

CompTIA DataX (DY0-001)

Get hands-on with AI, MLOps, and machine learning engineering in this CompTIA DataX (DY0-001) training for professional data scientists.

Learn core data science skills, including statistics, data modeling, and machine learning. Deploy deep learning models, build pipelines, and apply computer vision in real-world environments. This online DataX training prepares you to support advanced data science workflows and pass the CompTIA DataX certification exam. Train for the future of IT with cutting-edge AI skills.

[CBT Nuggets course material](#) →

WEEK 1

Explore Data Science and Resources for DataX

The Origin of Data Science?

What is Data Science?

Data Science Lifecycle & Machine Learning Approaches

Machine Learning & Data Science Application Examples

CHALLENGE

Assess your Data Science Knowledge Gaps for DataX

Introduction

The Five CompTIA DataX Domains

Domain 1

Domain 2

Section 3.0

Domain 4.0

Domain 5.0

Review Questions with Answers Explained

Explore Data Science Tools and Lifecycles

Introduction

Implementing Best Practices in the Data Science Lifecycle

CRISP-DM (Cross-Industry Standard Process for Data Mining)

TDSP (Team Data Science Process)

SEMMA (Sample, Explore, Modify, Model, Assess)

CHALLENGE

Examine Data Science Code Syntax and Workflows

Introduction

Python Libraries & Dependency Licenses for Data Science

Data Science Directory Structures

Basic Python Syntax

Basic R Syntax

CHALLENGE

WEEK 2

Solution Video

Review Best Practices, Composition, & Requirements

Introduction

DAMA and the DMBOK

Requirement Gathering

Software Composition Analysis

Documentation and Code Quality

API Integration

Challenge ☒

Solution Video (optional)

Explore Change Using Calculus for Data Science

Python Introduction

Derivatives

Chain Rule

Partial Derivatives

Integration

CHALLENGE ☒

Solution Video

Apply Probability & Statistics for Data Science

Introduction

Probability Distributions

Discrete, Continuous, and Cumulative Distributions

Types of Continuous Distributions

Central Limit Theorem, Monte Carlo Simulations

CHALLENGE ☒

Solution

Perform Statistical Testing for Data Science

Introduction

Inferential Statistics

Confidence Intervals and Z-Scores

WEEK 3

Hypothesis Testing

One-Sample, Two-Sample, and Paired t-Tests

ANOVA, Chi-Squared Tests, and Correlation

Challenge ☒

Apply Linear Algebra to Data Science Problems

Introductions

Scalars, Vectors, Matrices, and Tensors

Vector Operations

Dot Product & Vector Distance Metrics

Matrix Operations

Inverse & Solving Linear Systems

CHALLENGE ☒

Examine Key Data Sources for Data Science

Introduction

Generated Data

Records

Sensor Data

Transactions & Experiments

Synthetic Data

Public Data

Challenge ☒

Explore Data Ingestion & Storage for Data Science

Introduction

Data Ingestion

Ingestion Infrastructure

Data Ingestion Pipeline

Orchestration Vs Automation

Structured Storage

Flat Files, Semi-Structured, Unstructured, and Compression

CHALLENGE ☒

Explore Data Analysis & Variables for Data Science

Introduction

Exploratory Data Analysis Basics

Types of Analysis Variables

Analyzing Health Survey Data

Univariate Analysis

Bivariate Analysis

Real-world Scenarios

Challenge

Explore Multivariate Analysis and Quality in DS

Introduction

Univariate & Bivariate Analysis Review (Spaced Retrieval)

Multivariate Analysis

Data Quality

Nonlinearity

Challenge

Explore Data Transformation for Data Science

Introduction

Encoding Categorical Data

Label Encoding

Scaling and Normalization

Standardization

Transformation Functions

Structural Transformations

Feature Extraction

Challenge ☒

SOLUTION

WEEK 4

Augment and Feature Engineer Data for Data Science

Explore the Datasets
Data Augmentation and Enrichment
Ground Truth Labeling
Feature Engineering

WEEK 5

Merging and Combining Data with SQL JOINS
Left, Right, Full, Anti, and Fuzzy Joins (continued SQL JOINS)
Cleaning Data
Interpolation and Extrapolation
Outlier Detection
Challenge ☒

Explore Statistical and Machine Learning Models

Introduction
Regression Models
Classification Models
Modeling Time Series
Survival Analysis
Longitudinal Studies
Causal Inference & Bias-Variance Trade-Off
Loss Functions and Model Evaluations
Challenge ☒

Validate Models and Communicate Data Effectively

Introduction
What is Model Validation?

Performance Metrics
Design Constraints
Data Drift Vs Concepts Drift
Aligning Business Requirements
Benchmarking
Communicating Results to Stakeholders
CHALLENGE ☒

WEEK 6

Analyze Model Deployment and MLOps

Introduction
Containerization
Container Orchestration
Virtualization
On-Premises Deployment
Hybrid & Edge Deployment
Cluster & Cloud Deployment
MLOps
MLOps
MLOps
CHALLENGE ☒

Build a Supervised Learning Regression Model

Introduction
What is Supervised Learning?
What is Linear Regression?
Predicting Disease with Linear Regression
Predicting Disease

Predicting Disease

Linear regression Vs Logistic regression

CHALLENGE ☒

Build a Supervised Learning Classification Model

Introduction

Linear Regression Vs Logistic Regression

Implement Linear Regression Model

Implement Logistic Regression Model

Logistic Regression Breakdown

CHALLENGE ☒|

Solution Video

Explore Quadratic and Linear Discriminant Analysis

Introduction

Linear and Logistic Regression Vs. Discriminant Analysis (DA)

WEEK 7

What is Discriminant Analysis (DA)?

