Intro to MLOps

Quickly Deploying Data & ML Solutions via the Cloud

Victor Geislinger

What's Hardest about Data Science / Machine Learning?

What's Hardest about Data Science / Machine Learning?

- Getting something useful to stakeholders

What's Hardest about Data Science / Machine Learning?

- Getting something useful to stakeholders

- Taking "too much time" for perfect over good enough

What's Hardest about Data Science / Machine Learning?

- Getting something useful to stakeholders

- Taking "too much time" for perfect over good enough

Solutions?

What's Hardest about Data Science / Machine Learning?

- Getting something useful to stakeholders

- Taking "too much time" for perfect over good enough

Solutions?

- More focus on automation

What's Hardest about Data Science / Machine Learning?

- Getting something useful to stakeholders

- Taking "too much time" for perfect over good enough

Solutions?

- More focus on automation

- Iteratively improving
  - Rolling out
  - Working with SEs

What's Hardest about Data Science / Machine Learning?

Warning: Opinion Incoming...

What's Hardest about Data Science / Machine Learning?

- Moving trend in DS/ML

What's Hardest about Data Science / Machine Learning?

- Moving trend in DS/ML
- AutoML

What's Hardest about Data Science / Machine Learning?

- Moving trend in DS/ML
- AutoML
- Less time in developing algorithms

What's Hardest about Data Science / Machine Learning?

- Moving trend in DS/ML
- Reliance on understanding of fundamentals (what you've been working on!)
- AutoML
- Less time in developing algorithms

A Solution – MLOps: What is it?

A Solution – MLOps: What is it?

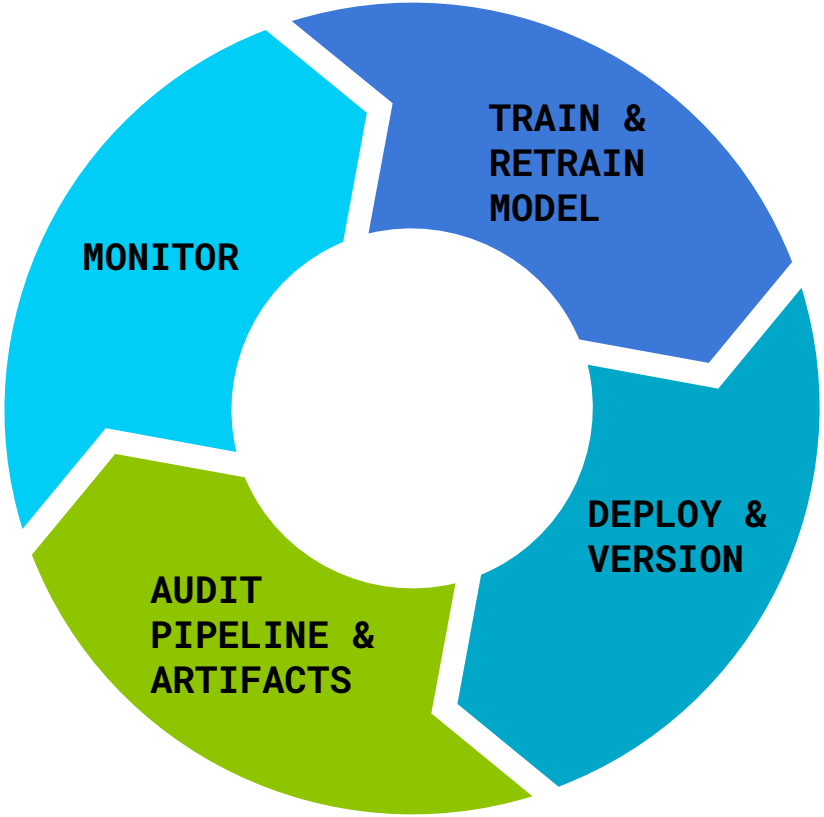
- Machine Learning Operations
  - (similar to DevOps)

Pictured: Performing “ML Operations”...

