RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

Credit Based Grading System

Computer Science and Engineering

VII-Semester CS-7005 Elective-IV (2) Data Science & Big data

1. Understanding Data: Data Wrangling and Exploratory Analysis, Data Transformation & Cleaning, Feature Extraction, Data Visualization. Introduction to contemporary tools and programming languages for data analysis like R and Python.
2. Statistical & Probabilistic analysis of Data: Multiple hypothesis testing, Parameter Estimation methods, Confidence intervals, Bayesian statistics and Data Distributions.
3. Introduction to machine learning: Supervised & unsupervised learning, classification & clustering Algorithms, Dimensionality reduction: PCA & SVD, Correlation & Regression analysis, Training & testing data: Over fitting & Under fitting.
4. Introduction to Information Retrieval: Boolean Model, Vector model, Probabilistic Model, Text based search: Tokenization,TF-IDF, stop words and n-grams, synonyms and parts of speech tagging.
5. Introduction to Web Search& Big data: Crawling and Indexes, Search Engine architectures, Link Analysis and ranking algorithms such as HITS and Page Rank, Hadoop File system & MapReduce Paradigm

Text Books:

1. Field Cady, "The Data Science Handbook" , 1/e ,2018,Publisher: Wiley
2. Sinan Ozdemir, "Principles of Data Science ", 1/e, 2016Packt Publishing Limited

References:

1. Peter Bruce, "Practical Statistics for Data Scientists: 50 Essential Concepts",Shroff/O'Reilly; First edition, 2017
2. Pang-Ning Tan, "Introduction to Data Mining", Pearson Edu.
3. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, "Modern Information Retrieval", Pearson Education