# Workbook 1.1



## Activity – Communication

The purpose of this activity is to strengthen the machine learning operations (MLOps) communications in your environment.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Roles | Business | Data | Develop | Deploy | Provide inference and monitor |
| Business stakeholder | ISIDS – Entomology Director | Identify legal resources/Fund Ground Truth project | Obtain Funding | Approve Staged deployments | Provide Business KPI/Metrics |
| Data engineer | DataOps Lead | DataOps – Pictures/  Ground Truth/ETL | Develop the DataOps Process | DataOps to MLOps integratiom | Meet batch delivery KPI’s / Data quality |
| Data scientist | Lead DS for Entomology | Feature Engineering & selection – Feature Store design | Model Selection & training | Approve Staged Deployment | Provide Model KPI/Metrics |
| ML engineer | Lead ML Engineer Computer Vision | Integrate DataOps with MLOps automation | Model Training  Automation/  Experimentation | MlOps Automated Deployment Design | MLOps Automation KPI/Metrics |
| DevOps engineer | Lead Debops | Integrate DataOps | MlOps-DevOps integration | CI/CD-Automated Deployments and Monitor |  |
| MLOps engineer | Data Science? | ML engineering |  |  |  |
| Software engineer | Lead Mobile Developer | Data Capture and integration and Model integration | Mobile/API and Security & Model Integration | Mobile Application Deployment Design | Mobile Application KPI/Metrics |
| Project manager | None assigned | Scrum Master/Legal/  Business | Cloud and Infrastructure Procurement | Project Gate Approver | Project KPIs |
| Security engineer | Security Lead | Data & Model Security Design | DataOps/  MLOps/  SecOps integration | Deployment Gate Approver | Security KPI/Metrics |

## Do you see any gaps between roles and ML stages? If so, do you think this work is someone’s responsibility, and you might need to investigate further?

|  |
| --- |
| This is a new project so we are sending a team to AWS training to help us identify gaps. We do have a current version in production so we have some success but we could not react fast enough to the MH invasion |

## What opportunities, if any, do you have to improve operations by seeking to assign someone to cover gaps in this table?

|  |
| --- |
| Speed to Market – Manual procedures |

## Action plan:

|  |  |  |
| --- | --- | --- |
| Action / outcome description | Stakeholders | Considerations |
| AWS Training | All | Time / Scheduling |
| Process review | All | All |
| Improved Automation | AllOps |  |
| Data Scientist Tooling | ML Engineering/Data Science |  |
| Improved Security | AllOps – SecOps Lead |  |
|  |  |  |
|  |  |  |

# Workbook 1.2





## Activity – From DevOps to MLOps

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fill out the following table.Operations tools and processes | Do you have this?  [Y/N] | Who is responsible for it? | Could it be used better in an MLOps context? | If you do not have it, what are your next steps (action plan)? |
| Code building pipeline | Y | DevOps | Y |  |
| Code version control | Y | DevOps/SecOps | Y | Better triggers when data changes |
| Model version control | N | MlOps/DevOps | Y | Development is siloed |
| Data version control | Y/N | DataOps | Y | DataOps initiative |
| Model building pipeline | N | MLOps | Y | Hope to leverage current tech |
| Data pipeline | Y/N |  |  | DataOps initiative |
| Approval process | Y | DevOps/SecOps/MLOps  BusOps | Y | Cumbersome w/out value |
| Key performance indicators | Y/N | DevOps/SecOps/MLOps  BusOps | Y | Focused effort / AWS Training |
| Baseline metrics | Y | DevOps/SecOps/MLOps  BusOps |  | Based on current app |

## Who approves models to be deployed to production in your ML environment?

|  |
| --- |
| It is basically a committee with inconsistent representation with conflicting priorities. Much could be automated with better KPI’s/Metrics. Business / Data Science / AllOps teams have different terminology. |

## Action plan:

|  |  |  |
| --- | --- | --- |
| Action / outcome description | Stakeholders | Considerations |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Workbook 2.1



