

MLOps

CI/CD for Machine Learning

SASHA ROSENBAUM



Sasha Rosenbaum

Sr. Program Manager

 Microsoft

@DivineOps

What is MLOps and WHY should you care?

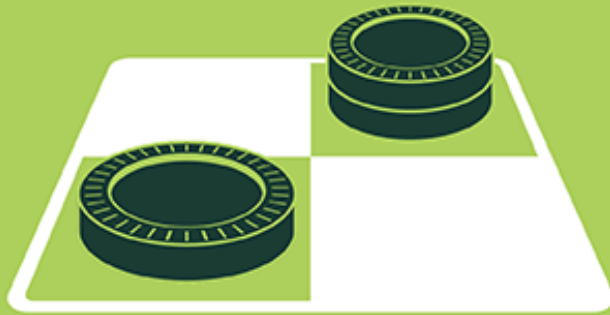
Machine Learning (ML)

Is the science of getting computers to act

Without being explicitly programmed

ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



MACHINE LEARNING

Machine learning begins to flourish.

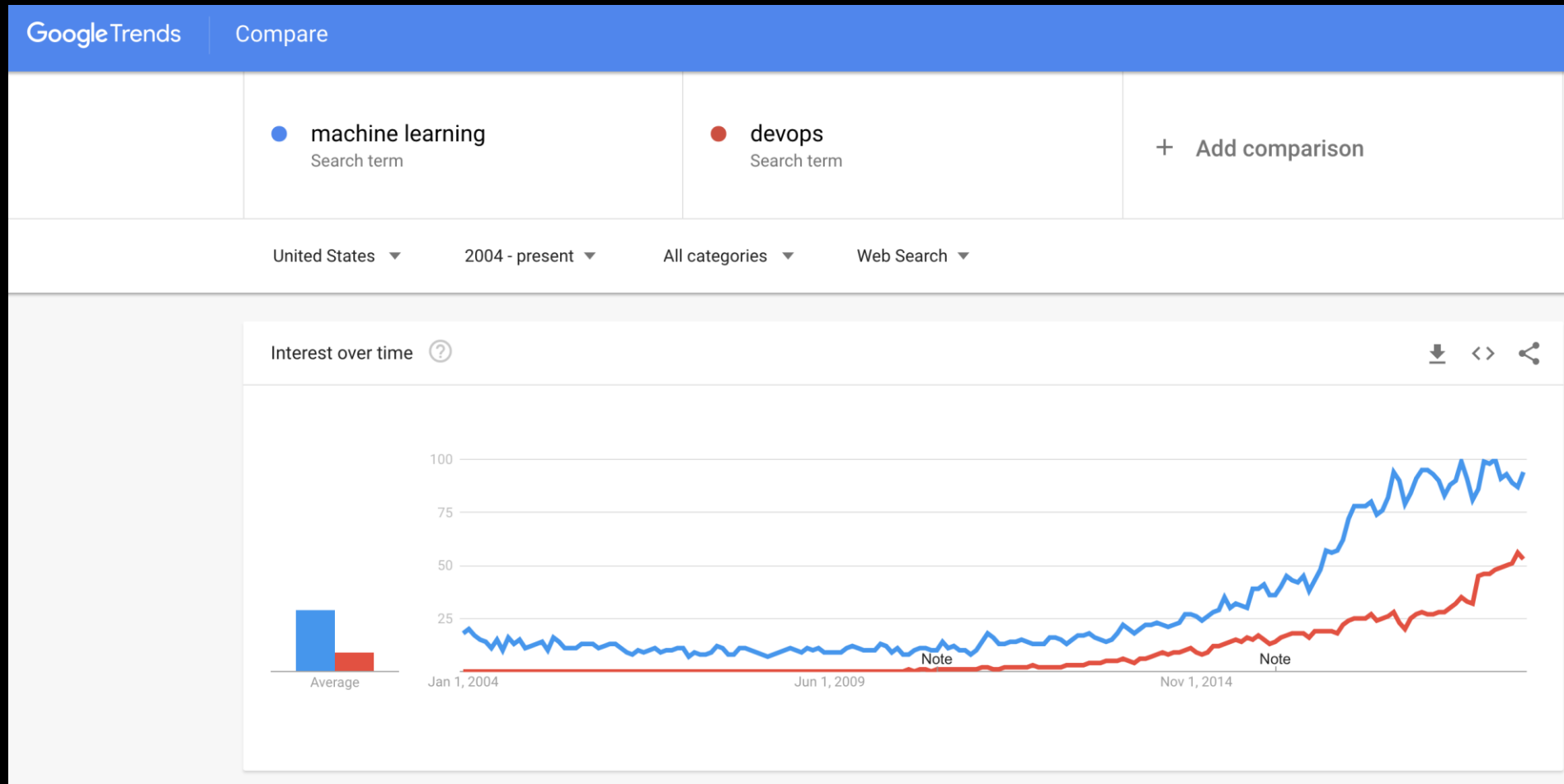


