TEXT BOOKS

1. Charu C.Aggarwal. *Data Mining: The Text Book*, Springer International Publishing, Switzerland, 2015.
2. Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques,* Morgan Kaufman, Third Edition, 2012.

REFERENCES

1. Mohammed J. Zaki, and Wagner Meira Jr. *Data Mining and Analysis - Fundamental Concepts and Algorithms,* Cambridge University Press, 2014.
2. Alex Berson, and Stephen J. Smith. *Data Warehousing, Data Mining and OLAP,* TataMcGraw Hill, 2012.
3. Ian H. Witten, Eibe Frank and Mark A. Hall. *Data Mining - Practical Machine LearningTools and Techniques,* Morgan Kaufman, Third Edition, 2011.
4. Vikram Pudi and P.Radhakrishna. *Data Mining.* Oxford University Press: 2009.

5.   Margaret H. Dunham. *Data Mining: Introductory and Advanced Topics.* Pearson Education, Sixth Impression, 2009.

ONLINE MATERIAL

1.   https://nptel.ac.in/courses/106105174/

UNITWISE LEARNING OUTCOMES

Upon successful completion of each unit, the learner will be able to

Unit I	• Describe the basics of data warehousing, data cube computation • Evaluate preprocessing techniques for sample data
Unit II	• Illustrate frequent patterns from data • Analyze data mining functionality such as classification
Unit III	• Apply any kind of basic clustering techniques for knowledge discovery • Illustrate outlier detection techniques to identify abnormal data
Unit IV	• Choose mining techniques to other kinds of data such as text data, time series data and Graph data • Analyze merits and demerits of mining techniques to other kinds of data

COURSE LEARNING OUTCOMES

Upon successful completion of this course, the learner will be able to

- Build a data warehouse, data cube and generate results and insights for analytical queries
- Evaluate preprocessing techniques on data of a given task.
- Illustrate frequent item sets and patterns from any structured data
- Compare different classification and clustering techniques
- Demonstrate outlier detection techniques to identify anomalies
- Recommend mining techniques to other kinds of data such as text, time series and Graph data