<div align="center">

V Semester
Course 14 B: Foundations of Data Science
Credits -3

</div>

Learning Objectives:

To enable students to develop IoT solutions for real-world problems

Learning Outcomes: On successful completion of the course, students will be able to

1. Identify the need for data science and understand various data collection strategies

2. Understand about NoSQL and Descriptive Statistics

3. Apply Numpy methods to process the data in an array.

4. Summarize and Compute Descriptive Statistics using Pandas.

5. Apply powerful data manipulations visualization using Pandas

UNIT-I

Introduction to Data Science: Need for Data Science – What is Data Science - Evolution of Data Science, Data Science Process – Business Intelligence and Data Science – Prerequisites for a Data Scientist – Tools and Skills required. Applications of Data Science in various fields – Data Security Issues.

Data Collection Strategies, Data Pre-Processing Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization, Data Munging, Filtering

UNIT-II

Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis; Box Plots – Pivot Table – Heat Map – Correlation Statistics –ANOVA.

No-SQL: Document Databases, Wide-column Databases and Graphical Databases.

UNIT-III

Python for Data Science –Python Libraries, Python integrated Development Environments (IDE)for Data Science, **NumPy Basics:** Arrays and Vectorized Computation- The NumPy ndarray-

Creating ndarrays- Data Types for ndarrays- Arithmetic with NumPy Arrays- Basic Indexing and Slicing - Boolean Indexing-Transposing Arrays and Swapping Axes.

Universal Functions: Fast Element-Wise Array Functions- Mathematical and Statistical Methods- Sorting- Unique and Other Set Logic.

UNIT-IV

Introduction to pandas Data Structures: Series, Data Frame and Essential Functionality: Dropping Entries- Indexing, Selection, and Filtering- Function Application and Mapping- Sorting and Ranking.
Summarizing and Computing Descriptive Statistics- Unique Values, Value Counts, and Membership. Reading and Writing Data in Text Format.

UNIT-V

Data Cleaning and Preparation: Handling Missing Data - Data Transformation: Removing Duplicates, Transforming Data Using a Function or Mapping, Replacing Values, Detecting and Filtering Outliers-

Plotting with pandas: Line Plots, Bar Plots, Histograms and Density Plots, Scatter or Point Plots.

Text Book(s)

1. Y. Daniel Liang, "Introduction to Programming using Python", Pearson, 2012.
2. Wes McKinney, "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", O'Reilly, 2nd Edition, 2018.

Reference Books

1. Sanjeev Wagh, Manisha Bhende, Anuradha Thakare, 'Fundamentals of Data Science, CRC Press, 1st Edition, 2022
2. Jake VanderPlas, "Python Data Science Handbook: Essential Tools for Working with Data", O'Reilly, 2017.

SUGGESTED CO-CURRICULAR ACTIVITIES & EVALUATION METHODS:

Unit 1: Activity: Seminar on Role of Data Science in Politics

   Evaluation Method: Content knowledge, organization, clarity, presentation skills, visualaids, audience engagement

Unit 2: Activity: Exercises on Descriptive Statistics

   Evaluation Method: Problem Solving, Accuracy

Unit 3: Activity: Hands-on Lab using Numpy

   Evaluation Method: Lab Performance and Correctness of solution Implementation

Unit 4: Activity: Hands-on Lab Activity on Pandas

   Evaluation Method: Lab Performance and Correctness of solution Implementation.

Unit 5: Activity: Group Activity to visualize college performance records using various plots

   Evaluation Method: Active Participation, Post Talk report presentation

<div align="center">

V Semester
Course 14 B : Foundations of Data Science
Credits -1

</div>

List of Experiments:

1. Study on various python IDEs for Data Science

2. Create NumPy arrays from Python Data Structures, Intrinsic NumPy objects and Random Functions.

3. Manipulation of NumPy arrays- Indexing, Slicing, Reshaping, Joining and Splitting.

4. Computation on NumPy arrays using Universal Functions and Mathematical methods.

5. Create Pandas Series and Data Frame from various inputs.

6. Import any CSV file to Pandas Data Frame and perform the following:

   a. Visualize the first and last 10 records

   b. Get the shape, index and column details

   c. Select/Delete the records (rows)/columns based on conditions.

   d. Perform ranking and sorting operations.

   e. Do required statistical operations on the given column

7. Import any CSV file to Pandas Data Frame and perform the following:

a. Handle missing data by detecting and dropping/ filling missing values.

b. Transform data using apply () and map() method.

c. Detect and filter outliers.

d. Perform Vectorized String operations on Pandas Series.

e. Visualize data using Line Plots, Bar Plots, Histograms, Density Plots and Scatter Plots.