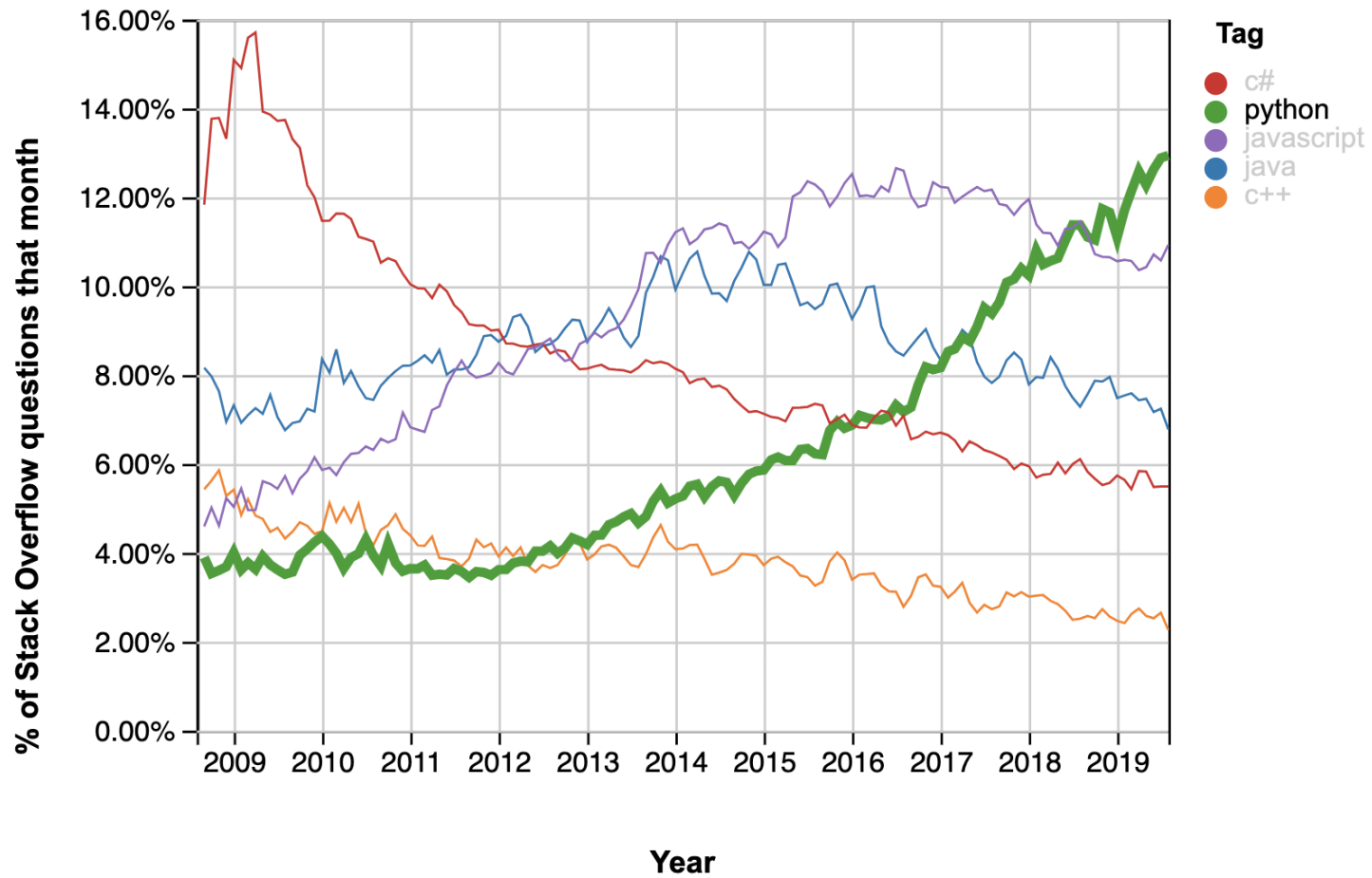
DEEP LEARNING

Deep learning breakthroughs drive AI boom.



Machine Learning vs DevOps Google searches





Python questions on Stack Overflow

OK, but why should YOU care?



ginablaber

@ginablaber

Follow



The story of enterprise Machine Learning: “It took me 3 weeks to develop the model. It’s been >11 months, and it’s still not deployed.”
[@DineshNirmalIBM](#) [#StrataData](#) [#strataconf](#)

