

CDMR1: Introduction to Data Science and Machine Learning

Minor Degree

Course Objectives:

1. To provide the knowledge and expertise to become a proficient data scientist.
2. Demonstrate an understanding of statistics and machine learning concepts that are vital for data science
3. Produce Python code to statistically analyze a dataset
4. Critically evaluate data visualizations based on their design and use for communicating stories from data

Course Outcomes:

After completion of course, students would be able:

1. To explain how data is collected, managed and stored for data science
2. To understand the key concepts in data science, including their real-world applications and the toolkit used by data scientists
3. To understand the supervised learning
4. To understand the unsupervised learning

UNIT-I

[CO1] (12 periods)

Introduction to data science: Data science process, Applications of Data Science, types of data, exploring the structure of data, data quality and remediation, data pre-processing.

Python libraries for Data Science: Numpy, Pandas, Matplotlib.

UNIT-II

[CO2] (12 periods)

Modeling and evaluation- Selecting and training a model, representation and interpretability, performance evaluation, performance improvisation.

Feature engineering- Introduction, feature transformation, subset selection.

Bayesian Decision Theory- Bayes' theorem and concept learning.

UNIT-III

[CO3] (12 periods)

Introduction to Supervised Machine Learning

Classification- Classification model, learning steps, Algorithms- KNN, Decision tree, Random Forest model, Support vector machines, Applications, and examples.

Regression- Simple linear regression, Multiple linear regression, Polynomial regression, Logistic regression, Maximum likelihood estimation, Applications, and examples.

UNIT-IV

[CO4] (12 periods)

Introduction to Unsupervised Machine Learning

Supervised vs Unsupervised learning, types of clustering techniques, partitioning methods, k-medoids, hierarchical clustering, density-based methods-DBSCAN, finding pattern using association rule, Applications, and examples.

Textbooks:

1. Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining, Glenn J. Myatt.
2. Machine Learning, Amit Kumar Das, Saikat Dutt, Subramanian Chandramouli, Pearson publications (2018).
3. Introduction to Machine Learning, Ethem Alpaydın, Fourth Edition, The MIT Press Cambridge, Massachusetts (2020).

References:

1. Making Sense of Data Making Sense of Data, A Practical Guide to Exploratory Data Analysis and Data Mining.
2. Hands-On Exploratory Data Analysis with Python, Suresh Kumar Mukhiya, Usman Ahmed, March 2020, Packt Publisher.
3. The Elements of Statistical Learning, Trevor Hastie, Robert Tibshirani, Jerome Friedman, Springer Series in Statistics, Springer (2009).