edureka!
Discover Learning

Python Certification Training for Data Science

https://www.edureka.co/data-science-python-certification-course

About Edureka

Edureka is a leading e-learning platform providing live instructor-led interactive online training. We cater to professionals and students across the globe in categories like Big Data & Hadoop, Business Analytics, NoSQL Databases, Java & Mobile Technologies, System Engineering, Project Management and Programming. We have an easy and affordable learning solution that is accessible to millions of learners. With our students spread across countries like the US, India, UK, Canada, Singapore, Australia, Middle East, Brazil and many others, we have built a community of over 1 million learners across the globe.

About Course

Edureka's Data Science with Python course enables you to learn data science concepts from scratch. This Python Course will also help you master important Python programming concepts such as data operations, file operations, object-oriented programming and various Python libraries such as Pandas, Numpy, Matplotlib which are essential for Data Science. Edureka's Python Certification Training course is also a gateway towards your Data Science career.

Python Course Curriculum

Introduction to Python

Learning Objectives: You will get a brief idea of what Python is and touch on the basics.

Topics:

- Overview of Python
- The Companies using Python
- Different Applications where Python is used
- Discuss Python Scripts on UNIX/Windows
- Values, Types, Variables
- Operands and Expressions
- Conditional Statements
- Loops
- Command Line Arguments
- Writing to the screen

Hands On/Demo:

- Creating "Hello World" code
- Variables
- Demonstrating Conditional Statements
- Demonstrating Loops

Skills:

- Fundamentals of Python programming

Deep Dive – Functions, OOPs, Modules, Errors and Exceptions

Learning Objectives: In this Module, you will learn how to create generic python scripts, how to address errors/exceptions in code and finally how to extract/filter content using regex.

Topics:

- Functions
- Function Parameters
- Global Variables
- Variable Scope and Returning Values
- Lambda Functions
- Object-Oriented Concepts
- Standard Libraries
- Modules Used in Python
- The Import Statements
- Module Search Path
- Package Installation Ways
- Errors and Exception Handling
- Handling Multiple Exceptions

Hands On/Demo:

- Functions - Syntax, Arguments, Keyword Arguments, Return Values
- Lambda - Features, Syntax, Options, Compared with the Functions
- Sorting - Sequences, Dictionaries, Limitations of Sorting
- Errors and Exceptions - Types of Issues, Remediation
- Packages and Module - Modules, Import Options, sys Path

Skills:

- Error and Exception management in Python

- Working with functions in Python

Data Manipulation

Learning Objective: Through this Module, you will understand in detail about Data Manipulation

Topics:

- Basic Functionalities of a data object

- Merging of Data objects

- Concatenation of data objects

- Types of Joins on data objects

- Exploring a Dataset

- Analysing a dataset

Hands On/Demo:

- Pandas Function- Ndim(), axes(), values(), head(), tail(), sum(), std(), iteritems(), iterrows(), itertuples()

- GroupBy operations

- Aggregation

- Concatenation

- Merging

- Joining

Skills:

- Python in Data Manipulation

Introduction to Machine Learning with Python

Learning Objectives: In this module, you will learn the concept of Machine Learning and its types.

Topics:

- Python Revision (numpy, Pandas, scikit learn, matplotlib)
- What is Machine Learning?
- Machine Learning Use-Cases
- Machine Learning Process Flow
- Machine Learning Categories
- Linear regression
- Gradient descent

Hands On/Demo:

- Linear Regression – Boston Dataset

Skills:

- Machine Learning concepts
- Machine Learning types
- Linear Regression Implementation

Supervised Learning - I

Learning Objectives: In this module, you will learn Supervised Learning Techniques and their implementation, for example, Decision Trees, Random Forest Classifier etc.

Topics:

- What are Classification and its use cases?

- What is Decision Tree?

- Algorithm for Decision Tree Induction

- Creating a Perfect Decision Tree

- Confusion Matrix

- What is Random Forest?

Hands On/Demo:

- Implementation of Logistic regression

- Decision tree

- Random forest

Skills:

- Supervised Learning concepts

- Implementing different types of Supervised Learning algorithms

- Evaluating model output

Dimensionality Reduction

Learning Objectives: In this module, you will learn about the impact of dimensions within data. You will be taught to perform factor analysis using PCA and compress dimensions. Also, you will be developing LDA model.

Topics:

- Introduction to Dimensionality

- Why Dimensionality Reduction

- PCA

- Factor Analysis

- Scaling dimensional model

- LDA

Hands-On/Demo:

- PCA

- Scaling

Skills:

- Implementing Dimensionality Reduction Technique

Supervised Learning - II

Learning Objectives: In this module, you will learn Supervised Learning Techniques and their implementation, for example, Decision Trees, Random Forest Classifier etc.