10:19 AM - 7 Mar 2018

7 Retweets 19 Likes

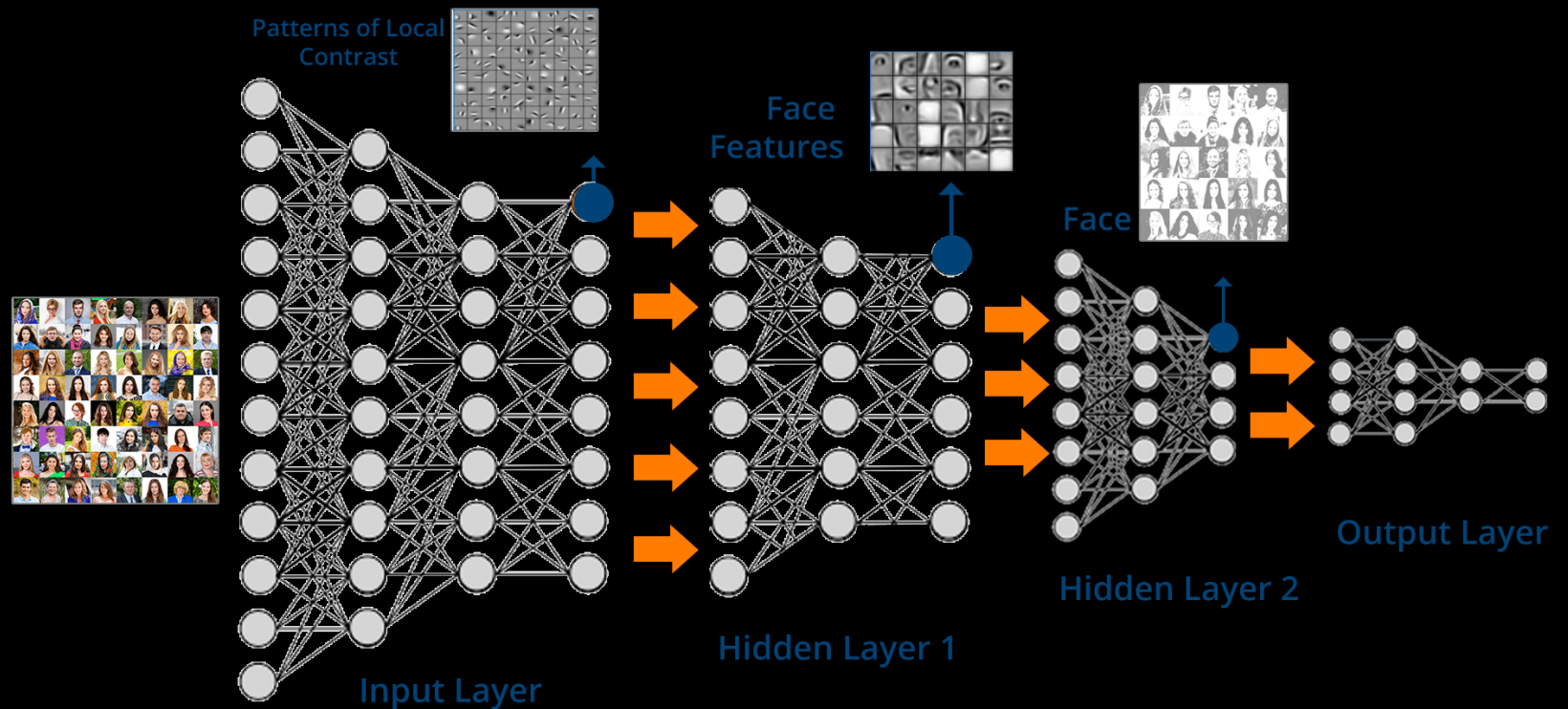


19

Data Scientists just want to Data Science

Deep Learning

Some ML training algorithms are complex



Deep Learning - Backpropagation

Some ML training algorithms are complex

Back-propagation

1. receive new observation $\mathbf{x} = [x_1 \dots x_d]$ and target y^*
2. **feed forward:** for each unit g_j in each layer $1 \dots L$
compute g_j based on units f_k from previous layer: $g_j = \sigma \left(u_{j0} + \sum_k u_{jk} f_k \right)$
3. get prediction y and error $(y - y^*)$
4. **back-propagate error:** for each unit g_j in each layer $L \dots 1$

(a) compute error on g_j

$$\frac{\partial E}{\partial g_j} = \sum_i \underbrace{\sigma'(h_i)}_{\substack{\text{how } h_i \text{ will} \\ \text{change as} \\ g_j \text{ changes}}} \underbrace{v_{ij}}_{\substack{\text{was } h_i \text{ too} \\ \text{high or} \\ \text{too low?}}} \underbrace{\frac{\partial E}{\partial h_i}}_{\substack{\text{should } g_j \\ \text{be higher} \\ \text{or lower?}}}$$

(b) for each u_{jk} that affects g_j

(i) compute error on u_{jk}

$$\frac{\partial E}{\partial u_{jk}} = \frac{\partial E}{\partial g_j} \underbrace{\sigma'(g_j)}_{\substack{\text{do we want } g_j \text{ to} \\ \text{be higher/lower}}} \underbrace{f_k}_{\substack{\text{how } g_j \text{ will change} \\ \text{if } u_{jk} \text{ is higher/lower}}$$

(ii) update the weight

$$u_{jk} \leftarrow u_{jk} - \eta \frac{\partial E}{\partial u_{jk}}$$

Copyright © 2014 Victor Lavrenko

The screenshot shows a Jupyter Notebook interface in a web browser. The browser address bar shows 'localhost:8888/notebooks/Untitled.ipynb'. The Jupyter interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Help) and a toolbar with icons for file operations and execution. The main content area displays a code cell with the following text:

```
In [1]: import xgboost
```

