

MACHINE LEARNING AND DATA SCIENCES

III-B.Tech.-II-Sem.

Subject Code: CS-PCC-322

L T P C

3 - - 3

Course Outcomes: Upon completion of the course, the student will be able to

1. demonstrate the mathematical foundations needed for Machine Learning and Data Sciences
2. outline the functionalities of machine learning
3. illustrate learning algorithms & data science basics
4. build data science applications using Python based toolkits
5. make use of Recommender Systems and Sentiment Analysis in real time applications

CO – PO Mapping				
POs	PO2	PO3	PO6	PO8
CO1	3	2	1	1
CO2	3	3	3	3
CO3	3	2	2	2
CO4	3	3	3	3
CO5	3	3	3	3
3-Strong; 2-Medium; 1-Weak				

Unit-I: Mathematical Foundations**10 hours**

Linear Algebra: Vectors, Matrices, Statistics: Describing a Single Set of Data, Correlation, Simpson's Paradox, Correlation and Causation, Probability: Dependence and Independence, Conditional Probability, Bayes's Theorem, Random Variables, Continuous Distributions, The Normal Distribution, The Central Limit Theorem, Hypothesis and Inference: Statistical Hypothesis Testing, Confidence Intervals, Phacking, Bayesian Inference.

Unit-II: Machine Learning**10 hours**

Overview of Machine learning concepts – Over fitting and train/test splits, Types of Machine learning – Supervised, Unsupervised, Reinforced learning, Introduction to Bayes Theorem, Linear Regression-model assumptions, regularization (lasso, ridge, elastic net), Classification and Regression algorithms- Naïve Bayes, K-Nearest Neighbors, logistic regression, support vector machines (SVM), decision trees, and random forest, Classification Errors.

Unit-III: Advanced Machine Learning and Introduction to Data Sciences (4 + 5) 10 hours

Part-A: Find-S: finding a maximally specific hypothesis, Analysis of Time Series- Linear Systems Analysis, Nonlinear Dynamics, Rule Induction, Neural Networks - Learning and Generalization, Overview of Deep Learning.

Part-B: Introduction to Data Science: Concept of Data Science, Traits of Big data, Web Scraping, Analysis vs reporting, Data Science in business.

Unit-IV: Programming Tools for Data Science**9 hours**

Toolkits using Python: Matplotlib, NumPy, Scikit-learn, NLTK, Visualizing Data: Bar Charts, Line Charts, Scatterplots, Working with data: Reading Files, Scraping the Web, Using APIs (Example: Using the Twitter APIs), Cleaning and Munging, Manipulating Data, Rescaling, Dimensionality Reduction.

Unit-V: Recommender Systems and Sentiment Analysis**10 hours**

Recommender Systems: Introduction, Content-Based Filtering, Collaborative Filtering, Hybrid Recommenders.

Sentiment Analysis: Introduction, Data Cleaning, Text Representation.

Textbooks:

1. Joel Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media(unit-1)
2. Jeeva Jose, "Machine Learning", Khanna Publishing House, Delhi. (unit-2&3)
3. Chopra Rajiv, "Machine Learning", Khanna Publishing House, Delhi. (unit2&4)
4. Introduction to data science by Igual, Laura & Seguí, Santi, Springer. (unit-5)

References:

1. Machine Learning – Tom M. Mitchell, TMH.
2. Jain V.K., "Data Sciences", Khanna Publishing House, Delhi.