Data Science and Machine Learning with Python

Course Details

Days   5.00

Course Code   QADSMLP

Overview

This five day course is aimed at those who are familiar with data analysis and are interested in learning about how Data Science, Analytics, Machine Learning, and Artificial Intelligence (AI) can be used to yield value from data assets.

This course will be of interest if you are interested in developing your own skills to move from analytics to Data Science, or if you are working with Data Scientists and want to learn more about what's possible.

You will be introduced to key concepts and tools for use in Data Science, including typical Data Science Project lifecycles, potential applications & project pitfalls, relevant aspects of data governance and ethics, roles and responsibilities, Machine Learning and AI model development, exploratory analysis and visualisation, as well as techniques and strategies for model deployment.

Throughout the course you will engage in activities and discussions with one of our Data Science technical specialists. Theoretical modules are complimented with comprehensive practical labs.

Prerequisites

Delegates must be existing Python Data users who have attended;

Python for Data Handling (QADHPYTHON)

or have a similar level of knowledge with NumPy and Pandas.

Additionally, we recommend that delegates have attended;

Introduction to Data Science for Data Professionals (QAIDSDP)

in order to understand key data science, machine learning, and AI governance requirements before developing Machine Learning models.

Target Audience

Members of the audience are required to have a some technical expertise such as table structure, working with tabular data in Python, and simple data analysis.

They may be Mid/Senior Leadership seeking a greater understanding of how to implement Data Science within their organization.

They may come from other technical backgrounds such as Data Analysts, Software Developers, and Data Engineers who either work with Data Scientists or are using this course in their journey towards training as a Data Scientist.

In the latter case, audience members may ask for recommendations for their next steps in training towards becoming Data Scientists. We recommend the following refreshed courses which are due to launch in 2023 and 2024 in this suggested sequence:

   Statistics for Data Analysis in Python

   Time Series and Forecasting with Python

   Maths and Statistics for Data Science with Python

   Practical Big Data Analytics (with Python and Spark)

   Generative AI Essentials

   Fundamentals of Deep Learning with Python (followed by selected Python & NVIDIA training)

Delegates will learn how to:

   Introduction to Data Science & Machine Learning

Course Outline

Introduction to Data Science & Machine Learning

Explain the role of the Data Scientist and the skillset it requires

Describe common application areas of Data Science, and examples of its usage in industry

Outline the Data Science process detailed in the CRISP-DM methodology

Detail the characteristics of problems which Data Science can be used to solve

Define how to evaluate the success of a Data Science Project

Introduction to Python for Data Science

Understand why notebooks are often used in Data Science projects

Use Python and associated libraries to manipulate datasets.

Describe why virtual environments are used

Visualise data using Python

Descriptive & Inferential Statistics with Python

Understand the role that descriptive and inferential statistics play in Data Science

Use measures of central tendency, variation, and correlation to understand data

Use hypothesis tests to establish the significance of effects

Use statistical visualisations to understand data distributions

Describe the role of Exploratory Data Analysis in a Data Science project

Preprocessing Data for Analysis

Appropriately process duplicated data, missing values & outliers

Understand the importance of scaling, encoding, and feature selection

Describe the importance of training, testing & validation sets

Engineer novel features to analyse

Supervised Learning: Regression

Describe regression in the context of machine learning

Build simple and multiple linear regression models

Understand non-linear regression approaches

Evaluate & compare regression models

Supervised Learning: Classification

Describe classification in the context of machine learning

Build simple and multiple logistic regression models for classification

Build Decision Tree & Random forest models for Classification

Evaluate and compare classification models

Model Selection & Evaluation

Understand how to choose the best model for regression and classification problems

Consider tests & baselines that can be used to evaluate model performance & behaviour

Evaluate 'how good is good enough'

Unsupervised Learning

Describe clustering and dimensionality reduction in the context of machine learning

Apply and evaluate KMeans clustering

Apply and evaluate dimensionality reduction techniques

Ethics for Data Scientists

Be aware of the legislation and standards Data Scientists must adhere to

Discuss the importance of legal, ethical, and moral considerations in Data Analytics projects and identify applicable UK

legislation for which employees should receive training

Discuss ethical considerations for data handling

Recognise ethical considerations in examples of machine learning, deep learning, and AI

Deploying Models & Insights

Understand how analytical models can be deployed

Evaluate how best to deploy a given model

Define checks which can be used to prevent model failures

Use Python and associated libraries to deploy a machine learning model

Describe which metrics can be used to monitor deployed machine learning models

Where to Go Next

Understand the role of deep learning in modern Artificial Intelligence

Know which qualifications and professional memberships can benefit data scientists