Below the code cell, a detailed Windows error traceback is shown:

```
WindowsErrorTraceback (most recent call last)
<ipython-input-1-afdaff4619ce> in <module>()
----> 1 import xgboost

D:\Anaconda2\lib\site-packages\xgboost-0.6-py2.7.egg\xgboost\__init__.py in <module>()
     9 import os
    10
----> 11 from .core import DMatrix, Booster
    12 from .training import train, cv
    13 from . import rabit                # noqa

D:\Anaconda2\lib\site-packages\xgboost-0.6-py2.7.egg\xgboost\core.py in <module>()
    110
    111 # load the XGBoost library globally
-> 112 _LIB = _load_lib()
    113
    114

D:\Anaconda2\lib\site-packages\xgboost-0.6-py2.7.egg\xgboost\core.py in _load_lib()
    104 if len(lib_path) == 0:
    105     return None
-> 106 lib = ctypes.cdll.LoadLibrary(lib_path[0])
    107 lib.XGBGetLastError.restype = ctypes.c_char_p
    108 return lib

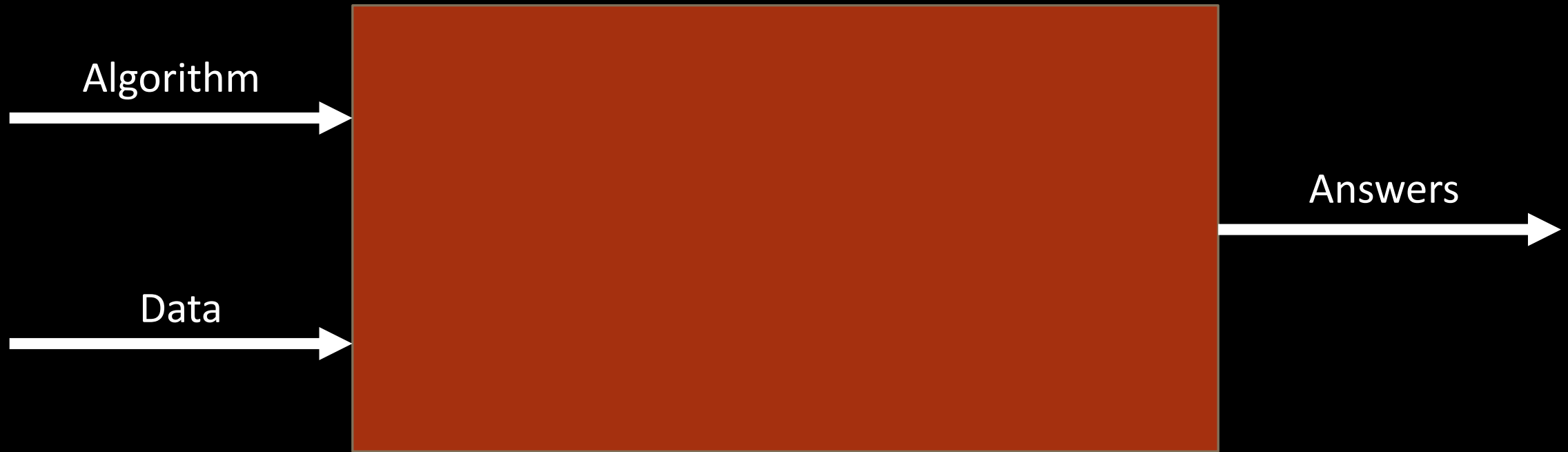
D:\Anaconda2\lib\ctypes\__init__.pyc in LoadLibrary(self, name)
    438
    439 def LoadLibrary(self, name):
-> 440     return self._dlltype(name)
```

Typical data
scientist work
environment

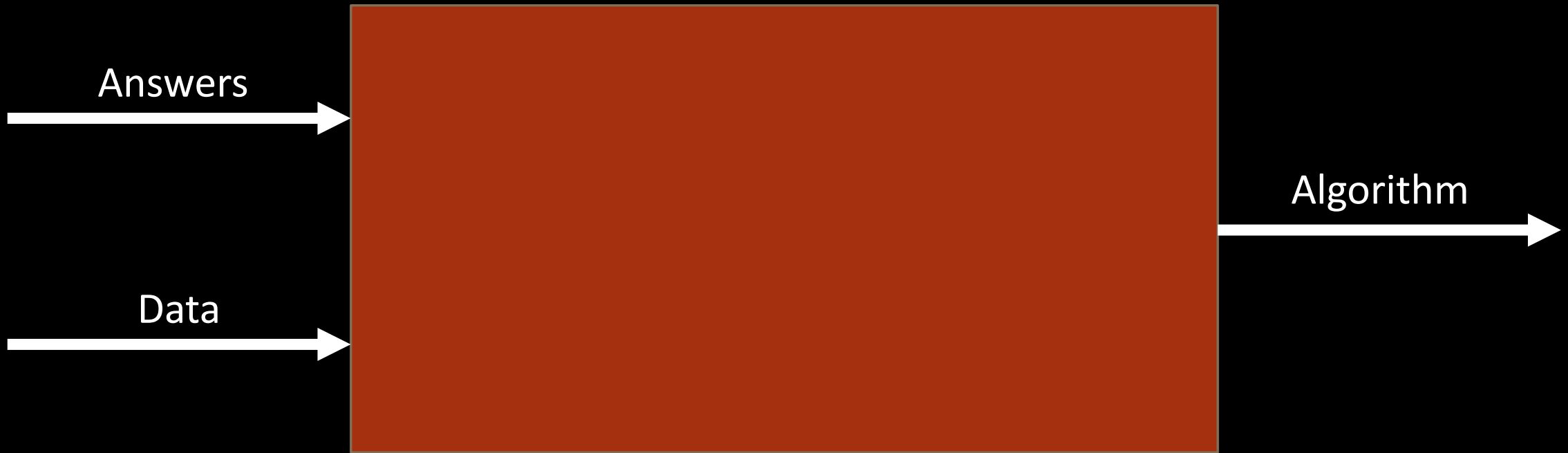


We've got
the notebook
into source
control!

