An MLOps Approach to Address the Complexities of Delivering an ML/AI Product

Pamela Perez[1,2], Shanna Sampson[1,2], Walter Wolf[2]

[1]GAMA-1 Technologies, College Park, MD, USA 20740, [2]NOAA/NESDIS/STAR, College Park, MD, USA 20740

Background

- **Accelerate the Transition of AI Research to Applications**: One of five goals reported in *The NOAA Artificial Intelligence Strategy* released February 2020.
- Machine learning (ML) solutions are complex and operationalizing them presents challenges not addressed in traditional software deployment.
- Unlike traditional software, ML models are automatically created from training data. Thus the data is part of the application and the rules are often not explainable. (Figure 1)
- Operationalizing ML products includes a multitude of additional activities from data collection and cleaning, feature engineering and hyperparameter tuning, all of the way to deployment and monitoring.
- The objective of this presentation is to introduce an MLOps approach for orchestrating the many components of ML based products to provide more robust, accurate, and rapidly available ML applications.

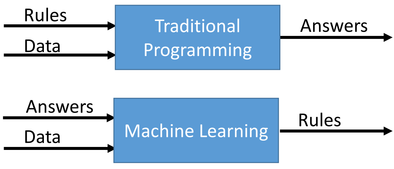
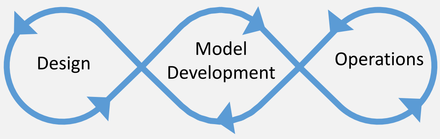


Figure 1. Difference in traditional programming and machine learning development process

Addressing these complexities to get ML models into operations in the shortest time possible while reducing risk

Machine Learning to Operations (MLOps)

MLOps is a set of best practices for the management of the ML Model life cycle.



- Iterative-incremental end-to-end ML workflow
- Process can be divided into three phases.
- All three phases iterate independently yet are interconnected and influence each other.

From "MLOps Principles"
https://ml-ops.org/content/mlops-principles

Design:
- Requirements
- Data exploration

Model Development:
- Data engineering
- ML Model Engineering
- Model Testing and Validation

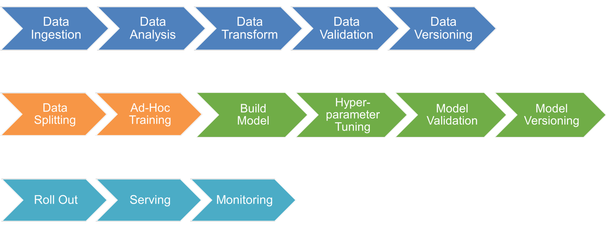
Operations:
- ML Model Deployment
- CI/CD Pipelines
- Monitoring and Triggering

Apply MLOps principles to expedite the transition of ML research to application, improve reliability and reduce rosts.

MLOps Principles

- ❑ Automation
- ❑ Versioning
  - Data Models
  - Pipelines
  - Feature Stores
  - Metadata
- ❑ Testing
  - Data validation
  - Feature
  - Reliable model development
  - ML infrastructure – integration
  - Canary, stress testing, algorithm correctness
- ❑ Reproducibility
- ❑ Experiment Tracking
- ❑ Monitoring
  - Computational performance
  - Data drift
  - Numerical stability of models
  - Degradation of predictions
- ❑ Managing infrastructure
- ❑ Security

Pipeline Examples



The power of an MLOps platform comes from the ability to automate and track the orchestration of machine learning solutions through pipelines. Pipelines can be run independently or fed into other pipelines. They provide modularity that decouples tasks and allows for improvements in smaller more manageable incremental tasks. The ability to track processes through versioning provides repeatability and knowledge transfer.

Current MLOps Landscape

- MLOps is a relatively new discipline
- End-to-end workflows are not here yet
- Many platforms are emerging but each has its own shortcomings
- Data pipelines mostly separate from rest of workflow

Amazon SageMaker
Kubeflow
ALGORITHMIA
data iku
TensorFlow
DataRobot
Valohai
databricks
mlflow
Azure
iguazio
tecton