ClearML Data Scientist Certification

- 3 Days
- Lecture and Hands-on Labs

Course Overview

Master machine learning (ML) experimentation and reproducibility with ClearML in this hands-on, scenario-driven course. Designed for data scientists you will learn to log and compare experiments, version datasets and models, and perform exploratory data analysis (EDA) in an existing ClearML environment. The course covers ClearML SDK usage, data-centric workflows, and model interpretability, with labs progressing from basic tracking to advanced experimentation. Culminating in a capstone project building a predictive model (e.g., inventory forecasting), this course ensures collaboration with ML Engineers and Developers. Ideal for data scientists with Python and ML experience, it offers ClearML Data Scientist Certification from Alta3 Research, empowering rapid iteration without operational overhead.

Review this course online at https://www.alta3.com/courses/00-template

Who Should Attend

- Data Scientists
- ML Researchers
- Data Analysts
- AI Specialists

What You'll Learn

- Log and compare ML experiments with auto-logged artifacts in $\leq$15 minutes.
- Version datasets and models for reproducibility and drift handling.
- Perform EDA, feature engineering, and model interpretability using ClearML.
- Collaborate with ML Engineers and Developers for seamless handoffs.
- Optimize models with hyperparameter tuning and visualization.
- Apply governance practices (e.g., model cards) for compliance.

Outline

All Outlines are subject to change dependent on changes in the underlying technologies

Introduction to ClearML for Data Scientists

- Lecture: What is MLOps? Role in Experimentation Workflows
- Lecture: ClearML Overview for Data Scientists
- Lecture: Introduction to ML Experimentation and Reproducibility
- Lecture + Lab: Set Up Python Environment with ClearML SDK
- Lecture + Lab: Connect to ClearML Server and Configure SDK
- Lecture + Lab: Log a Baseline ML Experiment
- Lecture + Lab: Explore ClearML Dashboards for Experiment Metrics

Dataset and Model Versioning

- Lecture: Treating Data as Code: Lineage and Reproducibility
- Lecture: Detecting Data Drift Using External Tools (e.g., Evidently)
- Lecture: Model Versioning and Formats
- Lecture + Lab: Version a Dataset
- Lecture + Lab: Detect Drift and Log Reports
- Lecture + Lab: Register and Query a Model Artifact

Collaboration and Governance

- Lecture: Collaborating with ML Engineers and Developers
- Lecture: ClearML Projects, Sharing, and Visibility
- Lecture: Documentation Strategies with Model Metadata
- Lecture + Lab: Share a Project with Simulated ML Engineer
- Lecture + Lab: Upload Model Documentation as Metadata
- Lecture + Lab: Troubleshoot Experiment Failures

Advanced Experimentation

- Lecture: Hyperparameter Tuning with ClearML and Optuna Wrapper
- Lecture: Model Interpretability Workflows
- Lecture: Visualizing Metrics in ClearML Dashboards
- Lecture + Lab: Tune Hyperparameters for a Neural Network
- Lecture + Lab: Log SHAP Plots for Model Insights
- Lecture + Lab: Compare Experiment Variants in Dashboard

Data-Centric ML

- Lecture: Importance of Data Quality in ML
- Lecture: Feature Engineering and Preprocessing Techniques
- Lecture: EDA Workflows with ClearML Auto-Logging
- Lecture + Lab: Preprocess a Messy Dataset
- Lecture + Lab: Engineer Features for Improved Model Accuracy
- Lecture + Lab: Track EDA Visualizations with Auto-Logger

Capstone: Build a Real-World AI Model

- Lecture: Capstone Overview: Inventory Forecasting Model
- Lecture: Simulating Handoff to ML Engineer
- Lecture: Certification Prep: Scenarios and Best Practices
- Lecture + Lab: Build and Log Forecasting Model
- Lecture + Lab: Detect Simulated Data Drift and Alert
- Lecture + Lab: Prepare Model for Production Handoff
- Lecture + Lab: Practice Certification Exam Tasks

Prerequisites

- Python - PCEP Certification or Equivalent Experience
- Familiarity with ML frameworks (e.g., PyTorch, scikit-learn) and basic statistics
- Basic Linux command-line skills