

DATA SCIENCE

BIG DATA IS THE FOUNDATION OF ALL CURRENT MEGATRENDS, FROM SOCIAL TO MOBILE TO THE CLOUD TO GAMING

! The future belongs to people who can't transform data into strategic business decisions and value-driven products. Data Science is a field of Big Data which seeks to provide meaningful information from large amounts of complex data using various tools, algorithms and machine learning principles.

i Worldwide revenues for Big Data and business analytics will grow from \$130.1 billion in 2016 to more than \$203 billion in 2020, at a Compound Annual Growth Rate (CAGR) of 11.7%.
Source: IDC

Certified Python Programmer for Data Science

Duration: Beginner to Intermediate (5 Days)

Course Overview

Are you planning to become a data scientist? If yes, then you have to learn Python programming language. Why? Python is the number one programming language in the world of data scientists. It emphasises on code readability and clear programming on both small and large scales, allowing you to focus on your research, product, or project.

In this 4-day journey, you will be exposed to multiple development environments so you can choose the best one for you. You will be taught step-by-step how to program in Python. You will go through all the steps of a Data Science project starting from data importing, data cleaning, data analysing, to data visualisation which reveals new insights.

In summary, you will gain a complete understanding of Python with Data Science from the ground up.

Prerequisites

All participants should have a basic knowledge of programming in any language (Java, C, C++, Pascal, Fortran, Javascript, PHP, Python, etc.)

Who Should Attend

This workshop is intended for individuals who are interested in learning Data Science, or who want to begin their career as a data scientist.

Exam Format

The Certified Python for Data Science certification exam duration is 2 hours, consisting of 50 Multiple-Choice Questions, with a passing score of 70%. You will receive a professional Certified Python for Data Science certification upon passing the exam.

Learning Outcomes

Upon completion of this course, you will be able to:

- Recognise the meaning of the terms "Data Science" and "Machine Learning".
- Understand the basics of Python.
- Develop and write code easily in Python.
- Deal easily with files and file systems.
- Deal with different sources of data.
- Analyse and visualise data to gain new insights.

Course Outline

Beginner to Intermediate (5 Days)

Day 1

Introduction to Programming

- What is Algorithm?
- What is Programming?
- The Natural Language of the Computer
- Machine Language
- Programming Language Levels
- Translators

Python Basics

- Identifiers, Lists, and Tuples
- Dictionaries, Sets, Strings, Operators, Control Structures, Loops

Day 2

Jupyter Notebook

- Installing and Running Jupyter
- User Interface
- Checkpoints

Functions

- Functions
- Lambda and Map Functions
- Globals and Locals

Pythonic Programming

- List Comprehension
- Generator Expressions
- Exceptions Handling

Modules and Packages

- Modules
- Documentation
- Packages and Namespaces

Working with Files

- Create, Read, Update, Delete (CRUD) a File

Day 3

Object-Oriented Programming

- OOP in General
- Classes
- Objects
- Constructors
- Instance/Class Data
- Instance/Class Method
- Inheritance

OS Module

- Working with File Systems
- Walking Directory Trees
- Paths
- Filenames
- Directories

Working with Files

- Creating a File
- Reading a File
- Updating a File
- Deleting a File

Working with JSON Data

- What is JSON and Why Is It Important?
- Module, Serialisation and Deserialisation

Web Scraping (BeautifulSoup)

- What is Web Scraping?
- HTML Tags
- BeautifulSoup Module
- Webpage Scraping Phase

Day 4

Introduction to Matrix Processing (NumPy)

- What is NumPy?
- Narray Object, Data Types
- Array Attributes, Array Creation Routines
- Indexing and Slicing
- Array Manipulation
- Mathematical Functions

Data Analysis (Pandas)

- What is Pandas?
- Series
- DataFrame
- Data Importing
- Data Pre-Processing
- Data Grouping

Data Visualisation (Matplotlib)

- What is Matplotlib?
- Line Graphs
- Bar Graphs
- Pie Charts
- Histograms
- Scatter Plots
- Graph Attributes
- Text Annotation

Day 5

Introduction to Applied Machine Learning (Scikit-learn)

- What is Machine Learning?
- Machine Learning Algorithm Types
- Main Steps in Machine Learning Projects
- Introduction to Scikit-learn Module

Capstone Project

Final Evaluation (Exam)

Certified Python Programmer for Data Science

Duration: Advanced (4 Days)

Course Overview

Python is a general-purpose programming language that is becoming more and more popular for analysing datasets and conducting Data Science processes. Companies worldwide are using Python to harvest insights from their data and gain a competitive edge.

