Machine Learning Certification Training using Python

Course Curriculum : Your 12 module Learning Plan

https://www.edureka.co/machine-learning-certification-training

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 Machine Learning Certification Training using Python helps you gain expertise in various machine learning algorithms such as regression, clustering, decision trees, random forest, Naïve Bayes and Q-Learning. This Machine Learning using Python Training exposes you to concepts of Statistics, Time Series and different classes of machine learning algorithms like supervised, unsupervised and reinforcement algorithms. Throughout the Data Science Certification Course, you'll be solving real-life case studies on Media, Healthcare, Social Media, Aviation, HR.

Machine Learning Course Curriculum

Introduction to Data Science

Goal: Get an introduction to Data Science in this Module and see how Data Science helps to analyze large and unstructured data with different tools.

Objectives: At the end of this Module, you should be able to:

- Define Data Science
- Discuss the era of Data Science
- Describe the Role of a Data Scientist
- Illustrate the Life cycle of Data Science
- List the Tools used in Data Science
- State what role Big Data and Hadoop, Python, R and Machine Learning play in Data Science

Topics:
- What is Data Science?
- What does Data Science involve?
- Era of Data Science
- Business Intelligence vs Data Science
- Life cycle of Data Science
- Tools of Data Science
- Introduction to Python

Data Extraction, Wrangling, & Visualization

Goal: Discuss the different sources available to extract data, arrange the data in structured form, analyze the data, and represent the data in a graphical format.

Objectives: At the end of this Module, you should be able to:

- Discuss Data Acquisition techniques
- List the different types of Data
- Evaluate Input Data
- Explain the Data Wrangling techniques
- Discuss Data Exploration

Topics:

- Data Analysis Pipeline
- What is Data Extraction
- Types of Data
- Raw and Processed Data
- Data Wrangling
- Exploratory Data Analysis
- Visualization of Data

Hands-On/Demo:

- Loading different types of dataset in Python
- Arranging the data
- Plotting the graphs

Introduction to Machine Learning with Python

Goal: In this module, you will learn the concept of Machine Learning and it's types.

Objective: At the end of this module, you should be able to:

- Essential Python Revision
- Necessary Machine Learning Python libraries

- Define Machine Learning

- Discuss Machine Learning Use cases

- List the categories of Machine Learning

- Illustrate Supervised Learning Algorithms

- Identify and recognize machine learning algorithms around us

- Understand the various elements of machine learning algorithm like parameters, hyper parameters, loss function and optimization.

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:

- Linear Regression – Using Boston Dataset

Supervised Learning - I

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

Objective: At the end of this module, you should be able to:

- Understand What is Supervised Learning?

- Illustrate Logistic Regression

- Define Classification

- Explain different Types of Classifiers such as Decision Tree and Random Forest

Topics:

- What is 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:

- Implementation of Logistic regression, Decision tree, Random forest

Dimensionality Reduction

Goal: In this module you will learn about 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.

Objective: At the end of this module, you should be able to:

- Define the importance of Dimensions

- Explore PCA and its implementation

- Discuss LDA and its implementation

Topics:

- Introduction to Dimensionality

- Why Dimensionality Reduction

- PCA

- Factor Analysis

- Scaling dimensional model

- LDA

Hands On:

- PCA

- Scaling

Supervised Learning - II

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

Objective: At the end of this module, you should be able to:

- Understand What is Naïve Bayes Classifier

- How Naïve Bayes Classifier works?

- Understand Support Vector Machine

- Illustrate How Support Vector Machine works?

- Hyperparameter optimization

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:

- Implementation of Naïve Bayes, SVM

Unsupervised Learning

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

Objective: At the end of this module, you should be able to:

- Define Unsupervised Learning

- Discuss the following Cluster Analysis

o K - means Clustering

o C - means Clustering

o Hierarchical Clustering

Topics:

- What is Clustering & its Use Cases?

- What is K-means Clustering?

- How K-means algorithm works?

- How to do optimal clustering

- What is C-means Clustering?

- What is Hierarchical Clustering?

- How Hierarchical Clustering works?

Hands On:

- Implementing K-means Clustering

- Implementing Hierarchical Clustering

Association Rules Mining and Recommendation Systems

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

Objective: At the end of this module, you should be able to:

- Define Association Rules

- Learn the backend of recommendation engines and develop your own using python

Topics:

- What are Association Rules?

- Association Rule Parameters

- Calculating Association Rule Parameters

- Recommendation Engines

- How Recommendation Engines work?

- Collaborative Filtering

- Content Based Filtering

Hands On:

- Apriori Algorithm

- Market Basket Analysis

Reinforcement Learning

Goal: 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.

Objective: At the end of this module, you should be able to

- Explain the concept of Reinforcement Learning

- Generalize a problem using Reinforcement Learning

- Explain Markov's Decision Process

- Demonstrate Q Learning

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:

- Calculating Reward

- Discounted Reward

- Calculating Optimal quantities

- Implementing Q Learning

- Setting up an Optimal Action

Time Series Analysis

Goal: 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 modelling such that you analyse a real time dependent data for forecasting.

Objective: At the end of this module, you should be able to:

- Explain Time Series Analysis (TSA)
- Discuss the need of TSA
- Describe ARIMA modelling
- Forecast the time series model

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:

- Checking Stationarity
- Converting a non-stationary data to stationary
- Implementing Dickey Fuller Test

- Plot ACF and PACF

- Generating the ARIMA plot

- TSA Forecasting

Model Selection and Boosting

Goal: 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 to stronger ones.

Objective: At the end of this module, you should be able to:

- Discuss Model Selection

- Define Boosting

- Express the need of Boosting

- Explain the working of Boosting algorithm

Topics:

- What is Model Selection?

- Need of Model Selection

- Cross – Validation

- What is Boosting?

- How Boosting Algorithms work?

- Types of Boosting Algorithms

- Adaptive Boosting

Hands on:

- Cross Validation

- AdaBoost

In-Class Project

Goal: In this module, you will learn how to approach and implement a Project end to end, and a Subject Matter Expert will share his experience and insights from the industry to help you kickstart your career in this domain. Finally, we will be having a Q&A and doubt clearing session.

Objectives: At the end of this module, you should be able to:

- How to approach a project

- Hands-On project implementation

- What Industry expects

- Industry insights for the Machine Learning domain

- QA and Doubt Clearing Session

Machine Learning with Python Project

What are the system requirements for our Machine Learning Certification Training using Python?

You don't have to worry about the System Requirements as you will be doing your Practical on a Cloud LAB environment. This environment already contains all the necessary software that will be required to execute your practicals.

How will I execute the practicals?

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.

Which case studies will be a part of this Machine Learning Certification Training using Python?

This course comprises of 34 case studies that will enrich your learning experience. In addition, we also have 3 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. Company charges

a fixed fee per month plus rental per book. So, company makes more money when user rent more books.

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

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 has gone wrong. For each of the wrong sample, plot the digit along with 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 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 Machine Learning Certification Training using Python?

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 goal. 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.