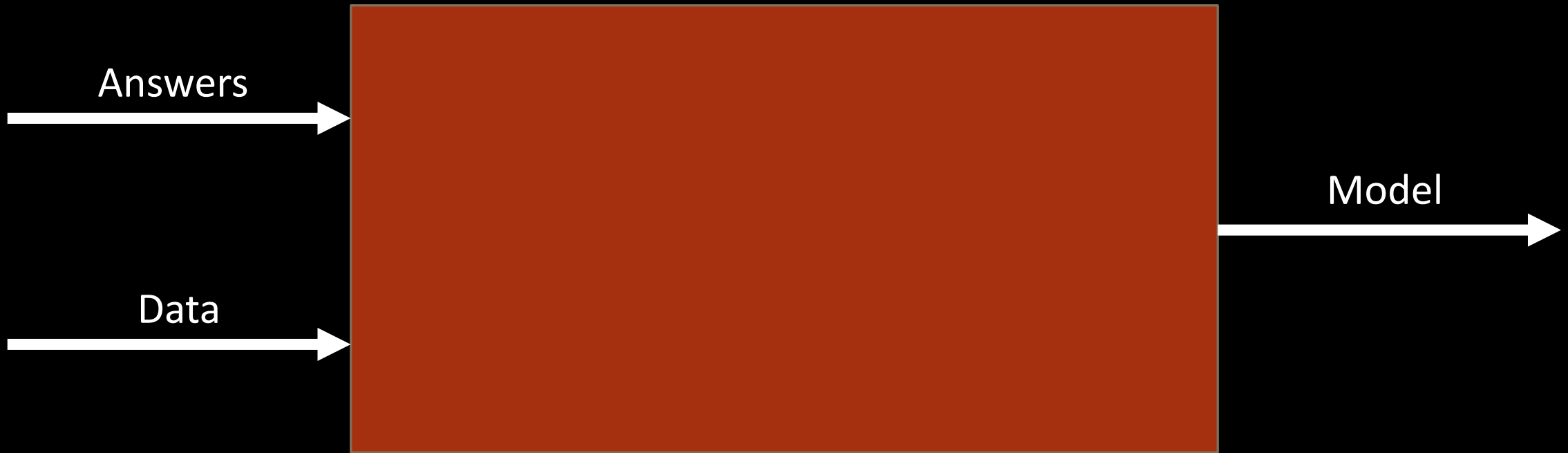
Programming



Machine Learning



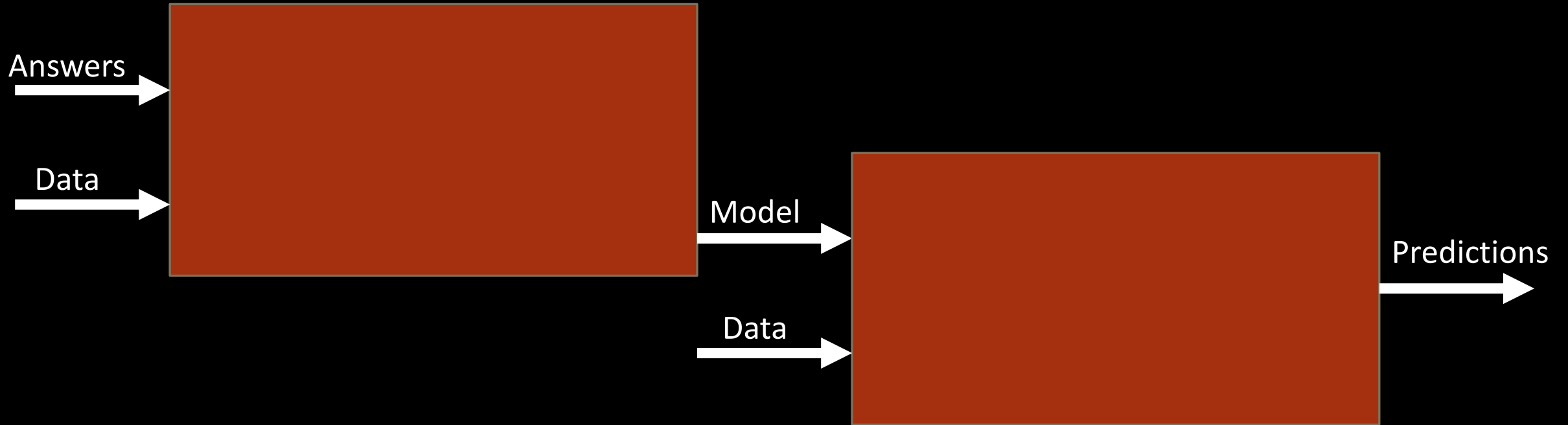
Machine Learning



