latabricks

Machine Learning Package

Summary

The potential for machine learning (ML) and deep learning practitioners to make breakthroughs and drive positive outcomes is unprecedented. But how do you take advantage of the myriad data and ML tools now available? What's the most efficient way to streamline processes, speed up discovery and scale implementations for real-life scenarios?

Based on your use case and where you are in your data science journey, we can help you with the following:

- 1. Migrate an existing ML model from central/parallel computing models to distributed modeling on Apache Spark,[™] enabling data discovery, model training and inference at scale
- 2. Build ML solutions to address business problems by
 - Translating business problems to ML solutions
 - Performing exploratory data analysis and feature engineering
- 3. Optimize machine learning pipelines
- 4. Perform advisory role to productionize machine learning models

Key outcomes

- Build a reference implementation for one ML pipeline jointly determined by the customer and Databricks
- Apply machine learning best practices at scale
- Optimize machine learning pipelines
- Extend options for ML pipeline automation and MLOps by leveraging MLflow for experiment tracking, reproducibility and rollbacks

Strategy

The package offers two tiers: ML Model MVP and MLOps Optimized. The milestones and outcomes for each tier are produced by our prescriptive methodology, and each tier can be chained to have a greater impact on bolstering your enterprise ML initiatives and adoption.

See the **Resources and schedule** section for details.

DATA LAKE MODEL LIFECYCLE MANAGEMENT mlflow Reproducible Experiments (\mathbf{r}) PKĽ Serialized Models MLflow Tracking Server Feature Store Streaming + Batch Jobs EXPLORATORY ANALYSIS MODEL MODEL DEPLOYMENT O PyTorch Notebooks Streaming + Batch Jobs K Keras anding / Ra Tables Refined Tables Enriched Tables R Feature Eng SQL docker DATABRICKS RUNTIME 🕺 NVIDIA Spark Single Node Data Science GPU Access DATABRICKS FILE SYSTEM / S3 / BLOB

- Building ML workflows customized to your needs
- Enabling internal team with regard to data science and ML

😂 databricks

Challenges of building and deploying ML models

- Diversity and number of ML tools
- Transitioning from experiments to production without rewriting the code
- Data and model drift
- Translating business problems to ML problem statements
- Fully understanding ML limitations and where ML can be applied

Key benefits

- Data discovery and machine learning at scale
- Optimized machine learning pipelines
- Best machine learning practices
- Increased data science productivity
- Repeatable ML

Μ

Databricks ML pipeline workflow

ML PROBLEM Formulation	FEATURE ENGINEERING	BUILD MODEL	MODEL SERVING PRODUCTIONIZE	AUTOMATE AND SCALE ML OPERATION
 Understand business problem 	 Exploratory data analysis 	 Build and train Deployment Model 	 Deploy: Automated batch 	 MLflow: Experiment tracking and
 Translating the problem to ML solution 	 Data wrangling and featurization 	 Hyperparameter tuning 	or streaming jobs Productionize models at scale	reproducibilityML pipeline automation/MLOps

Resources and schedule

L Model MVP — 3 weeks	MLOps Optimized — 3 weeks
ck one option:	ML pipeline automation/MLOps
Reference implementation of one ML pipeline Migration of one ML pipeline from single node to Spark	 Machine learning best practices Optimize machine learning pipeline Repeatable ML with MLflow and Delta Productionize machine learning models

Out of scope

- Configuration and integration of non-Databricks products and systems
- Data cleansing associated with building broader data lake
- ETL unrelated to ML

1 data scientist supporting the activity over a 3-week sprint Prior to kickoff, be sure to review the readiness checklist and complete required tasks