IT496: Introduction to Data Mining
Autumn AY 2023-24

Instructor:	Dr. Arpit Rana, arpit_rana@daiict.ac.in
Prerequisites:	Programming in Python
Slot:	B.Tech. V or VII Sem / M.Tech. (ICT) I Sem / M.Sc. (DS) III Sem
Category:	Technical Elective [class size < 120]
Course Credits:	3--0--2--4 (L--T--P--Cr)
Lectures [LT-01]:	Tuesday: 10:00 – 11:00 Hrs.	Thursday, Friday: 11:00 – 12:00 Hrs.
Lab Session[LT-02]:	Friday, 14:00 – 16:00 Hrs.
TA contact info:	Himanshu Beniwal (himanshubeniwal@iitgn.ac.in) – Head TA Vipasha Vaghela (202211002@daiict.ac.in) – DA-IICT TA-1 Aksh Patel (202211037@daiict.ac.in) – DA-IICT TA-2

Course Description:

Data mining is the study of algorithms for finding patterns in large data sets. Data mining is an interdisciplinary course involving topics from statistics, databases, machine learning, and algorithms. This course will cover the basic algorithms for data preprocessing, association rules, regression, classification, clustering, anomaly detection, and visualization. We will also explain implementations on real-world datasets. Finally, course projects on industrial problems will be assigned to the student groups (size of 4) for their practical understanding.

Suggested Books:

- Introduction to Data Mining, Tan, Steinbach, and Vipin Kumar, Pearson Education, 2019
- Data Mining: Concepts and Techniques. Han, J., Kamber, M., and Pei, J. Waltham: Morgan Kaufmann, 2011

Course Outcomes:

After successfully completing the course, the student will be able to -

- Understand the fundamentals of Data Mining.
- Develop skills in using recent Data Mining techniques for solving practical problems.
- Gain experience in doing independent study and research
- Apply all of the above in Python (using Python ML Stack: Scikit-learn, Numpy, Pandas)

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
X	X	X		X				X			X

IT496: Introduction to Data Mining
Autumn AY 2023-24

Evaluation Scheme:

Ability to Understand and Apply Theoretical Concepts		Ability to Implement Mining Techniques on Real Data	
Surprise Quizzes (25 %)	End-Term Exam (25 %)	Course Projects (40 %)	Case Study (10%)

Tentative Course Plan[1]:

Units	Topics	# Lectures
Introduction	Definition, objectives, components, approaches, and challenges	1
Statistics for Data Mining	Data Objects and Attribute Types, Basic Statistical Measures of Data, Measures of Data Proximity	3
Data Preprocessing	Data Cleaning, Data Integration, Data Reduction, Data Transformation, and Discretization	1
Regression and Classification: Basic Concepts	Binary Classifier, Multiclass, Multilabel, and Multi-output Classification; Performance Measures; Error Analysis, Data Partitioning: Holdout, Random sampling, and Cross-validation Regression Analysis: Lasso, Ridge, and Elastic; Loss functions, Optimising Loss Functions using GD, SGD, BGD, ALS; Bias-Variance Trade-off; Model Overfitting and Regularization; Class-imbalance: Undersampling, Oversampling, and Stratification	12
Classification Techniques	Eager Classifiers: Decision Trees, Support Vector Machines, and Neural Networks. Lazy Classifiers: K-Nearest Neighbor, Case-based Reasoning Ensemble Techniques: Voting Classifiers, Bagging and Pasting, Random Forests, Boosting	12
Dimensionality Reduction	The Curse of Dimensionality, PCA, SVD	2
Cluster Analysis	Basic concepts, Partitioning-based methods, Hierarchical methods, and Density-based methods; Evaluation of Clustering	3
Outlier Analysis	Basic concepts, Outlier Detection Approaches: Statistical, Proximity-based, Clustering-based	3
Frequent Itemset Mining	Mining Frequent Patterns, Association, and Correlations, Market-Basket Analysis, Apriori Algorithm, Evaluation of Frequent Patterns	3

1 — This course plan is subject to change without notice.

IT496: Introduction to Data Mining
Autumn AY 2023-24

Preliminary Schedule of the Course:

Week	Lecture	Lab	Due[1]
Week-1 [24 July 2023]	Introduction	– No lab –	-
Week-2 [31 July 2023]	Statistics for Data Mining	Environment Setup and Data Scraping	-
Week-3 [07 Aug 2023]	Data Preprocessing	End-to-End ML Project in Python	-
Week-4 [14 Aug 2023]	Fundamentals of Predictive Analytics - I Holidays: 15 Aug (Tues), 16 Aug (Wed)	CP - 1	Sunday, 10 Sept. 2023
Week-5 [21 Aug 2023]	Fundamentals of Predictive Analytics – II		
Week-6 [28 Aug 2023]	Regression Techniques Holidays: 30 Aug (Wed) | 28 Aug (Mon) to be treated as Tues		
Week-7 [04 Sept 2023]	Dimensionality Reduction Holidays: 07 Sept (Thurs)		
Week-8 [11 Sept 2023]	First In-Semester Exam Week	Evaluation: CP - 1	
Week-9 [18 Sept 2023]	Eager Classifiers – I: Support Vector Machines and Decision Trees Holidays: 19 Sept (Tues)	CP - 2	Sunday, 15 Oct 2023
Week-10 [25 Sept 2023]	Eager Classifiers – II: Neural Networks Holidays: 28 Sept (Thurs) | 29 Sept (Fri) to be treated as Thurs		
Week-11 [02 Oct 2023]	Eager Classifiers – III: Neural Networks Contd. Holidays: 02 Oct (Mon)		
Week-12 [09 Oct 2023]	Lazy Classifiers and Ensemble Techniques		
Week-13 [16 Oct 2023]	Second In-Semester Exam Week	Evaluation: CP - 2	

IT496: Introduction to Data Mining
Autumn AY 2023-24

Week	Lecture	Lab	Due[1]
Week-14 [23 Oct 2023]	Cluster Analysis – I Holidays: 24 Oct (Tues)	CP - 3	Sunday, 19 Nov 2023
Week-15 [30 Oct 2023]	Cluster Analysis – II Holidays: 31 Oct (Tues) | 03 Nov (Fri) to be treated as Tues		
Week-16 [6 Nov 2023]	Outlier Analysis		
Week-17 [13 Nov 2023]	In-Semester Break		
Week-18 [20 Nov 2023]	Association Rule Mining	Evaluation: CP - 3	
Week-19 [27 Nov 2023]	End-semester Examination		

1 – Course Projects (CPs) are due at 11:59 PM on the due date listed.

Note-1: CPs will be allocated to groups, with each group consisting of 4-5 members. Additionally, the assigned group will be responsible for delivering a case study on an AI-focused startup that addresses issues of the nation's priority and global interest.

Note-2: Further instructions related to the labs and case study presentations will be provided later on the classroom portal.