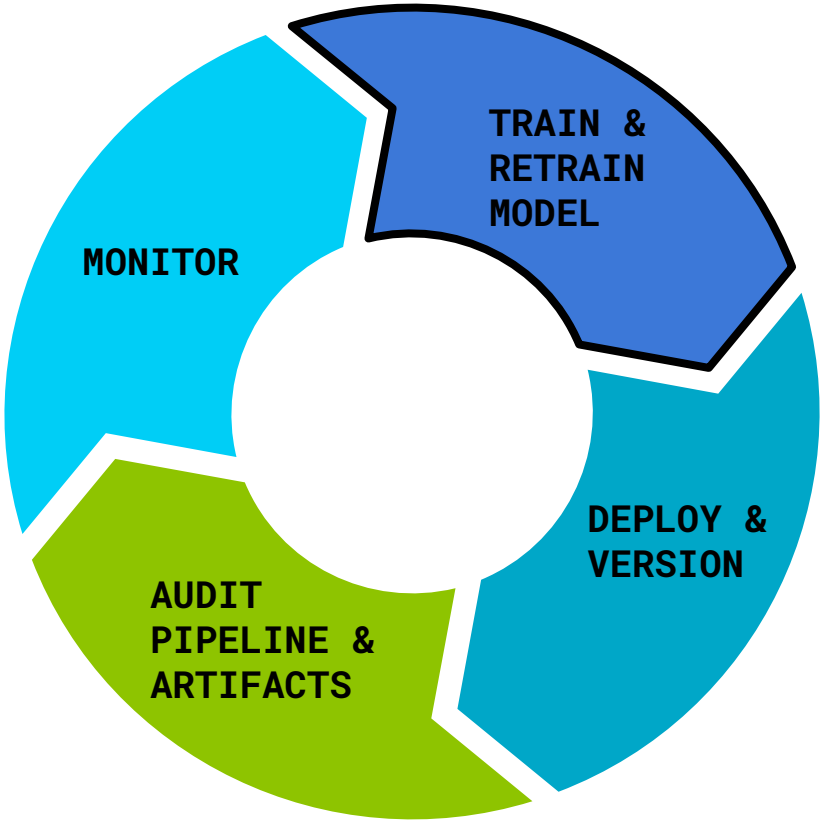
A Solution – MLOps: What is it?

- Machine Learning Operations
  - (similar to DevOps)
- Methods to incorporate the ML lifecycle
  - (move faster into production/usefulness)

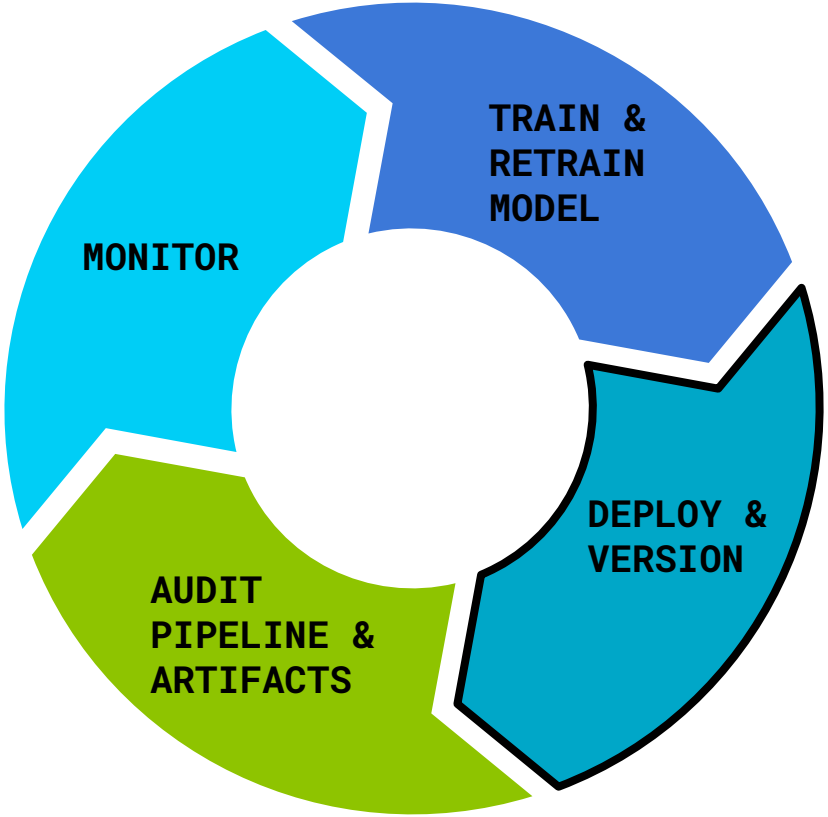
Pictured: Performing “ML Operations”...



