Course Instructor: Dr Chiranjib Sur

Course Webpage: https://chiranjibsuruf.github.io/courses/da205.html

Email: chiranjib@iitg.ac.in

Objectives: Data Mining and Warehousing teaches you how to turn scattered, messy data into a reliable, analysis-ready foundation for decision-making. It equips you with the end-to-end skills, ETL, modeling, storage, and mining, to extract real value from big, real-world datasets.

Prerequisites: Algorithms, Data Structure and Basic Programming.

Course Code: DA205

Course Name: Data Mining and Warehousing

Credits: 3-0-0-6

Syllabus: Introduction: Definitions, Review of basics of data analysis, applications; Data Warehousing: Definition, architectures, dimensional modeling and star schema, ETL (extract, transform, load) processes; Finding Similar Items: Similarity measures and distance metrics, Shingling and locality-sensitive hashing; Frequent Pattern Mining: Itemset, substring, sequence, pattern evaluation and interestingness measures; Graph Mining: Graph data and representations, link analysis, pattern mining, graph clustering techniques; Mining Data Streams: characteristics of data streams, sliding window models, approximate and sketching techniques, change detection and concept drift.

Textbooks:

- Han, Jiawei, Jian Pei, and Hanghang Tong, Data mining: Concepts and Techniques, 4th Edition, Morgan Kaufmann, 2022.
- Robinson, Ian, Jim Webber, and Emil Eifrem, Graph Databases: New Opportunities for Connected Data, 2nd Edition, O'Reilly Media, Inc., 2015.
- Ponniah, Paulraj, Data Warehousing Fundamentals for IT Professionals, 2nd Edition, John Wiley & Sons, 2012..

References:

- Jure Leskovec, Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, 3rd Edition, Cambridge University Press, 2020.
- Kimball, Ralph, and Margy Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Wiley, 2013.
- Reis, Joe, and Matt Housley, Fundamentals of Data Engineering, O'Reilly Media, Inc., 2022.