Databricks Certified Machine Learning Professional Training

Curriculum 3 Days

Day - 1

- **Overview of Databricks and Machine Learning on Databricks**
  - Introduction to Databricks: Platform overview, key components, and features
  - Machine learning on Databricks: Databricks runtime for ML, Spark MLlib, and MLflow
  - Setting up the environment for machine learning workflows
  - Hands-on Lab: Set up a Databricks workspace for ML experiments.
- **Deep Dive into Supervised Learning**
  - Advanced Regression Techniques: Lasso, Ridge, ElasticNet, and Polynomial Regression
  - Classification Algorithms: Logistic Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN)
  - Ensemble Learning: Random Forests, Gradient Boosting Machines (GBM), and XGBoost
  - Hands-on Lab: Building and tuning classification and regression models with XGBoost.
- **Feature Engineering for Machine Learning**
  - Feature selection techniques: Univariate selection, recursive feature elimination, and principal component analysis (PCA)
  - Feature transformation: Normalization, Standardization, Log Transform, and Polynomial Features
  - Handling categorical data: One-Hot Encoding, Label Encoding
  - Hands-on Lab: Implementing feature engineering on a real dataset (e.g., customer segmentation).
- **Hyperparameter Tuning & Model Optimization**
  - Introduction to Hyperparameter Tuning: GridSearchCV, RandomSearchCV, and Bayesian optimization
  - Tuning XGBoost and Random Forests: Best practices for performance optimization
  - Techniques for preventing overfitting: Cross-validation, Regularization
  - Hands-on Lab: Hyperparameter tuning of classification models (e.g., XGBoost, Random Forest)

Day - 2

- **Clustering and Dimensionality Reduction**
  - Introduction to Clustering: K-Means, DBSCAN, Agglomerative Clustering
  - Dimensionality Reduction Techniques: Principal Component Analysis (PCA) and t-SNE
  - Model evaluation in clustering: Silhouette Score and Inertia
  - Hands-on Lab: Implementing K-Means and PCA for dimensionality reduction in a dataset.
- **Advanced Deep Learning Techniques**
  - Introduction to Deep Learning: Neural Networks basics
  - Convolutional Neural Networks (CNNs): Use cases in image classification
  - Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM): Use cases in time-series analysis
  - Hands-on Lab: Implementing a basic CNN and RNN for a classification task using Keras in Databricks.
- **Model Evaluation and Interpretation**
  - Advanced Model Evaluation Metrics: Precision-Recall curve, ROC-AUC, Log-Loss
  - Model Interpretability: SHAP values, LIME for model explainability
  - Feature Importance: Understanding and visualizing feature contribution to predictions
  - Hands-on Lab: Evaluating and interpreting a classification model using SHAP and LIME.
- **Deploying Machine Learning Models on Databricks**
  - Introduction to Model Deployment: Benefits of model serving and deployment at scale
  - Deploying models using MLflow: Model packaging and serving REST APIs
  - Scaling models using Databricks Model Serving
  - Hands-on Lab: Deploying a trained model as a REST API using MLflow and Databricks Model Serving

Day - 3

- **MLflow and Managing Machine Learning Lifecycles**
  - Introduction to MLflow: Managing the complete ML lifecycle—tracking, model versioning, and artifact storage
  - Experiment Tracking: Logging parameters, metrics, and artifacts
  - Model Registry: Registering, versioning, and managing models in MLflow
  - Hands-on Lab: Implementing MLflow to track machine learning experiments and register models.
- **AutoML with Databricks**
  - Introduction to AutoML: What is AutoML, and how it accelerates model development
  - AutoML in Databricks: AutoML for regression and classification tasks
  - Hands-on Lab: Running AutoML for a classification problem in Databricks
- **Building End-to-End ML Pipelines in Databricks**
  - Overview of an end-to-end ML pipeline: Data preprocessing, model training, hyperparameter tuning, and evaluation
  - Automating ML pipelines with Databricks Workflows
  - Integrating pipelines with Databricks Jobs and triggering automated workflows
  - Hands-on Lab: Building an end-to-end machine learning pipeline for batch processing data.
- **Exam Preparation and Review**
  - Exam structure: Review of the certification exam topics and question types
  - Time management strategies for the exam
  - Hands-on mock exam: 45 minutes of simulated exam questions
  - Answer key review and discussions
  - Tips for continued learning and certification readiness.