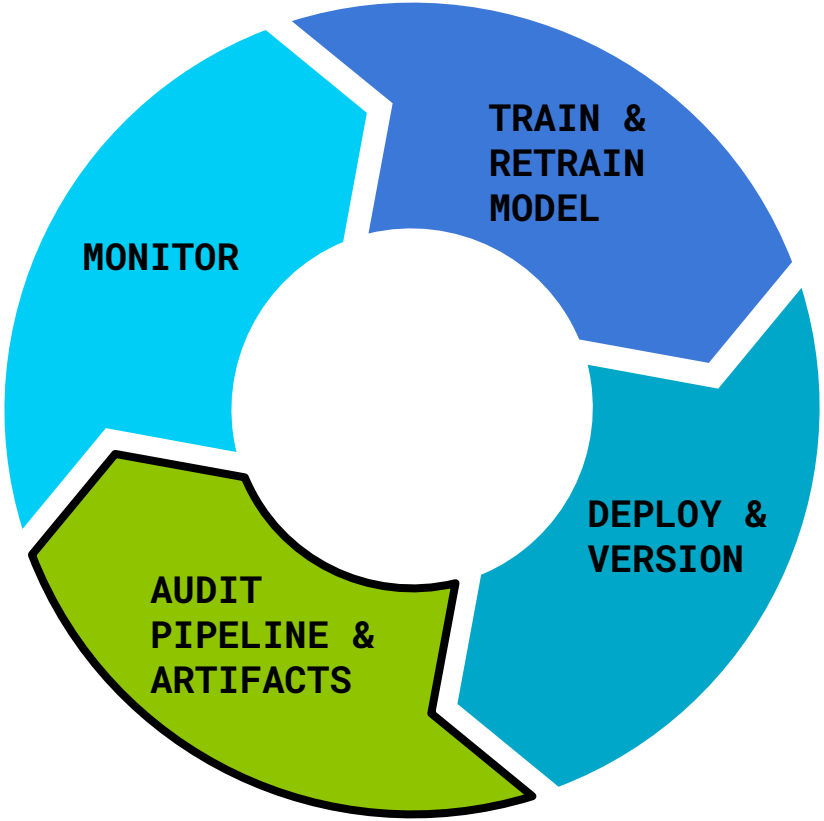
MLOps Feedback Loop



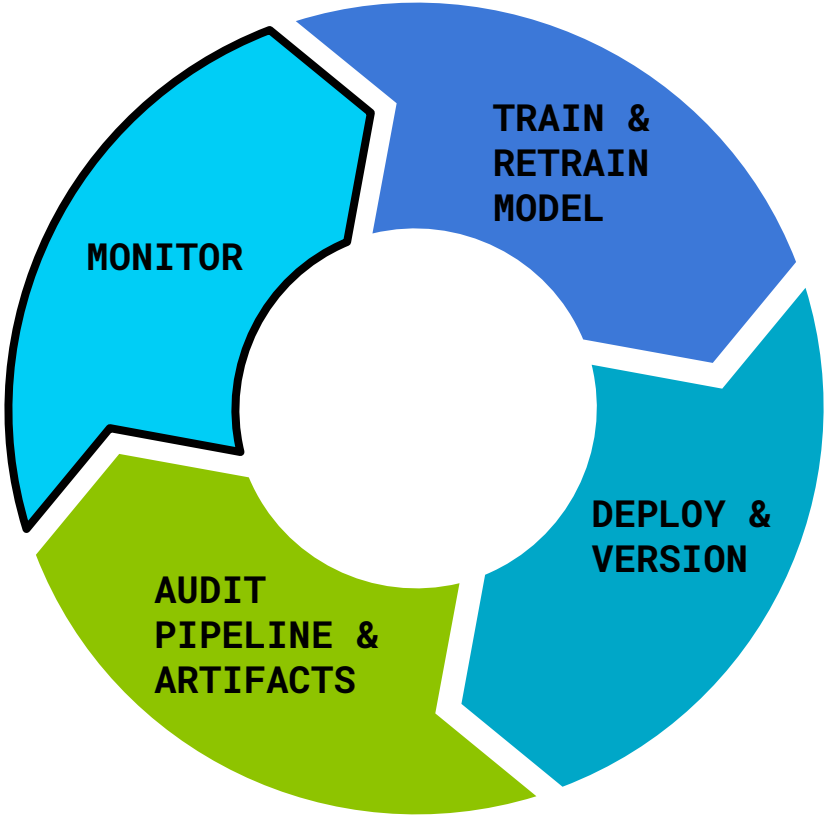
- Data can change
- People can change
- Reusable & automatic pipelines help



