ProTech protechtraining.com *"Charting the Course ... ... to Your Success!"*

Data Analysis & Machine Learning Using Python

Course Summary

Description

Python is one of the fastest-growing high-level programming languages which enables clear programs on small and large scales. Widely considered the programming language of choice for serious developers, it is easy to learn and deploy, with design features that emphasize clarity of syntax, easy readability, and easy comprehension. Python can be used for large and small applications - to create web apps, games, or even a search engine. As programming in Python is much simpler than C, C++ and Java, this is the preferred language in many engineering, science and business applications.

This course will introduce the participants to the design methodologies used to address Data Analysis/Machine Learning scenarios that use Python packages such as Scipy, Pandas, Statsmodels and Scikit-learn to build solutions. Participants will learn to manipulate, process, analyze and clean data using powerful libraries and tools. With the introduction of various reallife scenarios, they will learn the practical applications of Data Analysis using Python and how to integrate this modern and dynamic language with Hadoop.

Topics

- Data analysis - Why Python?
- Data Analysis - Application Scenarios
- Design methodologies in Data Analysis solutions
- Overview of available Frameworks for Data Analysis
- Matplotlib
- Scripy and Numpy
- Pandas
- IPython Toolkit
- Scikit-Learn
- Python in Hadoop Ecosystem

Audience

This course is designed for those wanting to learn the design methodologies used to address Data Analysis/Machine Learning scenarios that use Python packages such as Scipy, Pandas, Statsmodels and Scikit-learn to build solutions.

Prerequisites

There are no prerequisites for this course.

Duration

Five days

ProTech protechtraining.com *"Charting the Course ...*

... to Your Success!"

Data Analysis & Machine Learning Using Python

Course Outline

I. Data analysis - Why Python?

II. Data Analysis - Application Scenarios

III. Design methodologies in Data Analysis solutions
- A. Parallel processing.
- B. Data Classification.
- C. Building a kNN classifier.
- D. Reducing dimensions in data.
- E. Feature selection and extraction.
- F. Clustering.
- G. Machine Learning.

IV. Overview of available Frameworks for Data Analysis

V. Matplotlib
- A. Plot, Sub-plot, Figures, Axes, Ticks.
- B. Pylab.
- C. Plot configurations - changing color and line widths, limits, ticks, labels, spines, legends, annotations.
- D. Different kind of plots - scatter, bar, contour, images, pie, grid, polar axis.

VI. Scripy and Numpy
- A. File IO.
- B. Linear Algebra Operations.
- C. Fast Fourier Transforms.
- D. Optimization and Fit.
- E. Statistics and random numbers.
- F. Interpolations.
- G. Signal processing.
- H. Image processing.

VII. Pandas
- A. Data IO.
- B. Pandas Dataframes.
- C. Filtering/Aggregating Dataframes.
- D. Webscraping.
- E. Vectored String Operations.

VIII. IPython Toolkit
- A. Using IPython for interactive work.
- B. IPython notebooks.
- C. IPython for parallel computing.
- D. Configuration and Customization.

IX. Scikit-Learn
- A. Setting and Estimator Objects.
- B. Predictions.
- C. Model Selection.
- D. Training Classifiers.

X. Python in Hadoop Ecosystem
- A. Python in Hadoop MapReduce.
- B. Python with Pig.
- C. Python with Hive.