Topics:

- What is Naïve Bayes?

- How Naïve Bayes works?

- Implementing Naïve Bayes Classifier

- What is Support Vector Machine?

- Illustrate how Support Vector Machine works?

- Hyperparameter Optimization

- Grid Search vs Random Search

- Implementation of Support Vector Machine for Classification

Hands-On/Demo:

- Implementation of Naïve Bayes, SVM

Skills:

- Supervised Learning concepts

- Implementing different types of Supervised Learning algorithms

- Evaluating model output

Unsupervised Learning

Learning Objectives: In this module, you will learn about Unsupervised Learning and the various types of clustering that can be used to analyze the data.

Topics:

- What is Clustering & its Use Cases?

- What is K-means Clustering?

- How does K-means algorithm work?

- How to do optimal clustering

- What is C-means Clustering?

- What is Hierarchical Clustering?

- How Hierarchical Clustering works?

Hands-On/Demo:

- Implementing K-means Clustering

- Implementing Hierarchical Clustering

Skills:

- Unsupervised Learning

- Implementation of Clustering – various types

Association Rules Mining and Recommendation Systems

Learning Objectives: In this module, you will learn Association rules and their extension towards recommendation engines with Apriori algorithm.

Topics:

- What are Association Rules?

- Association Rule Parameters

- Calculating Association Rule Parameters

- Recommendation Engines

- How does Recommendation Engines work?

- Collaborative Filtering

- Content-Based Filtering

Hands-On/Demo:

- Apriori Algorithm

- Market Basket Analysis

Skills:

- Data Mining using python

- Recommender Systems using python

Reinforcement Learning

Learning Objectives: In this module, you will learn about developing a smart learning algorithm such that the learning becomes more and more accurate as time passes by. You will be able to define an optimal solution for an agent based on agent-environment interaction.

Topics:

- What is Reinforcement Learning

- Why Reinforcement Learning

- Elements of Reinforcement Learning

- Exploration vs Exploitation dilemma

- Epsilon Greedy Algorithm

- Markov Decision Process (MDP)

- Q values and V values

- Q – Learning

- α values

Hands-On/Demo:

- Calculating Reward

- Discounted Reward

- Calculating Optimal quantities

- Implementing Q Learning

- Setting up an Optimal Action

Skills:

- Implement Reinforcement Learning using python

- Developing Q Learning model in python

Time Series Analysis

Learning Objectives: In this module, you will learn about Time Series Analysis to forecast dependent variables based on time. You will be taught different models for time series modeling such that you analyze a real time-dependent data for forecasting.

Topics:

- What is Time Series Analysis?

- Importance of TSA

- Components of TSA

- White Noise

- AR model

- MA model

- ARMA model

- ARIMA model

- Stationarity

- ACF & PACF

Hands on/Demo:

- Checking Stationarity

- Converting a non-stationary data to stationary

- Implementing Dickey-Fuller Test

- Plot ACF and PACF

- Generating the ARIMA plot

- TSA Forecasting

Skills:

- TSA in Python

Model Selection and Boosting

Learning Objectives: In this module, you will learn about selecting one model over another. Also, you will learn about Boosting and its importance in Machine Learning. You will learn on how to convert weaker algorithms into stronger ones.

Topics:

- What is Model Selection?

- The need for Model Selection

- Cross-Validation

- What is Boosting?

- How Boosting Algorithms work?

- Types of Boosting Algorithms

- Adaptive Boosting

Hands on/Demo:

- Cross-Validation

- AdaBoost

Skills:

- Model Selection

- Boosting algorithm using python

Sequences and File Operations

Learning Objectives: Learn different types of sequence structures, related operations and their usage. Also learn diverse ways of opening, reading, and writing to files.

Topics:

- Python files I/O Functions

- Numbers

- Strings and related operations

- Tuples and related operations

- Lists and related operations

- Dictionaries and related operations

- Sets and related operations

Hands On/Demo:

- Tuple - properties, related operations, compared with a list

- List - properties, related operations

- Dictionary - properties, related operations

- Set - properties, related operations

Skills:

- File Operations using Python

- Working with data types of Python

Introduction to NumPy, Pandas and Matplotlib

Learning Objectives: This Module helps you get familiar with basics of statistics, different types of measures and probability distributions, and the supporting libraries in Python that assist in these operations. Also, you will learn in detail about data visualization.

