



Course Handout

Course Code: CS 451

L-T-P-C: 3-0-2-4

Session: 2017-2018

Course Title: Data Mining

Course In-charge: Pragya Verma

Contact Details:

Email: pragya.verma@niituniversity.in

Phone No: +91-1494-660-668

Course Objectives:

This course provides an introduction to fundamental aspects of data mining which is extraction of valid, novel, potentially useful, understandable patterns from data sources such as, databases, texts, web, images, etc. This course covers the importance of data preprocessing, major tasks in data preprocessing, association rule mining, different types of clustering techniques, decision tree construction, and evaluation of classifier performance.

Text Books:

1. Data Mining Concepts and Techniques, Jiawei Han, Michaeline Kamber, Jian Pei, Morgan Kaufmann publisher, 3rd Edition, 2011, ISBN-10: 9380931913.

Reference Books:

1. Introduction to Data Mining, Tan, Steinbach and Kumar, Pearson, 1st Edition, 2005, ISBN-10: 0321321367.
2. The Elements of Statistical Learning, Hastie, Tibshirani and Friedman, Springer, 7th Edition, 2013, ISBN-10: 0387848576.

Evaluation Scheme:

Component	Date	Time	Mode	Weightage
Mid Term Test I	TBA	1 hour	Pen paper	20%
Mid Term Test II	TBA	1 hour	Pen paper	20%
Attendance				10%
Comprehensive Test	TBA	2 hours	Pen paper	30%
Project	TBA	--	--	20%

Lecture Plan (Tentative):

Sl. No.	Topic	Number of Lectures
1.	Introduction: Definition of data mining; Knowledge discovery from databases (KDD); data mining tasks- regression, classification, association rule mining, clustering, sequential pattern mining, deviation detection, data visualization; Importance of data mining; data mining applications	02
2.	Data Preprocessing: Data types and forms (numerical, categorical); Importance of data preprocessing; Major tasks in data preprocessing- data cleaning, data integration, data transformation, data reduction, data discretization	07
3.	Association Rule Mining: Rules, Rule strength measures, support and confidence; Mining algorithms: Apriori algorithm, Frequent pattern growth method; Frequent patterns, maximal and closed frequent item sets; Rule generation- Naive algorithm, Apriori algorithm for rule generation; Pattern evaluation methods	08
4.	Clustering: Cluster analysis, application of cluster analysis; Dissimilarity measures- Types of data in clustering analysis- interval scaled variables, binary variables, nominal, ordinal and ratio variables, variables of mixed types; Clustering techniques- partitioning (K-means, K-medoids, Bisecting K-means), hierarchical (agglomerative, divisive), density based (DBSCAN);	11
5.	Advanced Cluster Analysis: Probabilistic Model-Based clustering, Clustering High-Dimensional Data, Clustering Graph and Network Data, Clustering with constraints	05
6.	Classification: Definition; Classification using distance (K Nearest Neighbors); Classification techniques- Decision tree based methods, rule based methods, memory based reasoning, Bayesian belief networks, Model evaluation and selection: metrics for evaluating classifier performance, cross validation, bootstrap, holdout methods and random subsampling	12
TOTAL		45

Lab Plan (Tentative)

Students will do project work during lab sessions. The project will involve programming, design & development, and validation of different data mining techniques. Project will be selected based on different data mining aspects. Students should complete the selected project within the same semester only. Students should submit a report based on their project and present the same.

Consultation Hour

The students can meet the Course-in-charge on matters related to the course being taught on all days from Monday to Friday by prior appointment fixed via email: write to: pragya.verma@niituniversity.in.

Make up Policy

Students who are likely to miss a component of evaluation on a **genuine reason** may be given a make-up of that component by the Course-in-charge. The students are required to approach the Course-in-charge immediately for the same *before the conduct of the evaluation component*. The decision of the Course-in-charge in this matter shall be final.

Plagiarism Policy

All work is supposed to be your own. There will be a zero-tolerance policy towards plagiarism. For any unfair means used, the university's policies will apply.

Grading Policy

The Marks obtained in **all the components of Evaluation shall be totalled** and the final marks shall be converted in the letter grades, namely, A, B, C, D and E. The grading is relative and normally it is centred around the average in a class.

University Attendance Policy

University Attendance Policy as in vogue will apply to this course as well.

Teaching Learning Process and Asian Lenses Connect

Lectures will be a combination of power point and whiteboard presentations. Several example problems will be solved in class to illustrate the theory. The course also covers the Asian contribution in the field of Data Mining.