Suggested Teaching Guidelines for

Advanced Analytics using Statistics
PG-DBDA March 2024

Objective: To perform advanced analytics using Python & R skills and important mathematical concepts.

Prerequisites: Good Knowledge of Basic Mathematics

List of Books / Other training materials

Text Book:

1. Business Analytics, James R Evans,Pearson Education, 3rd Edition.

Reference:

1. Beginning R – The Statistical Programming Languageby Dr. Mark Gardener PUB: WILEY
2. Art of Programming in R, by Norman Matloff
3. Statistics for Management by Levin
4. Business Analytics: Methods, Models, and Decisions by James R Evans
5. Introductory Statistics with R (Statistics and Computing) by Peter Dalgaard
6. R in a Nutshell by Joseph Adler (O'REILLY)
7. R Cookbook by Paul Teetor (O'REILLY)
8. The R Book, Second Edition
9. Statistics Using R, Shailaja Deshmukh, Sudha Purohit, Sharad Gore, Pub: Narosa

Session 1 & 2 :

- Introduction to Analytics
- Data analytics Life Cycle
  - Discovery
  - Data preparation
  - Model planning
  - Model building implementation
  - Quality assurance
  - Documentation
  - Management approval
  - Installation
  - Acceptance and operation
  - Intelligent data analysis

Assignment –Lab: Import csv file using R and perform ETL operation using dplyr package.

Suggested Teaching Guidelines for

Advanced Analytics using Statistics
PG-DBDA March 2024

Session 3 & 4

- Sample Spaces and Events
- Joint, Conditional and Marginal Probability
- Bayes' Theorem

Assignment –Lab: Load any dataset, apply Bayes' Theorem and predict the output

Session 5 & 6:

- Random Variable
- Concepts of Correlation
- Covariance
- Outliers

Assignment –Lab: Load any dataset and find out the covariance between two fields and also find the correlation and determine how two fields are correlated. Also handle the outliers in the data.

Session 7 & 8:

- Probability Distribution and Data
  - Continuous distribution – (Uniform, Exponential & Normal)
  - Discrete distribution – (Binomial, Poisson & Geometric distribution)

Assignment –Lab: generate random numbers and check if they are in normal distribution using scipy libraries.

Session 9 & 10 :

- Descriptive Statistical Measures
- Summary Statistics - Central Tendency & Dispersion (Mean, Median, Mode, Quartiles, Percentiles, Range, Interquartile Range, Standard Deviation, Variance, and Coefficient of Variation)

Assignment –Lab: Load any dataset and find out the mean, median mode and other central tendencies of the dataset.

Session 11 & 12 :

- Sampling and Estimation
- Sample & population, Uni-variate and bi-variate sampling, re-sampling
- Central Limit Theorem

Assignment –Lab: Load any dataset and Explore sampling techniques.

Session 13 & 14:

- Statistical Inference Terminology (types of errors, tails of test, confidence intervals etc.)
- Hypothesis Testing
- Parametric Tests: ANOVA, t-test
- Non-parametric Tests- chi-Square, U-Test

Assignment –Lab: Load any dataset and Perform the hypothesis testing on correlated variables.

Session 15 & 16:

- Predictive Modelling (From Correlation to Supervised Segmentation):
  - Identifying Informative Attributes,
  - Segmenting Data by Progressive Attributive,
  - Models,
  - Induction and Prediction,
  - Supervised Segmentation,
  - Visualizing Segmentations,
  - Trees as Set of Rules,
  - Probability Estimation;

Assignment –Lab: Explore predictive modelling techniques.

Session 17:

- Simulation and Risk Analysis
- Optimization, Linear

Assignment –Lab: Explore Monte Carlo simulation.

Session 18 & 19:

- Decision Analytics:
  - Evaluating Classifiers,
  - Analytical Framework,
  - Evaluation,
  - Baseline,
  - Performance and Implications for Investments in Data;

Session 20 & 21:

- Evidence and Probabilities:
  - Explicit Evidence Combination with Bayes Rule,
  - Probabilistic Reasoning;

Session 22:

- Business Strategy:
  - Achieving Competitive Advantages,
  - Sustaining Competitive Advantages

Session 23:

- Factor Analysis,
- Directional Data Analytics,

Assignment –Lab: Download dataset and perform factor analysis on it.