Machine Learning



Machine Learning





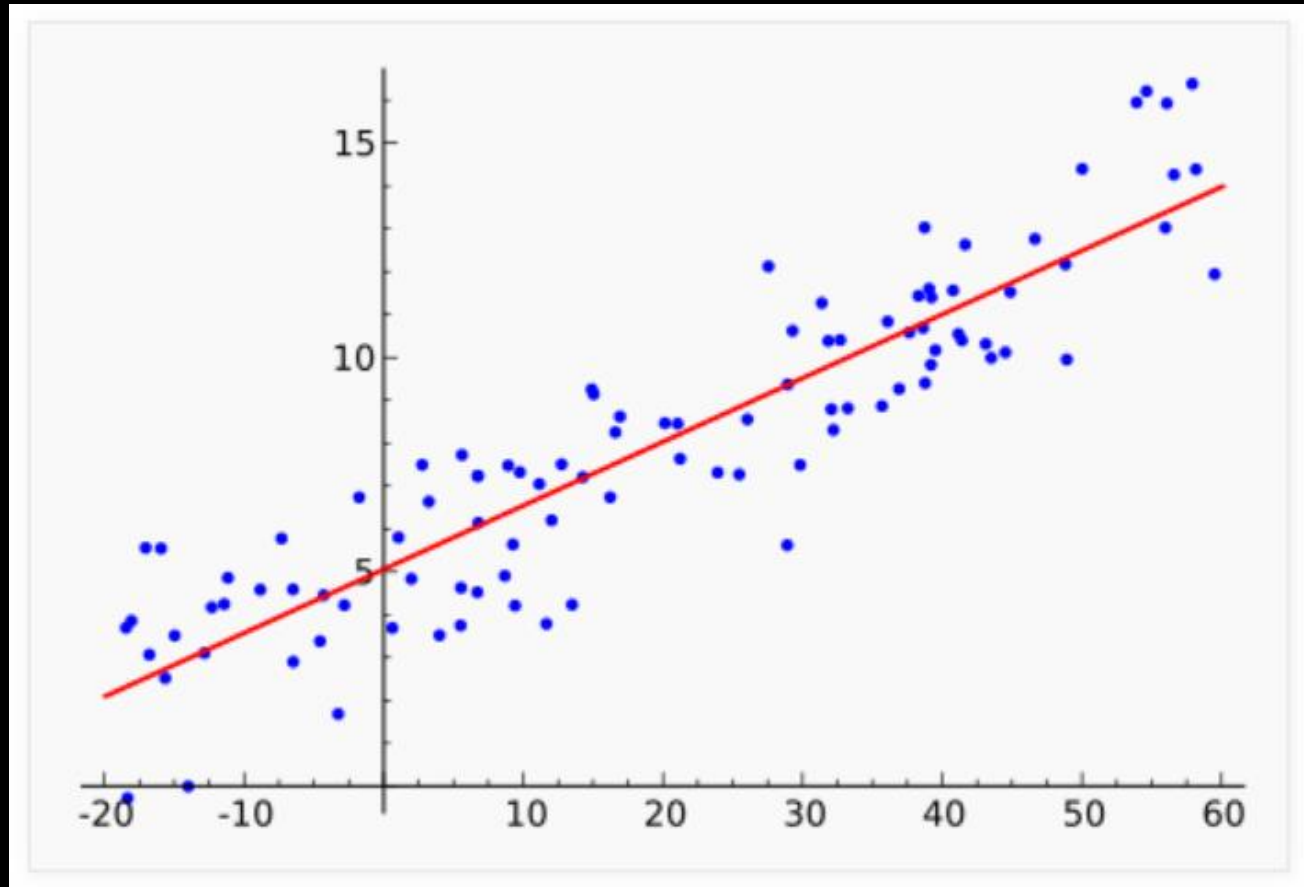
How do we
put the
model in
production?

What is an ML model?

