CBCA223 - Data Analysis using Python/ CSET214 - Data Analysis using Python

Course Type - Specialized Core – I L-T-P Format 2-0-4

Credits - 4

Course Type - CORE COURSES L-T-P Format 2-0-4

Credits – 4

COURSE SUMMARY This course combines the advantages of both Python and Data Science. Here, students will learn to apply the ideas of analytics in real-world problems. Moreover, multiple mathematical operations and scientific computing will be taught using existing and mature python-based libraries.

COURSE-SPECIFIC LEARNING OUTCOMES (CO)

CO1: To articulate the structured and unstructured data for extracting useful information.

CO2: To work with data through visualization and distributions.

CO3: To make use of real-world datasets utilizing various numerical libraries.

Detailed Syllabus

Module 1 (Contact hours: 8) Purpose of data analysis, Structured and Unstructured data, Steps of data analysis, Python Packages for Data Analysis: Numpy, Scipy, Matplotlib, Plotyly, NLTK. Data Frames, Usage of frames analytical roles, File handling and reading data for processing, Pre-processing data using multiple python frameworks, Data Formatting, Data Manipulation, Data normalization, Data Merging, Data reshaping, Data Wrangling, Aggregation functions.

Module 2 (Contact hours: 10) String Manipulations, Demonstrating string functions, A regular expression for data manipulation, Data Visualization, Using Histograms, Using Boxplots, Plotting data, Venn Diagram, Bar Chart, Pie Chart, Line Chart, Scatter Plots and R2, Grouped charts, Area Charts, Descriptive Statistics, Central tendencies, Analyzing variability, Data Distributions, Random Variables, Bernoulli Distribution, Binomial Distribution, Normal Distribution, Statistical Properties, Standard Normal Distribution.

Module 3 (Contact hours: 8) Exponential distribution, Statistical test, Hypothesis testing, Z-test, Right-tailed test, Two-tailed test, T-Test, Significance of p-value in t-test, Two-sample Z-test, Paired ttest, Visualizing statistical test analysis, Model building, Outlier Detection.

STUDIO WORK / LABORATORY EXPERIMENTS: Studio work focuses on Different Data Analysis Methods, Techniques, Algorithms using PythonData manipulation using numpy and scipy. Make use of numpy arrays, matrices, indexing and slicing options with the demonstration of numerical packages for data analysis.

TEXTBOOKS/LEARNING RESOURCES: a) Bharti Motwani, Data Analytics using Python (1 ed.), Wiley, 2020. ISBN 8126502959. b) Klosterman and Stephen, Data Science Projects with Python: A Case Study Approach to Successful Data Science Projects Using P (1 ed.), Packt Publishing Limited, 2019. ISBN 978-1838551025.

REFERENCE BOOKS/LEARNING RESOURCES: a) Suresh Kumar Mukhiya and Usman Ahmed, Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data (1 ed.), Packt Publishing Limited, 2020. ISBN 978-1789537253.