## Select the top challenges you are likely to encounter in your environment as it relates to ML data, models, and code.

|  |  |  |
| --- | --- | --- |
| Data | Model | Code |
| Providing mechanisms by which training / testing / validation datasets, training scripts, models, experiments, and service wrappers can all be versioned appropriately | Managing changes to training / testing / validation datasets, training scripts, models, experiments, and service wrappers, and ensuring each is auditable across the full lifecycle | Knowing the risks of trying to use Jupyter notebooks in production |
| Treating training / testing / validation datasets as managed assets under an MLOps workflow | Enabling all ML frameworks to be used within the scope of MLOps, regardless of language or platform | Methods for wrapping trained models as deployable services in scenarios where the data scientists training the models might not be experienced software developers with a background in service-oriented design |
| Applying MLOps to very large-scale problems at petabyte scale and beyond | Enabling MLOps to support a broad range of target platforms, including but not limited to CPU, GPU, TPU, custom ASICs, and neuromorphic silicon | Providing appropriate pipeline tools to manage MLOps workflows transparently as part of existing DevOps solutions |
| Access to key performance indicator business data for MLOps team members who need to evaluate model effectiveness | Ensuring efficient use of hardware in both training and operational scenarios | Model abstration |
|  | Testing ML assets appropriately | Maximize longevity of ML assets |
|  | Online ML | Ability to quickly cut out a model or roll back immediately to an earlier version |

# Workbook 2.2





## End-to-end MLOps frameworks are you using?

Not currently using an end-to-end MLOps solution

Apache Airflow

Kubeflow

Amazon SageMaker

AWS Step Functions

Other: \_\_\_Jenkins\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Would migrating to an end-to-end solution for MLOps be better for your organization than making adjustments to the existing ML process? Why or why not?

|  |
| --- |
| There are multiple different automation pipeline solutions in place through acquisition, philosophical difference. Seeing ways we could us AWS API’s effectively with any of the existing solutions. Perhaps a team could develop API’s to be used by any of the solutions for Model Training and development. Current evaluation underway to utilize Sagemaker Pipelines |

## Action plan:

|  |  |  |
| --- | --- | --- |
| Action / outcome description | Stakeholders | Considerations |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Workbook 2.3



## Activity – Common orchestration elements

Fill out the below chart. Indicate which elements are already present in your ML process. For missing elements or elements that have room for improvement, indicate

|  |  |  |
| --- | --- | --- |
|  | Included in your ML process [Y/N] | Priority for adoption |
| Version control | Y | Critical - Staged |
| CI/CD | Y | Critical |
| ML model builder | Both Local and SageMaker | Critical - Staged |
| Monitoring ML workloads | Y - Explainability, Model Monitoring, Infrastructure | Critical |
| Workflow security | Y – Model/Data/Pipleine/Endpoints | Critical |

Work with data science teams to integrate processes.

# Workbook 3.1



In this workbook activity, you will record characteristics of your ML environment. This will help you think about the methods you currently use for deployment. It will also help you consider options for MLOps improvements as you study deployment alternatives.

## How are you using ML in current and planned projects?

|  |
| --- |
| We have multiple computer vision solutions in place with more ideas than we have time to deliver. There is more interest in integrating ML in internal systems for client churn analysis and product pricing analysis. |

## For what purpose are models packaged in deployment?

Batch inference

Online inference

Both

Not using either yet

## At your organization, models are generally built using what ML framework (or these ML frameworks)?

|  |
| --- |
| All of our solutions are MXNet and Gluon currently. |

## Models are built using what key ML algorithms?

|  |
| --- |
| Most of our models use existing CNN’s with retraining. Some newer models are being developed based on existing CV models. |

# Workbook 3.2



## What scaling strategies do you use in your ML process?

|  |
| --- |
| We started on the cloud using AWS. We are familiar with Autoscaling/ Load Balancers and we use some serverless technology. Most of our deployments run on Mobile platforms but we are looking to use streaming and batch more as well as utilizing cache based solutions. |

## Of the different ways to implement inference that were discussed, which are most similar to those you use in production?

Basic hosting inference

Basic transform inference

**Basic transform** prediction in advance

Streaming with batch transform inference

Inference pipeline with multiple containers

Customized inference using AWS services

Customized inference using containers on AWS services

Online inference with online inference pipeline

Customized inference hosting using containers with managed services

**SageMaker hosted** endpoint

SageMaker hosted services

SageMaker batch transform

# Workbook 3.3



In this workbook activity, you will consider how deployment is currently implemented in your environment. You will also consider opportunities for employing MLOps deployment techniques described in this course. In the final step, you will create an action plan of items to investigate, discuss, and implement to improve MLOps in your ML workflow.

