

CTMTAIDS SI P4: Fundamentals of Data Science and Machine Learning

Teaching Scheme					Evaluation Scheme									
L	T	P	C	TCH	Theory							Practical		Total
					Internal Exams					University Exams		University Exams (LPW)		
					TA-1		MSE		TA-2 *	Marks	Hrs	Marks	Hrs	
					Marks	Hrs	Marks	Hrs	Marks					
03	00	00	03	03	25	00:45	50	01:30	25	100	03:00	-	-	200

* Note: TA-2 will be in form of assignments or workshops.

Objectives

1. To explore the fundamental concepts of data science and machine learning.
2. To understand data analysis techniques for applications handling large data.
3. To visualize and present the inference using various tools.

UNIT -I

Introduction to core concepts and technologies: Introduction, Terminology, data science process, data science toolkit, Types of data, Example applications, Introduction to Statistical Methods: basic and some advanced concepts of probability and statistics; Concepts of statistics in solving problems arising in data science.

UNIT -II

Data preprocessing, Data cleaning, data integration, Data Reduction Data Transformation and Data Discretization. Evaluation of classification methods, Confusion matrix, Students T-tests and ROC curves-Exploratory Data Analysis Basic tools (plots, graphs and summary statistics) of EDA, Philosophy of EDA.

UNIT -III

Basic Machine Learning Algorithms: Association Rule mining, Linear Regression, Logistic Regression. Classification, k-Nearest Neighbors (k-NN), k-means, Decision tree, Naive Bayes, Ensemble Methods Random Forest. Feature Generation and Feature Selection, Feature Selection algorithms, Filters, Wrappers, Decision Trees, Random Forests.

UNIT -IV

Clustering, Choosing distance metrics, Different clustering approaches, hierarchical agglomerative clustering, k-means, DBSCAN, Relative merits of each method, clustering tendency and quality. Computer science and engineering applications Data mining, Network protocols, analysis of Web traffic.

UNIT -V

Data visualization introduction, Types of data visualization, Data types. Applications of Data Science, Technologies for visualization, Tools for data visualization, recent trends in various data collection and analysis techniques, various visualization techniques, application development methods used in data science.

Reference Books: -

1. Cathy O’Neil, Rachel Schutt, “Doing Data Science”, Straight Talk from The Frontline, O’Reilly, 2013.
2. Han, J., Pei, J. and Tong, H., “Data mining: concepts and techniques.”, 2022.
3. Davy Cielen, Arno D. B. Meysman, Mohamed Ali, “Introducing Data Science,” Manning Publications Co., 3 rd edition, 2016.
4. Mohammed J. Zaki and Wagner Miera Jr, “Data Mining and Analysis: Fundamental Concepts and Algorithms”, Cambridge University Press, 2014.
5. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, “An Introduction to Statistical Learning: with Applications in R,” Springer, 1st edition, 2013.
6. Jure Leskovek, Anand Rajaraman, Jeffrey Ullman, “Mining of Massive Datasets”, v2.1, Cambridge University Press, 2014.
7. Joel Grus, O’Reilly, “Data Science from Scratch: First Principles with Python”, 1st edition, 2015.