- Continuous Delivery (CD)
- Automated steps (infrastructure too!)
- Minimal friction to deploy new models



- Problems in accuracy, bias, & security arise
- Identify where they occur
- Useful for continuous improvement



- Data drift!
- Prevents accuracy issues

Let’s Learn By Doing!

Scenario

Scenario

Scenario

Scenario

Scenario

- Business problem statement

Scenario

- Business problem statement
- Use case? What's important?
    - Parse Intent: What *needs* to be done?

Scenario

- Business problem statement
- Use case? What's important?
  - Parse Intent: What *needs* to be done?
- How difficult in solving problem?
  - Pre-packaged solution available?

Scenario

- Business problem statement
- Use case? What's important?
  - Parse Intent: What *needs* to be done?
- How difficult in solving problem?
  - Pre-packaged solution available?
- How do we evaluate?

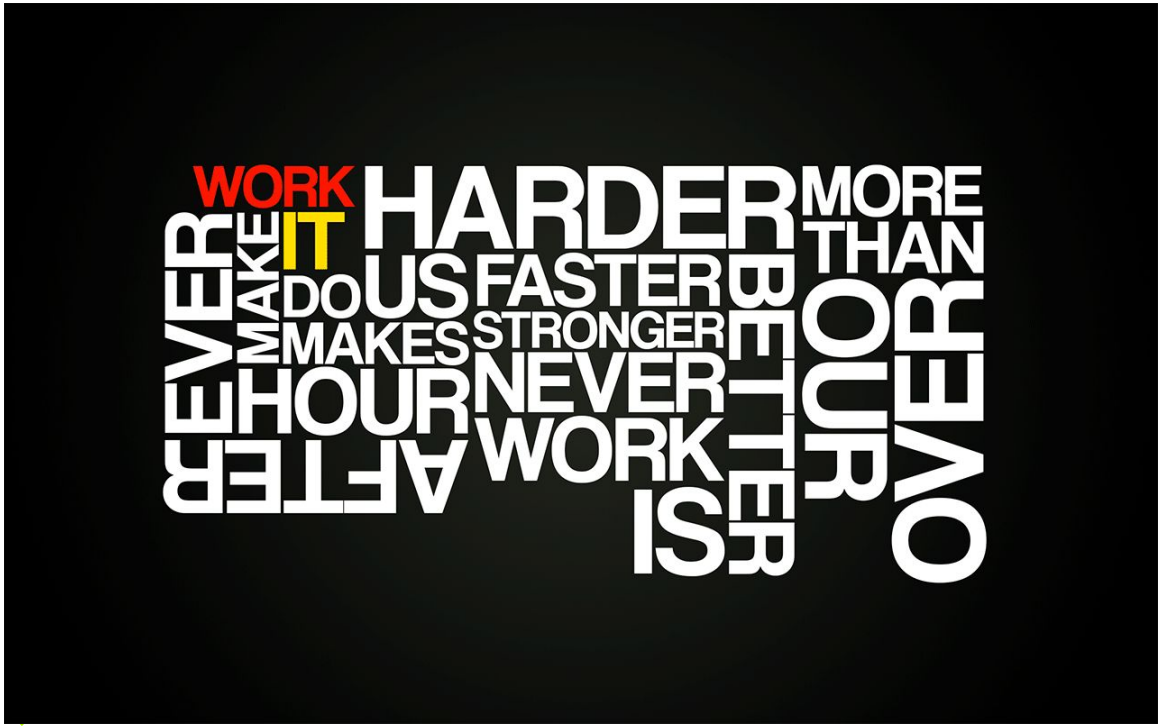
Scenario

- Business problem statement
- Use case? What's important?
  - Parse Intent: What *needs* to be done?
- How difficult in solving problem?
  - Pre-packaged solution available?
- How do we evaluate?
- How do I get data? Difficult?