Linear Regression – Housing Prices

The training finds a and b such that

$$Y = a + bX + \epsilon$$



Deep Learning

The input and output may be vectors \bar{X} , \bar{Y}

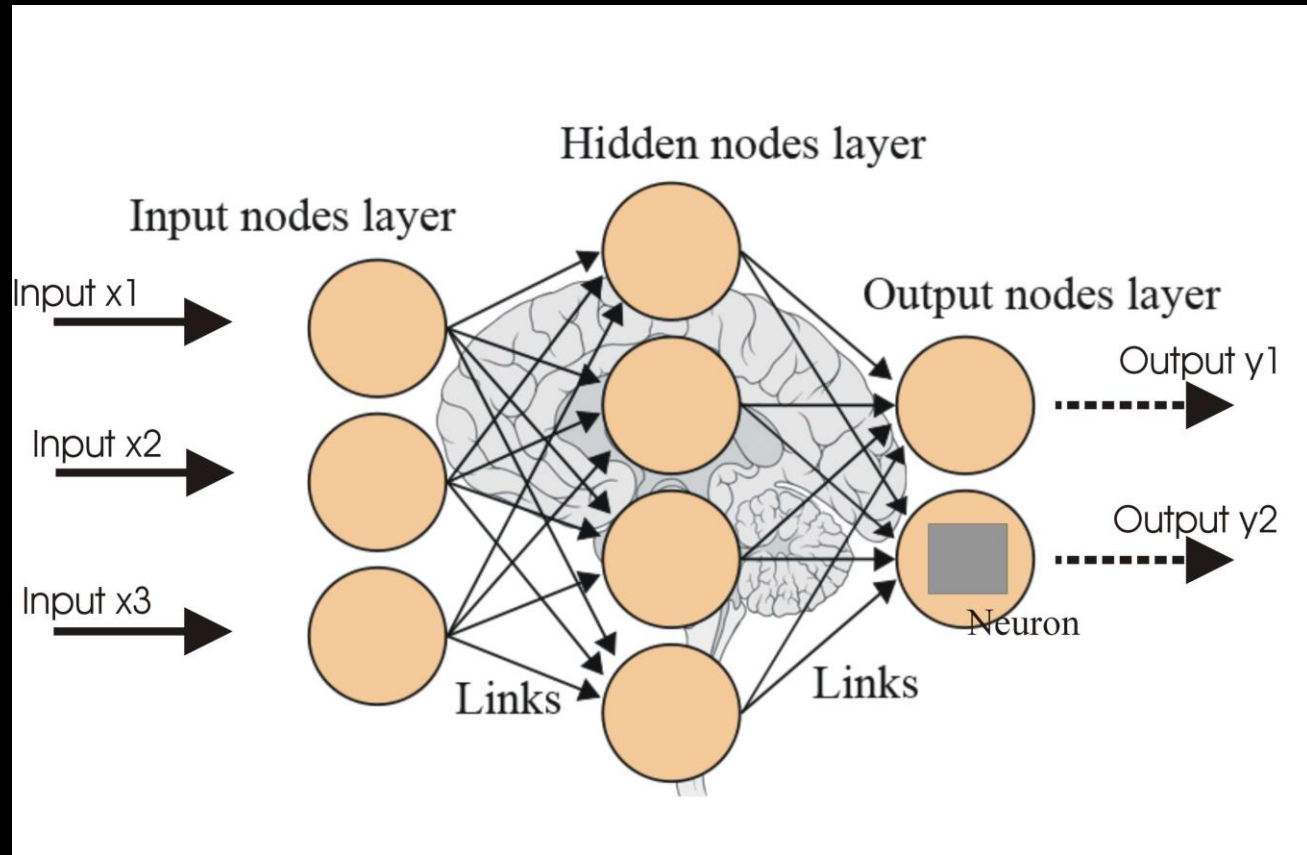
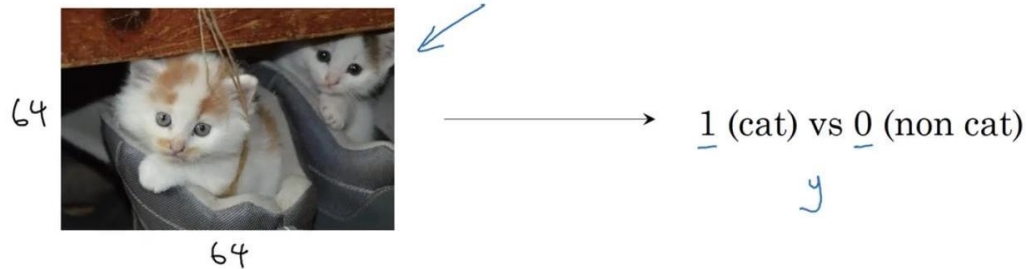
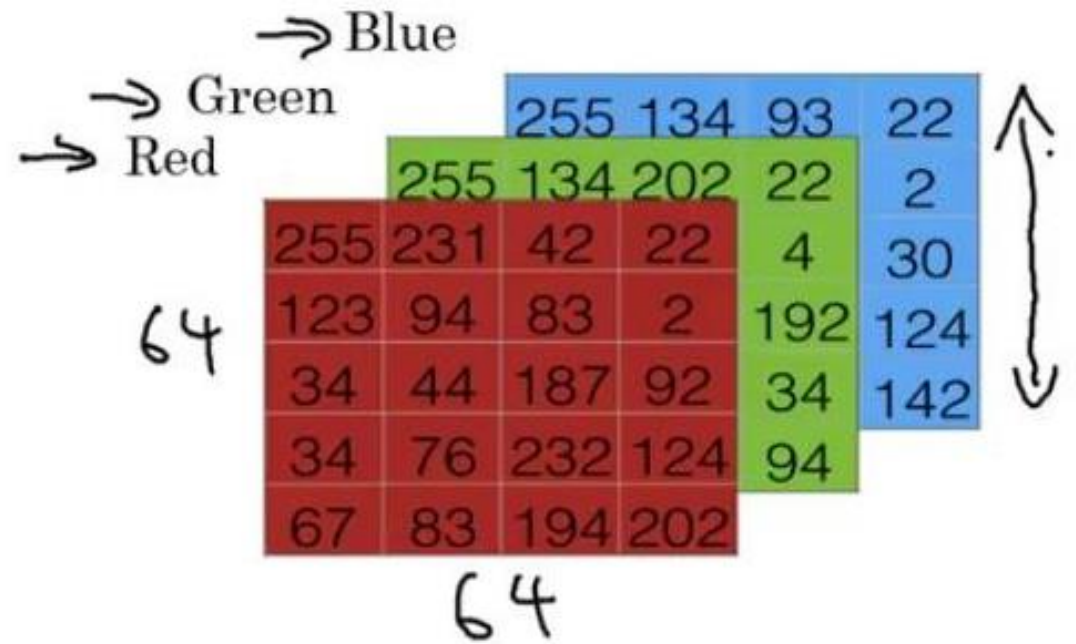


Image Classification

Binary Classification



To store an image your computer stores three separate matrices corresponding to the red, green, and blue color channels of this image.



