MLOps platforms to address the complexities of delivering a 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 the five goals reported in *The NOAA Artificial Intelligence Strategy* released in February 2020.
- Machine learning solutions are complex and operationalizing them presents challenges not addressed in traditional software deployment.
- The seminal paper "Hidden Technical Debt in Machine Learning Systems." explains ML models are only a small component of an ML solution. (Figure 1)
- Unlike traditional software, ML models are automatically created from training data. Thus the data is port of the application and the rules are often not explainable. (Figure 2)

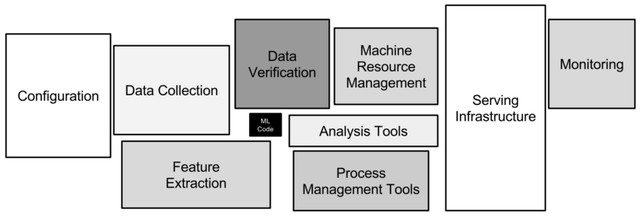
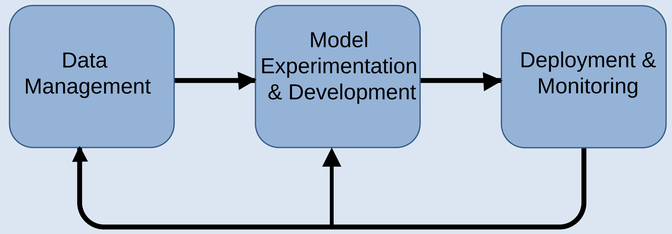


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

From: Sculley, D. et al., "Hidden Technical Debt in Machine Learning Systems." NIPS 2015



The components from Figure 1 can be combined into three broad categories described in detail below.

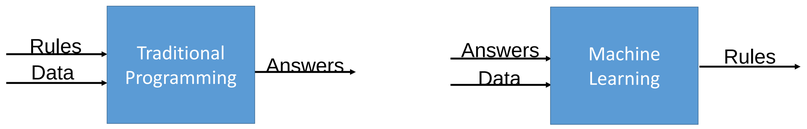


Figure 2. Difference in traditional programming and machine learning development process

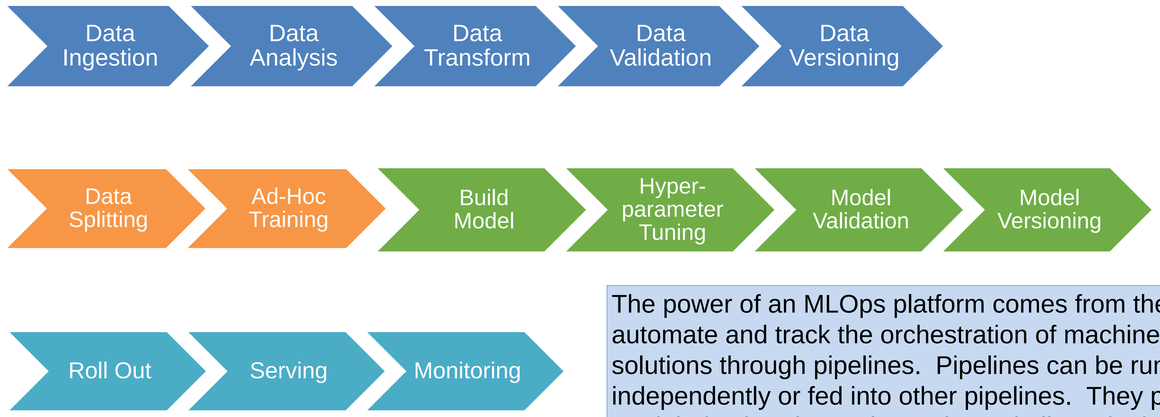
	Data Management	Model Experimentation & Development	Deployment and Monitoring
Steps	• Data ingestion • Data preparation & Transformation • Data exploration & Visualization	• Feature engineering • Model selection • Training, testing and evaluating	Serving Monitoring for accuracy
Issues	• Large data sets, possible multiple sources. • Storing or moving • Data quality • Security • Compliance	• Training and troubleshooting • Tracking experiments • Code quality – not written by software engineers • Model accuracy • Infrastructure	• Online vs. Offline prediction • Monitoring for performance degradation or drift Explainability • Ethics

Apply MLOps Principles to Expedite the Transition of ML Research to Application, Improve Reliability and Reduce Costs.

MLOps is the set of best practices for the management of model life cycle.
MLOps platforms provide a framework and tools to facilitate MLOps practices.

- Automation through pipelines
- Versioning for
  - Data - may reside on multiple systems, may not be immutable, ownership factors
  - Models - can be retrained with new data or training approaches, deployed in new applications, may be subject to attack and require revisions
  - Pipelines
  - Feature Stores
  - Metadata
- Testing of
  - Features and Data - identify features most significant in prediction, data validation
  - Reliable model development
  - ML infrastructure - reproducibility, canary, stress testing, algorithm correctness, integration
- Experiment Tracking
- Monitoring for
  - 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.