Scenario

- Business problem statement
- Use case? What's important?
  - Parse Intent: What *needs* to be done?
- How difficult in solving problem?
  - Pre-packaged solution available?
- How do we evaluate?
- How do I get data? Difficult?
- Think of CD/ improvement

Continuous Deployment / Continuous Improvement



Continuous Deployment / Continuous Improvement

- Proof of concept (MVP)
- What parts can be iterated?

Continuous Deployment / Continuous Improvement

- Proof of concept (MVP)
- What parts can be iterated?

- Work w/SE? Build our own solution?
  - Cost of maintenance?
- Evaluation, Monitoring, etc.

Looking for Data

Looking for Data

- Is this data open/accessible?
  - Solutions already exist? (Pleasssse!!!)

Looking for Data

- Is this data open/accessible?
  - Solutions already exist? (Pleasssse!!!)
- Is there “enough”?
  - Other sources + company’s data

Looking for Data

- Is this data open/accessible?
  - Solutions already exist? (Pleasssse!!!)
- Is there “enough”?
  - Other sources + company’s data

- Do we have a “special” use case
  - How flexible is this?
  - How inflexible is it?
  - Issues? Fixes?

Looking for Data

- We'll use this:

[UCI Machine Learning Repository: SMS Spam Collection Data Set]

Simple Model

- Why?
- Make a baseline!
- What if a simple solutions *is* the solution?!
    - (Story time)
-

Simple Model

- Why?
- Make a baseline!
- What if a simple solutions *is* the solution?!
  - (Story time)
-
- Does it have “spammy” words (free, win, etc.)

Checking Out Code!

We'll look at this repo:

https://github.com/MrGeislinger/mlops-example-spam-detector

Checking Out Code!

We'll look at this repo:

https://github.com/MrGeislinger/mlops-example-spam-detector

Let's go to a notebook!

Revisit for Modularization

- Focus on separating out the parts

- Allows us to make *different* UI or allow an interface we don’t know (prescribed from other team)

Checking Out Code!

Let’s get modularize!

Back to some code!!

Revisit for UI

I am but a lowly data scientist;

I don't know much about them UI design...

- Victor Geislinger (everyday)

Revisit for UI

- We’ll use Streamlit (Flask-based)

I am but a lowly data scientist; I don’t know much about them UI design...

Revisit for UI

- We’ll use Streamlit (Flask-based)
  - It’s “so hot”

Revisit for UI

- We'll use Streamlit (Flask-based)
  - It's "so hot"
  - Pretty easy to use with low-learning curve

I am but a lowly data scientist; I don't know much about them UI design...

Revisit for UI

- We'll use Streamlit (Flask-based)
  - It's "so hot"
  - Pretty easy to use with low-learning curve

- Flask might be better to make a simple API
  - Streamlit is arguably prettier and we're doing a presentation. Sooo…

I am but a lowly data scientist; I don't know much about them UI design…

Revisit for UI

- We'll use Streamlit (Flask-based)
  - It's "so hot"
  - Pretty easy to use with low-learning curve

- Flask might be better to make a simple API

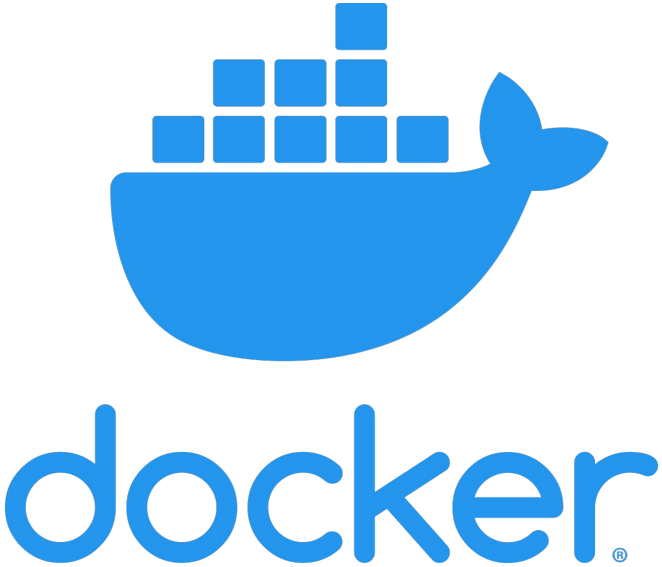
I am but a lowly data scientist; I don't know much about them UI design...

