Parter Academy Training Courses - Machine Learning Track

Step -1: Introductory Training

Course	Topics	Partner Academy Links
Databricks Lakehouse Fundamentals	• What is the Databricks Lakehouse Platform? • What is Databricks SQL? • What is Databricks Data Science and Data Engineering Workspace? • What is Databricks Machine Learning? • Fundamentals of the Databricks Lakehouse Platform Accreditation	Databricks Lakehouse Fundamentals Learning Plan

Step-2: Data Science Topics - Link

Courses related to data science that are not part of any role-based learning path.

Course	Topics	Partner Academy Links
Databricks with R	• Identify core features of Spark and Databricks. • Describe how DataFrames are created and evaluated in Spark. • Apply the DataFrame transformation API to process and analyze data. **Prerequisites** • Beginning experience working with R.	Databricks with R
Introduction to Hyperparameter Optimization	• Explain common machine learning techniques that are used to optimize machine learning models for unseen data. • Apply machine learning techniques to improve the fit of machine learning models. • Apply machine learning techniques to improve the generalization of machine learning models. **Prerequisites** • Intermediate level experience with machine learning (ex. feature engineering, feature selection, applying-tree-based models) ◦ We recommend taking the following courses prior to taking this course: Fundamentals of Machine Learning, Introduction to Feature Engineering and Selection with Databricks, Introduction to Applied Tree-based Models with Databricks.	Introduction to Hyperparameter Optimization
Introduction to Natural Language Processing	• Describe foundational concepts about how latent semantic analysis is used to analyze text data. • Perform latent semantic analysis using the Databricks Machine Learning Runtime with the Databricks Workspace. • Generate TFIDF vectors to reduce the noise in a dataset being used for latent semantic analysis in a Databricks Workspace. **Prerequisites** • Intermediate experience performing machine learning/data science workflows • Intermediate experience using the Databricks Data Science Workspace to perform machine learning workflows	Introduction to Natural Language Processing
Natural Language Processing with Databricks	• Explain the motivation behind using Natural Language Processing to analyze data. • Identify distributed Natural Language Processing libraries commonly used when analyzing data. • Perform a series of Natural Language Processing workflows in the Databricks Data Science Workspace **Prerequisites** • Experience working with PySpark DataFrames • Mastery of concepts presented in the Databricks Academy "Apache Spark Programming" course • Mastery of concepts presented in the Databricks Academy "Scalable Machine Learning with Apache Spark" course	Natural Language Processing with Databricks
New Capabilities Overview: Data Profiles in Databricks Notebooks	Exploratory data analysis is a key part of the repeating cycle of exploration, development, and validation that makes up data asset development	New Capabilities Overview: Data Profiles in Databricks Notebooks

How to Tune Models with Hyperopt and Apache Spark	**Prerequisites** • using Python for basic data wrangling processes, • writing simple Python functions and context managers, • how to apply machine learning models with Scikit-Learn and/or Spark MLlib, • why and how we apply various resampling procedures (i.e. train-validation-test sets vs. k-fold cross-validation), • why hyperparameter tuning is important.	How to Tune Models with Hyperopt and Apache Spark

Step-3: Data Engineering with Databricks

Course	Topics	Partner Academy Links
1.0B - Data Science and Engineering Workspace	Leverage the Databricks Lakehouse Platform to perform core responsibilities for data pipeline development • Use SQL and Python to write production data pipelines to extract, transform, and load data into tables and views in the Lakehouse • Simplify data ingestion and incremental change propagation using Databricks-native features and syntax, including Delta Live Tables • Orchestrate production pipelines to deliver fresh results for ad-hoc analytics and dashboarding	1.0B - Data Science and Engineering Workspace

Step-4: Getting Started with Databricks Data Science & Data Engineering Workspace

Course	Topics	Partner Academy Links
Getting Started with Databricks Data Science & Data Engineering Workspace	• Describe the Databricks architecture and the services it provides. • Navigate the Databricks Data Science and Engineering Workspace. • Create and manage Databricks clusters for running code. • Manage data using the Databricks File System and Delta Lake. • Create and run Databricks Notebooks. • Schedule non-interactive execution of Databricks Notebooks using Databricks Jobs. • Integrate a hosted Git service for revision control using Databricks Repos.	Getting Started with Databricks Data Science & Data Engineering Workspace

Step-5: Machine Learning at Databricks

Course	Topics	Partner Academy Links
2-02 Machine Learning at Databricks	• Platform, security, and technology mapping • Data and metadata migration • Code and workload migration • Recommended partner and 3rd party tools • Best practices	Machine Learning at Databricks

Step-6: Scalable Machine Learning with Apache Spark

Course	Topics	Partner Academy Links
Scalable Machine Learning with Apache Spark	• Create data processing pipelines with Spark. • Build and tune machine learning models with Spark ML. • Track, version, and deploy models with MLflow. • Perform distributed hyperparameter tuning with Hyperopt. • Use Spark to scale the inference of single-node models.	Scalable Machine Learning with Apache Spark

Step-7: Machine Learning in Production

Course	Topics	Partner Academy Links
Machine Learning in Production	• Machine learning engineers, Data Engineers & data scientists learn best practices for managing complete machine learning models/architectures on production	Machine Learning in Production

Step-8: Certification Overview: Databricks Certified Machine Learning Associate Exam

Course	Topics	Partner Academy Links
Certification Overview: Databricks Certified Machine Learning Associate Exam	• Describe the learning context, format, and structure behind the exam. • Describe the topics covered in the exam. • Recognize the different types of questions provided on the exam. • Identify resources that can be used to learn the material covered in the exam.	Certification Overview: Databricks Certified Machine Learning Associate Exam