Topics:

- NumPy - arrays

- Operations on arrays

- Indexing slicing and iterating

- Reading and writing arrays on files

- Pandas - data structures & index operations

- Reading and Writing data from Excel/CSV formats into Pandas

- matplotlib library

- Grids, axes, plots

- Markers, colours, fonts and styling

- Types of plots - bar graphs, pie charts, histograms

- Contour plots

Hands On/Demo:

- NumPy library- Creating NumPy array, operations performed on NumPy array

- Pandas library- Creating series and dataframes, Importing and exporting data

- Matplotlib - Using Scatterplot, histogram, bar graph, pie chart to show information, Styling of Plot

Skills:

- Probability Distributions in Python

- Python for Data Visualization

Python Course Project

How will I execute practical's in Edureka's Python Certification Course?

You will do your Assignments/Case Studies using Jupyter Notebook that is already installed on your **Cloud Lab environment** whose access details will be available on your LMS. You will be accessing your Cloud Lab environment from a browser. For any doubt, the 24*7 support team will promptly assist you.

What is CloudLab?

CloudLab is a cloud-based Jupyter Notebook which is pre-installed with Python packages on the cloud-lab environment. It is offered by Edureka as a part of Python Certification Course where you can execute all the in-class demos and work on real-life projects in a fluent manner.

You'll be able to access the CloudLab via your browser which requires minimal hardware configuration. In case, you get stuck in any step, our support ninja team is ready to assist 24x7.

Which case studies will be a part of this Python Certification Course ?

This course comprises of 40 case studies that will enrich your learning experience. In addition, we also have 4 Projects that will enhance your implementation skills. Below are few case studies, which are part of this course:

Case Study 1: Maple Leaves Ltd is a start-up company which makes herbs from different types of plants and its leaves. Currently, the system they use to classify the trees which they import in a batch is quite manual. A laborer from his experience decides the leaf type and subtype of plant family. They have asked us to automate this process and remove any manual intervention from this process.

You have to classify the plant leaves by various classifiers from different metrics of the leaves and to choose the best classifier for future reference.

Case Study 2: BookRent is the largest online and offline book rental chain in India. The company charges a fixed fee per month plus rental per book. So, the company makes more money when user rents more books.

You as an ML expert and must model recommendation engine so that user gets a recommendation of books based on the behavior of similar users. This will ensure that users are renting books based on their individual taste.

The company is still unprofitable and is looking to improve both revenue and profit. Compare the Error using two approaches – User Based Vs Item Based

Case Study 3: Handle missing values and fit a decision tree and compare its accuracy with random forest classifier.

Predict the survival of a horse based on various observed medical conditions. Load the data from 'horses.csv' and observe whether it contains missing values. Replace the missing values by the most frequent value in each column. Fit a decision tree classifier and observe the accuracy. Fit a random forest classifier and observe the accuracy.

Case Study 4: Principal component analysis using scikit learn.

Load the digits dataset from sklearn and write a helper function to plot the image. Fit a logistic regression model and observe the accuracy.

Using scikit learn perform a PCA transformation such that the transformed dataset can explain 95% of the variance in the original dataset. Compare it with a model and also comment on the accuracy. Compute the confusion matrix and count the number of instances that have gone wrong. For each of the wrong sample, plot the digit along with the predicted and original label.

Case Study 5: Read the datafile "letterCG.data" and set all the numerical attributes as features. Split the data in to train and test sets.

Fit a sequence of AdaBoostClassifier with varying number of weak learners ranging from 1 to 16, keeping the max_depth as 1. Plot the accuracy on the test set against the number of weak learners, using decision tree classifier as the base classifier.

Which kind of projects will be a part of this Python Certification Course?

Project #1:

Industry: Social Media

Problem Statement: You as ML expert have to do analysis and modeling to predict the number of shares of an article given the input parameters.

Actions to be performed:

Load the corresponding dataset. Perform data wrangling, visualization of the data and detect the outliers, if any. Use the plotly library in Python to draw useful insights out of data. Perform regression modeling on the dataset as well as decision tree regressor to achieve your Learning Objectives. Also, use scaling processes, PCA along with boosting techniques to optimize your model to the fullest.

Project #2:

Industry: FMCG

Problem Statement: You as an ML expert have to cluster the countries based on various sales data provided to you across years.

Actions to be performed:

You have to apply an unsupervised learning technique like K means or Hierarchical clustering so as to get the final solution. But before that, you have to bring the exports (in tons) of all countries down to the same scale across years. Plus, as this solution needs to be repeatable you will have to do PCA so as to get the principal components which explain the max variance.