Checking Out Code!

Back to some code!!

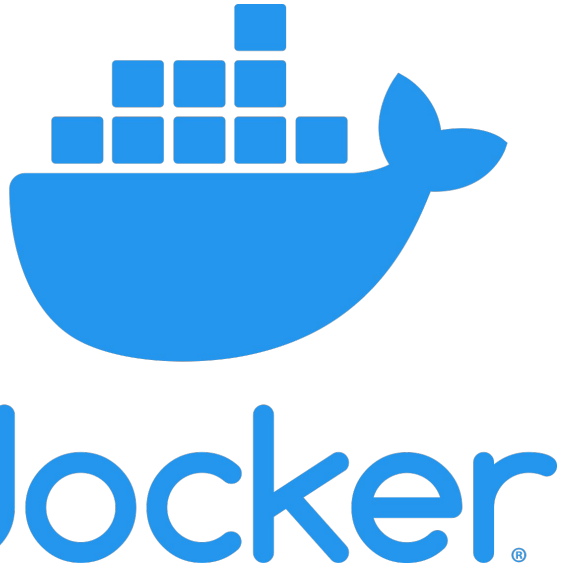
Gotta make it pretty for our stakeholders!

Docker-ize It!



Docker-ize It!

- Allows to “package” up a working environment (called a “container”)
- Makes deploying a lot easier/simpler
- Used a lot by software engineers



Checking Out Code!

Let’s make a Dockerfile and container-ize our code!

Cloud Deployment: Serverless

Cloud Deployment: Serverless

- Deploy to the “cloud”

- Past: Manage a server to run our code
- Future: “Serverless”
  - Run code only when needed
  - Easier to implement/rollout versions

Checking Out Code!

We'll use Google Cloud Platform (GCP)

Specifically using "Cloud Run" that uses Docker images

(But there are countless of different methods!)

Iterate (if we have time)

- Hey we need an improvement!
  - Can be from the DS/ML team
  - Can be a “signal”
- Ideally continuously monitor & train, automatically replacing the model

Iterate (if we have time)

- Hey we need an improvement!
  - Can be from the DS/ML team
  - Can be a “signal”
- Ideally continuously monitor & train, automatically replacing the model

- How you update model depends on use case
  - Emergency:
    - Stop everything; rollback to working version
  - Signal new model:
    - auto-rollout
  - Signal drift: investigate

Other Considerations

- Data & Concept Drift
- Monitoring & Logging
- CI / CD
- Tools
    - GitHub Action
    - GCP Vertex AI
    - AWS Sagemaker

My Advice

- Learn CLI
  - Get comfortable in the terminal

My Advice

- Learn CLI
  - Get comfortable in the terminal

- Learn to modularize your code!
  - Python files, Classes, packages, etc.

My Advice

- Learn CLI
  - Get comfortable in the terminal
- Learn to modularize your code!
  - Python files, Classes, packages, etc.
- Docker (see above points)

My Advice

- Learn CLI
  - Get comfortable in the terminal
- Learn to modularize your code!
  - Python files, Classes, packages, etc.
- Docker (see above points)
- Cloud - especially serverless!

Further Reading

- Practical MLOps
  - Focus on ideas
- Building Machine Learning Pipelines
  - How to automate model lifecycles
- Data Science at the Command Line
  - Useful way to introduce CLI
  - Free online
- Building Machine Learning Powered Applications
  - More about ideas and some use cases

Other books

- Build a Career in Data Science
- Hands-On ML w/ Scikit-Learn ,.. (2nd Ed).
- Python Data Science Handbook
- Machine Learning Design Patterns

THANKS!

@MrGeislinger / mrgeislinger.com