Summary

Machine learning has seen rapid growth and adoption across industries. According to MIT Sloan Management Review, 83% of CEOs report that artificial intelligence (AI) is a strategic priority. As ML becomes ubiquitous and the variety of ML tools increases, organizations face the challenge of integrating and streamlining tools, as well as developing and deploying ML applications, all while ensuring operational rigor throughout.

Based on your use case, we meet you where you are by:

- Building new ML solutions that involve:
  - Translating your business problems to practical ML solutions
  - Performing exploratory data analysis and feature engineering
  - Implementing scalable ML pipelines
  - Incorporating MLflow Tracking and MLflow Model Registry for reproducibility
  - Enabling internal teams
- Optimizing existing ML pipelines, allowing model training and inference at scale
- Productionizing and deploying ML models using robust MLOps practices

Overview

The package offers two tiers: ML Model MVP and MLOps Optimized. Milestones for each tier are produced by our prescriptive methodology, and each tier can be chained for greater impact in bolstering your enterprise ML initiatives and adoption. See Resources and schedule section for details.

Challenges building and deploying ML models

- Translating business problems to ML problems
- Integrating and maintaining diversity of ML tools
- Time from experimentation to production
- Versioning and promoting models
- Data discovery and scalable ML
- Optimized and reproducible ML pipelines
- Rigorous ML practices
- Increased ML practitioners’ productivity, in addition to upskilling team members
- Configuration and integration of non-Databricks products
- Data cleansing associated with building broader data lake
- ETL nonrelated to ML
Databricks ML pipeline workflow

**ML PROBLEM FORMULATION**
- Translate from business problems to ML solutions

**DATA PREPARATION**
- Data wrangling and feature engineering built on Delta Lake

**MODEL TRAINING AND EVALUATION**
- Build and tune models, incorporating MLflow for tracking and reproducibility

**DEPLOYMENT**
- Deployment and CI/CD with MLOps best practices

**Resources and schedule**

**ML MODEL MVP**
- Reference implementation of one ML pipeline

**MLOPS OPTIMIZED**
- Production-ready ML pipeline, incorporating MLOps best practices

*Up to 15 person-days, typically spread over 3-4 weeks

Key outcomes
- A reference implementation for one scalable ML pipeline jointly determined by customer and Databricks
- Optimized and rigorous ML pipelines that enable reliable, reproducible deployment