Semester VII

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Data Mining and Data Warehousing DSC-19	**4**	**3**	**1**	**0**	**Class XII pass**	**Database Management system**

Learning Objectives

This course objective is to introduce data mining principles and techniques. Introduce data mining as a cutting edge business intelligence tool. Develop and apply critical thinking, problem solving and decision making skills. Introduce the concepts of Data Warehousing, difference between database and data warehousing. Describe and demonstrate basic data mining algorithms, methods, and tools.

Learning outcomes

After completing this course, student should be able to;

- Design a data warehouse to present information needed by the end can be utilized for managing clients.
- Design and implement a quality data warehouse effectively and administer the data resources in such a way that it will truly meet management's requirements.
- Evaluate standards and new technologies to determine their potential impact on your information resource for a large complex data warehouse/data mart.
- Use data mining tools for projects and to build reliable products as per demand.

SYLLABUS OF DSC-19

Unit I: Overview: The process of knowledge discovery in databases, predictive and descriptive data mining techniques, supervised and unsupervised learning techniques. Data preprocessing: Data cleaning, Data transformation, Data reduction, Discretization. (9 hours)

Unit II: Classification: Supervised learning for predictive data mining, Basic issues in predictive data mining Decision trees, Decision rules, Statistical classification, Instance-based methods (nearest neighbor), Evaluation and Validation methods.

(9 hours)

Unit III: Clustering: Unsupervised learning for descriptive data mining, Basic issues in clustering, Partitioning methods, Hierarchical methods for clustering, Density-based methods, Cluster Validation methods and metrics. Association Rule Mining: Frequent item set, Maximal and Closed itemsets, Apriori property, Apriori algorithm. (12 hours)

Unit IV: Data Warehousing: Overview, Definition, Delivery Process, Difference between Database System and Data Warehouse, Multi-Dimensional Data Model, Data Cubes, Stars, Snow Flakes, Fact Constellations, Concept hierarchy, Process Architecture, 3 Tier Architecture, Data Marting. (12 hours)

Essential/recommended readings

1. Jiawei Han, Micheline Kamber, ”Data Mining Concepts & Techniques” Elsevier.
2. Mallach,”Data Warehousing System”,McGraw –Hill.
3. H.Dunham,”Data Mining:Introductory and Advanced Topics” Pearson Education.
4. Sam Anahory, Dennis Murray, “Data Warehousing in the Real World : A Practical Guide for Building Decision Support Systems, Pearson Education.
5. Charu C Agrawal, Data Mining: The Textbook, Springer, 2015.
6. J Zaki Mohammed and Wagner Meira, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, 2014.
7. P. Tan, M. Steinbach, A Karpatne, and V. Kumar, Introduction to Data Mining, 2nd Ed., Pearson Education, 2018.