## Which deployment strategies are relevant for your organization?

Update(standard update)

A/B testing

Blue/green

Canary

Champion/challenger

## How often do changes occur to the ML models in production?

|  |
| --- |
| We have a monthly/quartely release cycle. We need get this down to days and even weeks for some of our solutions |

## What are the key criteria that determine when a new or updated ML model is deployed to production?

|  |
| --- |
| Field testers/Partners report issues. We have some preventative models in place. Changes in mobile device camera resolutions are causing issues. We should be able to monitor app and model performance better |

## Which Amazon SageMaker integration strategies are relevant for your organization?

Not currently using any cloud services in our ML process

Originate the ML process on the cloud

Leverage a built-in algorithm to train models

Use a built-in framework for the entire ML process

Bring your own model (The model is built on premises and then brought to the cloud for hosting.)

Bring your own container

Use a hosted ML model (model as a service, such as Amazon Rekognition)

## What methods do you use, or are considering, in your ML process?

|  |
| --- |
|  |

## Action plan:

|  |  |  |
| --- | --- | --- |
| Action / outcome description | Stakeholders | Considerations |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Questions for the instructor:

|  |
| --- |
|  |

# Workbook 3.4



## Which Amazon SageMaker integration strategies are relevant for your organization?

* Not currently using any cloud services in our ML process
* Originate the ML process on the cloud

X Leverage a built-in algorithm to train models

* Use a built-in framework for the entire ML process
* Bring your own model (The model is built on premises and then brought to the cloud for hosting.)

X Bring your own container

* Buy and customize from Marketplace (Third party pre-built algorithms or models are used)
* Use a hosted ML model (model as a service, e.g. Amazon Rekognition)

## What methods are you using or are you considering to implement in your ML process?

We need a better mechanism to scale, and quickly change models. Pipelines would definitely help, and we need to better understand model lineage.

## Action plans:

# Workbook 4.1



## How and why might your ML model’s ability to provide accurate prediction decay over time?

# Workbook 4.2



## What are the most important business requirements?

|  |
| --- |
| All of our current solutions and image recognition solutions. Accuracy is the most important model measurement. Application Stability and version control are critical for the mobile applications and model deployment.  We just started doing batch. Meeting batch processing deadlines is critical for our new cache based design. |

# Workbook 4.1



## How and why might your ML model’s ability to provide accurate prediction decay over time?

|  |
| --- |
| New species of plants and insects are being included in our models. Camera resolutions are changing and vary widely. Seasonal changes in lighting, blooms etc are not represented in our training data. |

# Workbook 4.3



## Activity – What is your MLOps monitoring recommendation?

Fill out the following table to list what to monitor, how you’ll monitor it, and who’s responsible for each task. Note ideas for alerts and automation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | What to monitor | How to monitor | Who’s responsible | Type(s) of alerts | Automation recommendations |
| Business requirements | Acurracy/  Drift  Security | Earlier drift detection  Simulation  Live Performance/  Daily Ground Truth Evaluations | Product Owners for KPIs  MLOps for Implementaion | Drift  Performance  Stability | Earlier drift detection  Simulation  Live Performance/  Daily Ground Truth Evaluations  Security Analysis |
| Model performance | Accuracy (CV Apps) | End-to-end Monitoring  Compare against Ground Truth | Data Science/  MLOps Engineering | Drift  Training and Evalution  Post deployment (A/B) | Model Monitoring  Model Evaluation/Experiments |
| Hosting activities | Time to deliver predictions  Stability  Security | Infrastructure Monitoring  Resource Utilization Monitoring | Developers  DevOps  MLOps  SecOps  DataOps | Crash detection  Resource Usage  ThroughPut | Mobile MDM  Cloudwatch  Edge Management  CloudTrail  AWS Security best practices |

## Action plan:

|  |  |  |
| --- | --- | --- |
| Action / outcome description | Stakeholders | Considerations |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Questions for the instructor:

|  |
| --- |
|  |