Unlike any other Python tutorial, this class will touch on various environments for project development to let you choose your best one. All the steps to construct a Data Science project will be taught starting from data importing, data cleaning, data analysing, to data visualisation which reveals new insights.

In summary, you will gain a complete understanding of Python with Data Science from the ground up.

Learning Outcomes

Upon completion of this course, you will be able to:

- Understand different categories of Machine Learning science.
- Understand different algorithms under Machine Learning science.
- Understand how to build a Machine Learning model.
- Understand how to evaluate and improve Machine Learning models.
- Understand how to read and pre-process text data.
- Apply different methods to extract features from text data.
- Apply Machine Learning algorithms on text datasets.
- Understand and build language models from text datasets.
- Understand the main definitions under network science.
- Apply Graph Theory using Python to analyse the relationships between objects.
- Apply Graph Theory using Python to extract influencers from networks.
- Apply Graph Theory using Python to calculate different nodes similarity measures.

Prerequisites

All participants should have a basic knowledge of programming in any language (Java, C, C++, Pascal, Fortran, Javascript, PHP, Python, etc.)

Who Should Attend

This workshop is intended for individuals who are interested in learning Data Science, or who want to begin their career as a data scientist.

Exam Format

The Certified Python for Data Science certification exam duration is 2 hours, consisting of 50 Multiple-Choice Questions, with a passing score of 70%. You will receive a professional Certified Python for Data Science (Advanced) certification upon passing the exam.

Course Outline

Advanced (4 Days)

Day 1: Applied Machine Learning (ML)

Module 1: Introduction to Machine Learning

- What is Machine Learning?
- Introduction to SK Learn
- Machine Learning Steps

Module 2: Datasets

- What is Dataset?
- Iris Dataset
- Handwritten Digits Dataset
- Dataset Distribution

Module 3: Supervised Learning

- What is Supervised Learning?
- Key Classifiers Algorithms - KNN, SVM, DT
- Performance Metrics and Errors
- Regression

Module 4: Unsupervised Learning

- What is Unsupervised Learning?
- Key Clustering Algorithms - K-Means, Mean Shift
- Principal Component Analysis and Dimensionality Reduction

Module 5: Neural Network

- Introduction to Neural Network
- Multi-Layer Perceptron Classifier
- Hidden Layers
- Activation Function
- Solver

Day 2: Applied Natural Language Processing (NLP)

Module 1: Introduction to NLP

- What is NLP?
- Basic Text Analysis with Python
- Introduction to NLTK

Module 2: Text Analysis with NLTK

- Tokenise Words and Sentences
- Stop Words
- Regular Expression
- Stemming
- Part-of-Speech (POS) Tagging

Module 3: NLTK Corpus

- What is Corpus?
- Popular NLTK Corpus
- Build Your Own Corpus

Module 4: Text Classification

- Text Classification
- NLTK and Scikit-learn
- Save and Load the Model

Day 3: Social Network Analysis (SNA)

Module 1: Why Study Networks and Basics on NetworkX

- Why Networks are Very Important
- Graphs
- Nodes and Edges
- Introduction to NetworkX Module

Module 2: Network Connectivity

- Clustering Coefficient
- Distance Measures
- Connected Component
- Network Robustness

Module 3: Influence Measures and Network Centralisation

- Degree and Closeness Centrality
- Betweenness Centrality
- Hubs and Authorities

Module 4: Applications

- Power Law
- Small World Network
- Link Prediction
- Use Cases

Day 4

Capstone Project

Final Evaluation (Exam)



Testimonials

Hear What Our Students Have To Say



The trainer is really knowledgeable and it's great to learn this course for further career development.

Krystle Tee, UOB

The training's content was very useful and the trainer was very nice and helpful.

Farica Perdana Putri, Universitas Multimedia Nusantara

A big miss if you don't join!

Chan Ken Yew, Kenanga Investment Bank

The course is very good, quite easy to follow and interactive.

Balqis, Technical Data Analyst,
Sarawak Shell Berhad

That the teacher is great and very knowledgeable about the subject and the center has a good environment to study , i would definitely come back for further trainings and certifications.

Babacar Gaye, University of Science
and Technology BeijingZillionquest Sdn Bhd



Companies Who Learned From Us

Trusted by Public, Private and Education Sectors



8 Marina View, Asia Square Tower 1
Level 07-04, Singapore 018960



www.itrainasia.com



info@itrainasia.com