DATA SCIENCE WITH PYTHON

About Data Science with Python

- Data science - the process of deriving knowledge and insights from a huge and diverse set of data through organizing, processing and analysing the data.
- It involves many different disciplines like mathematical and statistical modelling, extracting data from it source and applying data visualization techniques.
- It also involves handling big data technologies to gather both structured and unstructured data.
- Python is a very popular general-purpose interpreted, interactive, object-oriented, and high-level programming language

Why Learn Data Science with Python?

- Python is relatively simple and easier to learn.
- Tools and Libraries of Python are used for Data analysing.
- Python packages and libraries like NumPy and Pandas help with data clean up and Analysis.

How to learn Data Science with Python?

- Data Science with Python tutorial from Vcare Technical Institute - We provide step by step Data Science with Python tutorials, examples, and references. Get started with Data Science with Python.

Topics :- **Course Duration – 6 Months**

1. INTRODUCTION TO DATA SCIENCE

I: Introduction

- Introduction to Data Science
- Evolution of Data Science
- Data Science Roles
- Stages in a Data Science Project
- Applications of Data Science in various fields
- Data Security Issues.

II: Data Collection and Data Pre-Processing

- Data Collection Strategies
- Data Pre-Processing Overview
- Data Cleaning
- Data Integration and Transformation
- Data Reduction
- Data Discretization.

III: Exploratory Data Analytics

- Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis

- Box Plots
- Pivot Table
- Heat Map
- Correlation Statistics
- ANOVA.

IV: Model Development

- Simple and Multiple Regression
- Model Evaluation using Visualization
- Residual Plot
- Distribution Plot
- Polynomial Regression and Pipelines
- Measures for In-sample Evaluation
- Prediction and Decision Making.

V: Model Evaluation

- Generalization Error
- Out-of-Sample Evaluation Metrics
- Cross Validation
- Overfitting – Under Fitting and Model Selection
- Prediction by using Ridge Regression
- Testing Multiple Parameters by using Grid Search.

2. STATISTICS FOR DATA SCIENCE

I: Descriptive Statistics

- Sampling Techniques
- Data Classification – Tabulation – Frequency and graphic Representation
- Measures of Central Tendency – Measures of Variation – Quartiles and Percentiles – Moments - Skewness and Kurtosis.

II: Correlation and Regression

- Scatter Diagram
- Karl Pearson's Correlation Coefficient
- Rank Correlation
- Correlation Coefficient for Bivariate Frequency Distribution
- Regression Coefficients
- Fitting of Regression Lines.

III: Probability Theory

- Random Experiment
- Sample Space – Events – Axiomatic Definition of Probability
- Addition Theorem – Multiplication Theorem – Baye's Theorem Applications.

IV: Distribution Function

- Continuous and Discrete Random Variables
- Distribution Function of a Random Variable
- Probability Mass Functions and Probability Density Functions
- Characteristic Functions

- Central Limit Theorems.

V: Probability Distributions

- Probability Distributions
- Recurrence Relationships
- Moment Generating Functions
- Cumulant Generating Functions
- Continuous Probability Distributions
- Rectangular Distribution
- Binomial Distribution
- Poisson Distribution
- Continuous Probability Distributions
- Uniform Distribution - Normal Distribution
- Exponential Distribution.

3. PYTHON FOR DATA SCIENCE

I: Data Structures and OOP

- Python Program Execution Procedure
- Statements
- Expressions
- Flow of Controls
- Functions
- Numeric Data Types
- Sequences
- Strings
- Tuples
- Lists
- Dictionaries
- Class – Constructors – Object Creation – Inheritance – Overloading
- Text Files and Binary Files – Reading and Writing.

II: Numpy and Pandas Packages

- NumPy ndarray
- Vectorization Operation
- Array Indexing and Slicing
- Transposing Array and Swapping Axes
- Saving and Loading Array
- Universal Functions
- Mathematical and Statistical Functions in Numpy
- Series and DataFrame data structures in pandas
- Creation of Data Frames
- Accessing the columns in a DataFrame - Accessing the rows in a DataFrame - Panda's Index Objects - Reindexing Series and DataFrames - Dropping entries from Series and Data Frames - Indexing, Selection and Filtering in Series and Data Frames - Arithmetic Operations between Data Frames and Series
- Function Application and Mapping.

III: Data Wrangling

- Combining and Merging Data Sets
- Reshaping and Pivoting
- Data Transformation
- String manipulations
- Regular Expressions.

IV: Data Aggregation and Group Operations

- GroupBy Mechanics
- Data Aggregation
- GroupWise Operations
- Transformations
- Pivot Tables
- Cross Tabulations
- Date and Time data types.

V: Visualization in Python

- Matplotlib and Seaborn Packages
- Plotting Graph - Controlling Graphs – Adding Text
- More Graph Types – Getting and Setting Values
- Patches.

4. DATABASE FOR DATA SCIENCE

I: SQL

- Introduction to Database
- Statements of Database

II: INTRODUCTION TO RELATIONAL DATABASES & TABLES

- Relational Database Concepts
- Types of SQL statements

III: INTERMEDIATE SQL

- Using string patterns & ranges
- Sets and functions
- Sub-Queries
- Multiple Tables

IV: ACCESSING DATABASE USING PYTHON

- Code using DB-API
- Connecting to a database
- Creating tables, loading data & querying data
- Analysing data with python