AI and Machine Learning Automation with Activeeon: from Models to MLOps
Andrews Cordolino Sobral, AI Solution Architect
Agenda

- Typical data science project (life cycle, roadmap, challenges, etc)
- What an enterprise level data science project needs?
- The evolution of the MLOps solutions
- Proactive Machine Learning (PML) at the core of data science projects
- Bringing up your data science project maturity level with PML
- Data science project pipelines with PML
- MLOps stack with PML
Data science project life cycle

Data Ingestion
- Data movement

Data Preparation
- Normalization
- Transformation
- Validation
- Featureization

Model Training
- Hyper-parameter tuning
- Automatic model selection
- Model testing
- Model validation

Model Deployment
- Deployment
- Batch scoring
- Context capture

Operationalization
- Instrumentation
- Monitoring
- Alerting

Data Engineers

Data Scientists

AI Architects
Data science project life cycle

Data Ingestion
- Data movement
- Normalization
- Transformation
- Validation
- Featureization

Data Preparation
- Hyper-parameter tuning
- Automatic model selection
- Model testing
- Model validation

Model Training
- Deployment
- Batch scoring
- Context capture

Model Deployment
- Instrumentation
- Monitoring
- Alerting

Operationalization

Automated ML Pipeline

IS NEEDED!
Data science common roadmap

NEED TO BE AUTOMATED!

Develop/Experiment

Package
- Dependencies
- Parameters
- Run scripts
- Build

Scale-out
- Load-balance
- Data partitions
- Model distribution
- Hyper params

Tune
- Parallelism
- GPU support
- Query tuning
- Caching

Instrument
- Monitoring
- Logging
- Versioning
- Security

Automate
- CI/CD
- Workflows
- Rolling upgrades
- A/B testing

Weeks with one data scientist

Months with a large team of developers, scientists, data engineers and DevOps

ml-ops-challenges-solutions-and-future-trends
Data science is... complex!

How do we find the AI pipeline that provide the best AI model?
Data science is... complex!
Data science is... complex!

Data Scientists, ML Engineers, ...

IT Specialists, System Administrators, ...

How do we bridge gaps between teams?

Programmers, Web developers, ...
Data science is... complex!

Hardware Abstraction Layer is needed!
Data science needs…

A Modern Automation Platform

A Modern Methodology
Data science needs…

A Modern Automation Platform

A Modern Methodology
**MLOps** is the DevOps for data science projects

- **MLOps** aims to unify the release cycle for machine learning and software application release.
- **MLOps** enables automated testing of machine learning artifacts (e.g. data validation, ML model testing, and ML model integration testing)
- **MLOps** enables the application of agile principles to machine learning projects.
- **MLOps** enables supporting machine learning models and datasets to build these models as first-class citizens within CI/CD systems.
- **MLOps** reduces technical debt across machine learning models.
- **MLOps** must be a language-, framework-, platform-, and infrastructure-agnostic practice.

ml-ops.org
Data science needs...
The Evolution of #MLOps

Proprietary Inference Servers

Using proprietary tools to perform modeling and inference

- SAS
- SPSS
- FICO

The Rise of Open Source Data Science Tools

...attempt to wrap the data science stack in a lightweight web service framework, and put it into production

Python:
- SciPy stack
- scikit-learn
- TensorFlow etc.

R:
- dplyr
- ggplot2
- etc.
- Spark, H2O, others...

Containerization to-the-rescue

Containerization of the “Stone Age” approach, making it easy to scale, robust, etc.

“MLOps Platforms”

- Dockerized open-source ML stacks
- Deployed them on-premise or in the cloud via Kubernetes
- And providing some manageability (“MLOps”).

Pre-History Age

Sketch: @visenger

Stone Age

Bronze Age

2000

2015

2018

MLOps

Gold Rush Age

Source: bit.ly/mlops-evolution
Proactive Machine Learning helps you to:

- Automate the complete Data Science lifecycle,
- Remove silos by creating a bridge between teams: Data, ML and Operations,
- Use your favorite tools by integrating a vast array of solutions,
- Use your infrastructure by linking heterogeneous working environments,

At last, you can industrialize your data science lifecycle project and scale-up seamlessly when needed
PML at the core of data science world
Data science project maturity level

Stage 1  **Manual process.** This is a typical data science process, which is performed at the beginning of implementing ML. This level has an experimental and iterative nature. Every step in each pipeline, such as data preparation and validation, model training and testing, are executed manually. The common way to process is to use Rapid Application Development (RAD) tools, such as Jupyter Notebooks.

Stage 2  **ML pipeline automation.** The next level includes the execution of model training automatically. We introduce here the continuous training of the model. Whenever new data is available, the process of model retraining is triggered. This level of automation also includes data and model validation steps.

Stage 3  **CI/CD pipeline automation.** In the final stage, we introduce a CI/CD system to perform fast and reliable ML model deployments in production. The core difference from the previous step is that we now automatically build, test, and deploy the Data, ML Model, and the ML training pipeline components.

More than 80% of the data science projects stays here!

Less than 20% of the data science projects comes here!

Less than 5% of the data science projects comes here!

ml-ops.org
Data science project pipeline

MLOps

Automated pipeline for model deployment
Thank you for watching
ProActive Machine Learning (PML)

Accelerate the development and deployment of AI models with scalability

www.activeeon.com/products/proactive-machine-learning