Course Code and Title: **IT563 - Data Mining**
Credit Structure (L-T-P-Cr): 3-0-2- 4

Instructor: PM Jat, 203 FB-2; extension: 641; email: pm_jat@daiict.ac.in

Course Category: **Group Core (Group-I)**

Prerequisites (if any)/Required skills: Databases, Statistics, Probability, and Linear Algebra

**Course objectives:**

This course builds foundation for large scale Data Mining and Predictive Analysis.

**Abstract Course Content:**

- Selected methods for following data mining tasks: Classification, Frequent Item Set, Association Rule Mining, Clustering, Link Analysis, Feature selection.
- Big Data Analysis Architectures like Hadoop ecosystem, Spark, No SQL databases.
- Challenges and Techniques in Mining Large Scale datasets. Map-Reduce/Spark based algorithms for selected data mining tasks.
- Certain case studies drawn from one and more verticals, for instance Text Mining, Web Mining, Recommendation Systems, Social Media Analysis, Opinion Mining.

During Lab hours, students will be asked to perform various experiments on open source data repositories. They will have option of using Python, Java, or R. Students will also be required to do a data mining project/study leading to a term paper.

**Grading:**

1. In-Semester exam: 20%
2. End-Semester Examination: 30%
3. Project Work: 50%

**Suggested textbook/references:**

1. Introduction to Data Mining by Pang-Ning Tan, Vipin Kumar, Michael Steinbach, Pearson Education, 2006