

Course Name: **DS- 311 Data Mining**

Credit Hours: 2-1

Contact Hours: 2-3

Pre-requisites: Statistics II, Introduction to Data Science

Course Introduction:

Data Mining has emerged at the confluence of artificial intelligence, statistics, and databases as a technique for automatically discovering hidden patterns in large datasets. The main purpose of this course is the ability to analyze and construct knowledge from data. The aims of this course are to:

- Expand on the student's understanding and awareness of the concepts of data mining basics, techniques, and application.
- Introduce the concepts of Data Pre-processing and Summary Statistics.
- Introduce the concepts of Frequent Item Set Generation, Associations and Correlations measures.
- Introduce the concepts of Classification, Prediction, and Clustering algorithms.

Build on the programming and problem-solving skills developed in previous subjects studied by the student, to achieve an understanding of the development of Classification, Prediction, and Clustering applications.

CLO No	Course Learning Outcomes	Bloom Taxonomy
CLO-1	Apply preprocessing techniques on any given raw data.	C3 (Apply)
CLO-2	Select and apply proper data mining algorithm to discover interesting patterns	C3 (Apply)
CLO-3	Analyze and extract patterns to solve problems and point out how to deploy solution	C4 (Analyze)
CLO-4	Evaluate systematically supervised, semi supervised and unsupervised models and algorithms with respect to their accuracy	C4 (Analyze)

Course Outline:

Introduction to data mining and basic concepts, Pre-Processing Techniques & Summary Statistics, Association Rule mining using Apriori Algorithm and Frequent Pattern Trees, Introduction to Classification Types, Supervised Classification (Decision trees, Naïve Bae Classification, K-Nearest Neighbors, Support Vector Machines etc.), Unsupervised Classification (K Means, K Median, Hieratical and Divisive Clustering, Kohonan Self Organizing maps), outlier & anomaly detection, Web and Social Network Mining, Data Mining Trends and Research Frontiers. Implementing concepts using Python.

Reference Materials:

1. Jiawei Han & Micheline Kamber, Jian Pei (2011). Data Mining: Concepts and Techniques, 3rd Edition.
2. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar (2005). Introduction to Data Mining.
3. Charu C. Aggarwal (2015). Data Mining: The Textbook
4. D. Hand, H. Mannila, P. Smyth (2001). Principles of Data Mining. MIT Press.