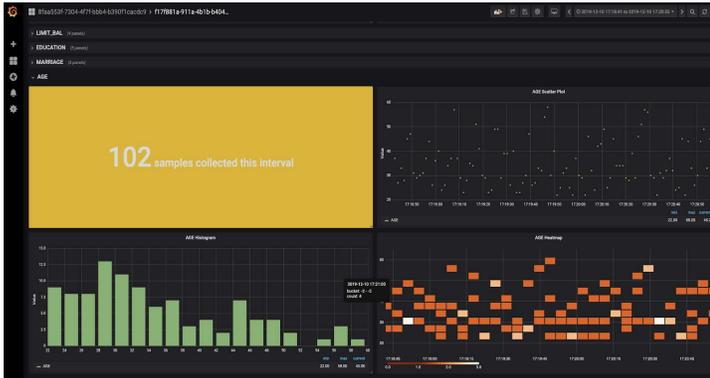




# H2O ModelOps



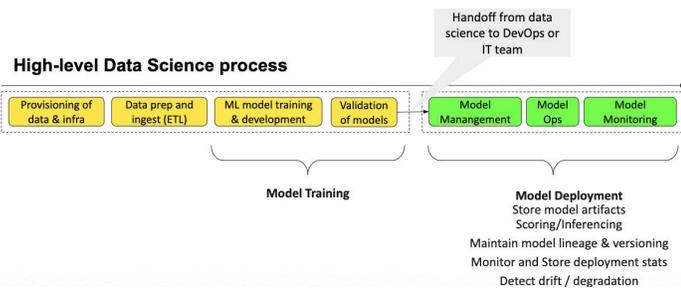
H2O ModelOps delivers a centralized catalog and management of models, model deployment, and monitoring capabilities for DevOps and data science teams.

## H2O ModelOps Overview

As enterprises “make their own AI”, a new set of challenges emerge. Maintaining reproducibility, traceability, and verifiability of machine learning models, as well as recording experiments, tracking insights and reproducing results, are key. Collaboration between teams is also necessary as “model factories” are created for enterprise-wide model data science efforts. Additionally, monitoring of models ensures that drift or performance degradation is addressed with either retraining or model updates. Finally, data and model lineage in case of rollbacks or addressing regulatory compliance is necessary. H2O ModelOps delivers centralized catalog and management, deployment, monitoring, collaboration and administration of machine learning models.

## The Data Science Workflow

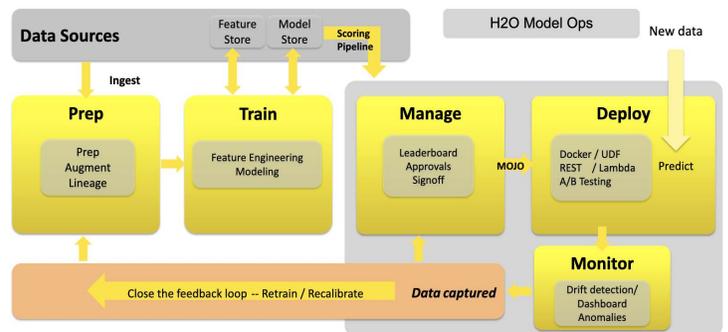
The data science workflow requires data scientists to iterate and collaborate on a number of steps that include training and optimizing of models and then deployment of the models. Once the models have been trained and tuned, the next step is to deploy the model into production.



## Deploying Models

Deployment of models takes a series of steps with the end goal to make predictions or inferences. Depending on the use case or the maturity of the organization, model results can be consumed locally, on a server or REST endpoint, or embedded within an application. Model deployment comes with its own set of challenges and often involves a team effort across data scientists, infrastructure (IT) and operations experts (DevOps). This can become even more complex as teams scale the number of models and the frequency of retraining. Model deployment is quintessential in delivering continuous business value. Model Ops consists of four major steps:

1. Scoring Pipeline
2. Model Management
3. Model Deployment
4. Model Monitoring



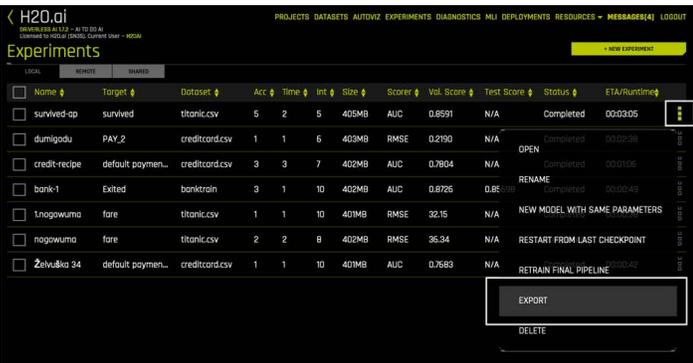
Once deployed, models need to be monitored to ensure they perform optimally within the thresholds defined by the business. Models will need to be retrained and replaced when a particular metric (e.g., accuracy) drift is detected.

