Data Warehousing and Data Mining

Code	Credit Hours
CS-423	3-1

Course Description

This course provides an overview of fundamental data warehousing and data mining concepts. It introduces the concepts and strategies necessary to build and deploy a data warehouse as a decision support tool for an enterprise. Different data mining techniques e.g. classification, clustering would also be covered in this course. The course objective is that its successful completion should enable students to engineer database warehouse and to apply mining on real-world data repositories.

Text Book:

1. Ralph Kimball and Margy Ross. 2002. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (2nd Edition)

2. Jiawei han & Micheline Kamber Harcourt. Data Mining Concepts and Techniques –

Reference Book:

1. Joe Reis and Matt Housley. June 2022. Fundamentals of Data Engineering: Plan and Build Robust Data Systems (First Edition), O'Reilly

Prerequisites

None

ASSESSMENT SYSTEM FOR THEORY

Quizzes	15%
Assignments	05%
Mid Term	30%
ESE	50%

ASSESSMENT SYSTEM FOR LAB

Quizzes	10%-15%
Assignments	5% - 10%
Lab Work and Report	70-80%
Lab ESE/Viva	20-30%

Teaching Plan

Week No	Topics	Learning Outcomes
1	Introduction	Course Outline, objectives, teaching plan, assessment method, concepts review
2-8	Data Preprocessing, OLAP and Cube Computation	Data summarization. Data cleaning. Data integration. Data Transformation. Data Reduction. Data discretization. Concept hierarchy generation. Data Warehouse and OLAP Technology: What is data warehouse? Multidimensional data model. Data warehouse architecture.
9	**MID TERM IN WEEK 9**	
10-17	Data Mining techniques	This part of this course is more focused on data mining techniques such as frequent pattern mining, correlation analysis, classification and prediction, clustering analysis, evaluation metrics.
18	**FINAL TERM IN WEEK 18**	

Practical:

Experiment No	Description
1	Setting up the ETL environment and basic operations
2	Advanced ETL operations
3	Performing ETL through Talend – Extract operations
4	Performing ETL through Talend – Transform and Load operations
5	Data Visualization (Matplot library)
6	Data Preprocessing using python
7	Data Visualization and correlation analysis

8	Project Proposal Discussion
9	**MID TERM IN WEEK 9**
10	Data Visualization using Seaborn
11	Data classification using Scikit learn
12	Cluster analysis using Scikit learn
13	Aprori and FP growth algorithms using MLXtend
14	PageRank algorithm in python
15	**Open-Ended Lab**
16	**Open-Ended Lab**
17	**Final Lab**
18	**FINAL TERM IN WEEK 18**