What is Linear Discriminant Analysis (LDA)?

LCA Vs. PCA

What is Quadratic Discriminant Analysis (QDA)?

Checking Normality for LDA

CHALLENGE ☒

Classify Data with the Naive Bayes Algorithm

Introduction

Bayes' Theorem

Naive Bayes

Naive Bayes

Naive Bayes

Naive Bayes

Handling Continuous, High-Dimensional, and Missing Data

CHALLENGE ☒

Explore Decision Trees and Ensemble Methods

Introduction

What is a Decision Tree?

Visualizing Decision Trees

How Do Decision Trees Grow?

Impurity Measures

Growing the Tree

A Practical Decision Tree

What is a Random Forest?

What Are Ensemble Methods?

Challenge ☒

Analyze Core Artificial Neural Network Concepts

Introduction

What is an Artificial Neural Network?

WEEK 8

What is a Perceptron?

Activation Function Breakdown

ANN Components

Different Activation Functions

Training Neural Networks

Challenge

Solution

Explore ANN Training Techniques & Gradient Descent

Introduction

Training Neural Networks

Training a Neural Network

Training a Neural Network

Training Techniques

Challenge

Solution

Apply Neural Network Concepts to Deep Learning

Introduction

Gradient Problems

Optimizers in Deep Learning

What is Deep Learning?

Recurrent Neural Networks (RNNs), LSTMS, and GRUs

The Transformer Architecture

GANs, Diffusion, and Special Learning Techniques

CHALLENGE ☒

Compare PyTorch and TensorFlow for Deep Learning

Introduction

WEEK 9

Compare

TensorFlow Multiclass Classifier

PyTorch Multiclass Classifier

CHALLENGE ☒

Explore Natural Language Processing (NLP) Concepts

Introduction

What is NLP?

What is Text Analysis? (aka Text Mining & Text Analytics)

Text Analysis

Text Analysis

Challenge ☒

Explore Tokenization, Gen AI, and LLMs in NLP

Introduction

What is Tokenization? (preprocessing)

Sentence and Subword Tokenization

What is Language Generation?

What are Large Language Models (LLMs)?

Challenge ☒

Prepare Text for Natural Language Processing

Introduction

Text Preparation and Representation

WEEK 10

Tokenization with N-grams
Stemming and Lemmatization
Removing Stop Words
CHALLENGE ☒

Use Advanced Text Preparation for Machine Learning

Introduction
POS Tagging
Regular Expressions (Regex)
Spelling Normalization
Data Augmentation
Text Representations
Challenge ☒

Apply NLP: One-Hot, BoW, TF-IDF, Word2Vec & GloVe

Introduction
To Vectorize or Not to Vectorize
One-Hot Encoding
Bag of Words, TF-IDF, and TensorFlow Embedding Projector
Word Embeddings with Word2Vec
GloVe
CHALLENGE ☒

Explore Foundations of Optimization

Introduction
What is Optimization?
Optimization in Machine Learning
Decision Variables, Objective Functions, and Constraints

WEEK 11

Feasibility Regions

Constrained and Unconstrained Optimization
Gradient Descent Vs Stochastic Gradient Descent
Challenge ☒

Imagine you are analyzing a food delivery startup. You are tasked with minimizing the loss function for a regression model using gradient descent. A small startup sells two SaaS tiers. In this scenario, you are minimizing a loss function for a regression model using gradient descent.

Compare Linear and Nonlinear Programming Methods

Introduction
Linear Programming Concepts
Linear Programming with Python
Visualize the Feasibility Region
Integer Programming
Linear Vs Nonlinear Programming
Visualize the Feasibility Region
CHALLENGE ☒

You've been hired to consult for a popup food truck startup in San Francisco. They are trying to minimize the cost of their ingredients while maximizing the number of customers they can serve.

Explore Specialized Machine Learning Optimization

Introduction
What is the Goal of ML Optimization?
Visualize Gradient Descent in 1D
Different Flavors of Gradient Descent
Constrained Vs Unconstrained Optimization
Bandit Problems

CHALLENGE ☒

Walkthrough/Solution Video

Apply Computer Vision for Image Understanding

Introduction

Image Preprocessing

Object Detection with Pretrained Models

OCR and Segmentation

Image Extraction

Sensor Fusion

Challenge

Apply Feature Extraction for Image Perception

Introduction

Image Preprocessing Deep Dive

More Geometric Transformations

Contrast, Normalization, Color Space Conversion & Hole Filling

Feature Extraction Deep Dive

Challenge ☒

Identify Knowledge Gaps with the DataX SkillScan

Introduction

Question 1

Question 2

Question 3

Question 4

Question 5

Question 6

Question 7

Question 8

Question 9

Question 10

Question 11

Question 12

Question 13

Question 14

Question 15

Question 16

Question 17

Question 18

Question 19

Question 20

Question 21

Question 22

Question 23

Question 24

Question 25

Question 26

Question 27

Question 28

Question 29

Question 30

Question 31

Question 32

WEEK 12

Question 33

Question 34

Question 35

Question 36