

# RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

## New Scheme Based On AICTE Flexible Curricula

### CSE-Artificial Intelligence and Machine Learning/ Artificial Intelligence and Machine Learning, VI-Semester

#### Departmental Elective AL603 (B) Data and Visual Analytics

**Unit 1 : Data Definitions and Analysis Techniques:** Elements, Variables, and Data categorization Levels of Measurement Data management and indexing Introduction to Statistical Concepts: Sampling Distributions, Resampling, Statistical Inference and Descriptive Statistics, Measures of central tendency, Measures of location of dispersions

**Unit 2: Advance Data analysis techniques:** Statistical hypothesis generation and testing, Chi-Square test, t-Test, Analysis of variance, Correlation analysis, Maximum likelihood test, Regression Modelling, Multivariate Analysis, Bayesian Modelling, Inference and Bayesian Network, Regression analysis

**Unit 3: Data Wrangling:** Intro to Data Wrangling, Gathering Data, Assessing Data, Cleaning Data.

**Data Visualization in Data Analysis:** Design of Visualizations, Univariate Exploration of Data, Bivariate Exploration of Data, Multivariate Exploration of Data, Explanatory Visualizations.

**Unit 4: Data Ecosystem:** Overview of the Data Analyst Ecosystem, Types of Data, Understanding Different Types of File Formats, Sources of Data, Overview of Data Repositories, NoSQL, Data Marts, Data Lakes, ETL, and Data Pipelines, Foundations of Big Data, Big Data processing tools such as Hadoop, Hadoop Distributed File System (HDFS), Hive, and Spark

**Unit 5: Data Visualization tools:** Python visualization libraries (matplotlib, pandas, seaborn, ggplot, plotly), Introduction to PowerBI tools, Examples of inspiring (industry) projects- Exercise: create your own visualization of a complex dataset.

Text Books/References:

1. Joel Grus, Data Science from Scratch, Shroff Publisher Publisher /O'Reilly Publisher Media
2. Annalyn Ng, Kenneth Soo, Numsense! Data Science for the Layman, Shroff Publisher Publisher
3. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly Publisher Media.
4. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press.
5. Jake VanderPlas, Python Data Science Handbook, Shroff Publisher Publisher /O'Reilly Publisher Media
6. Philipp Janert, Data Analysis with Open Source Tools, Shroff Publisher Publisher /O'Reilly Publisher Media.