## Scoring Pipeline

The key step in moving from training to production is packaging the model and key artifacts into a scoring pipeline that is “production-ready”. The scoring pipeline needs to include the model along with the feature engineering transformation used in the development of the model. To be production-ready this scoring pipeline needs to meet the latency requirements of inferencing and environment execution requirements (Java, Python, R, C++). H2O Driverless AI produces a production-ready scoring pipeline that provides low latency inferencing. This scoring pipeline can be used locally within H2O Driverless AI or in data science environments in Python or R. It can also be deployed to a variety of environments on-prem and in the cloud. Lastly, it can be embedded within applications or deployed as a function or procedure in a DBMS.

## Model Management

Oftentimes data scientists work in teams on a particular use case. Driverless AI has the notion of an experiment that can package data and various experiments for a particular use case. Driverless AI allows data scientists to share datasets and collaborate on experiments. Once the team of data scientists is ready to move to production, they can export their experiments to a deployment staging area in H2O ModelOps.



LOCAL	RENAME	SHARE	Name	Target	Dataset	Acc	Time	Int	Size	Scorer	Val. Score	Test Score	Status	ETA/Runtime
<input type="checkbox"/>			survived-op	survived	titanic.csv	5	2	5	405MB	AUC	0.8591	N/A	Completed	00:03:05
<input type="checkbox"/>			dumigodu	PAY_2	creditcard.csv	1	1	6	403MB	RMSE	0.2190	N/A	OPEN	00:02:58
<input type="checkbox"/>			credit-recipe	default paymen...	creditcard.csv	3	3	7	402MB	AUC	0.7804	N/A	Completed	00:01:06
<input type="checkbox"/>			bank-1	Exited	banktrain	3	1	10	402MB	AUC	0.8726	0.8699	Completed	00:02:45
<input type="checkbox"/>			1nogawuma	fare	titanic.csv	1	1	10	401MB	RMSE	32.15	N/A	NEW MODEL WITH SAME PARAMETERS	
<input type="checkbox"/>			nogawuma	fare	titanic.csv	2	2	8	402MB	RMSE	36.34	N/A	RESTART FROM LAST CHECKPOINT	
<input type="checkbox"/>			2elvu8kq 34	default paymen...	creditcard.csv	1	1	10	401MB	AUC	0.7683	N/A	RETRAIN FINAL PIPELINE	00:02:42

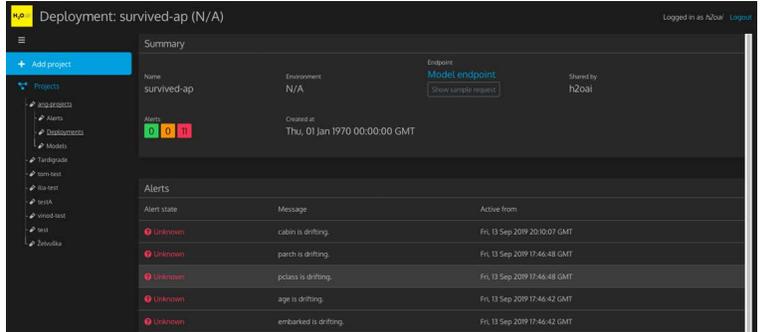
H2O ModelOps provides senior data scientists the ability to evaluate the various metrics from the experiments and all the associated experiment summaries and validate which models to promote to test or production.

## Model Deployment

Model performance is known to degrade over time. Businesses wishing to maximize the performance of their applications need to detect the optimal moment for swapping in new models without exposing their production environments to those that are unproven.

H2O ModelOps allows admins to promote models from staging to test and production environments and capturing key metrics..

With H2O ModelOps’ Champion/Challenger mode, teams can deploy models in shadow mode, where they operate on the same data as the one currently in production. This allows organizations to evaluate the performance and resilience of new models before promotion to production. All predictions carry metadata about the model responsible for the inference, and is used to trace a prediction back to details of a model and how it was trained.



## Model Monitoring

H2O ModelOps includes real-time monitoring of models for detecting anomalies, feature drift, and model performance degradation. Metrics are presented in a real-time dashboard.

The alerting capabilities of H2O ModelOps allow teams to be notified when discoveries are made. With built-in integrations for email, Slack and PagerDuty, teams are kept in the know with real-time alerts. With support for webhook, alerts can be streamed into an organization’s existing alerting systems.



When the threshold and anomalies are triggered alerts will be sent to the dashboard. The data scientist will have the option to retrain a model when the model exceeds a defined level of instability.

## H2O ModelOps for the Data Science Workflow

Today, as organizations increase their usage of machine learning models many organizations are facing challenges that limit the impact and scale of consumption of these models because of the lack of automation of this process. H2O ModelOps aims to assist with operationalizing, scaling and managing production deployments. H2O.ai customers can expect a full range of support, training, and expertise to assist them with their AI journey.