Andrew Ng

ML Model

A definition of the **mathematical formula** with a number of parameters that are learned from the data



Isn't this
just an API
endpoint?!

Do models really change that often?

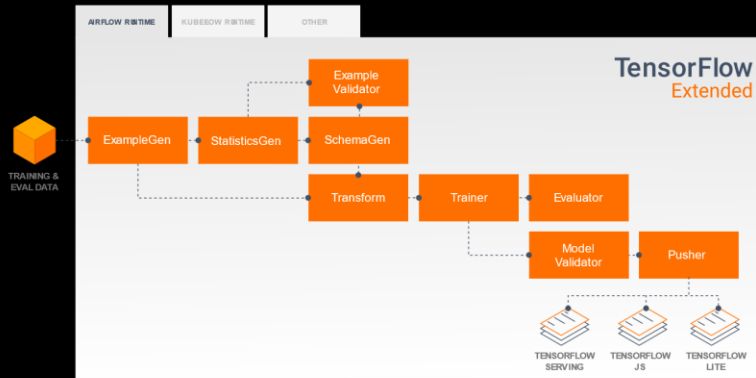
Models must be improved continuously

↑ Posted by [REDACTED] 3 years ago 📄

12.7k ↓ Facebook's list of "suggested friends" is quite literally a list of people I've been avoiding my entire life.

💬 523 Comments ➦ Share 📌 Save 🚫 Hide 🚩 Report

95% Upvoted

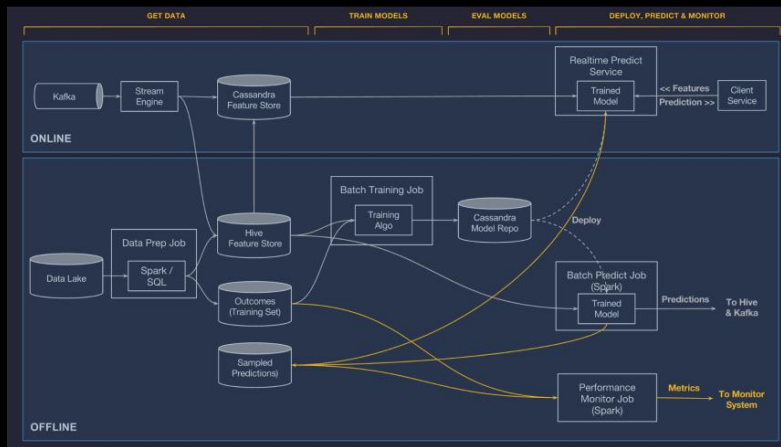


TensorFlow Extended

The screenshot shows the FBLearner interface with a table of workflow runs. The table has the following columns: ID, Owner, Workflow, Name, Progress, Start Time, Tags, Log Loss, and AUC. The data rows are as follows:

ID	Owner	Workflow	Name	Progress	Start Time	Tags	Log Loss	AUC
1047155	Mahaveer Jain	Parameter Sweep Example		59, 9 06pm		@@@-Demo	-	-
1047268	Mahaveer Jain	Gradient Boosted Decision Tree Training	Learning Rate: 0.35	100%	9/6, 9 19pm	-	0.00105	0.95524
1047297	Mahaveer Jain	Gradient Boosted Decision Tree Training	Learning Rate: 0.25	100%	9/6, 9 19pm	-	0.00107	0.95776
1047295	Mahaveer Jain	Gradient Boosted Decision Tree Training	Learning Rate: 0.3	100%	9/6, 9 19pm	-	0.00104	0.95719
1047295	Mahaveer Jain	Gradient Boosted Decision Tree Training	Learning Rate: 0.1	100%	9/6, 9 19pm	-	0.00122	0.95871
1047294	Mahaveer Jain	Gradient Boosted Decision Tree Training	Learning Rate: 0.2	100%	9/6, 9 19pm	-	0.00109	0.95796
1047293	Mahaveer Jain	Gradient Boosted Decision Tree Training	Learning Rate: 0.15	100%	9/6, 9 19pm	-	0.00115	0.95887
1047292	Mahaveer Jain	Gradient Boosted Decision Tree Training	Learning Rate: 0.4	100%	9/6, 9 19pm	-	0.00106	0.95355
1047291	Mahaveer Jain	Gradient Boosted Decision Tree Training	Learning Rate: 0.45	100%	9/6, 9 19pm	-	0.00110	0.95293
1037778	Jason Bencio	Parameter Sweep Example		100%	9/6, 2 30pm	-	-	-
950428	Li Zhang	Parameter Sweep Example		100%	8/21, 2 40pm	-	-	-
800873	Jawel Chen	Parameter Sweep Example		100%	8/6, 9 11pm	-	-	-
832281	Giri Rajaram	Parameter Sweep Example		100%	7/24, 12 56pm	-	-	-
832027	Giri Rajaram	Parameter Sweep Example		100%	7/24, 12 34pm	-	-	-

FBLearner Flow



Uber's Michelangelo

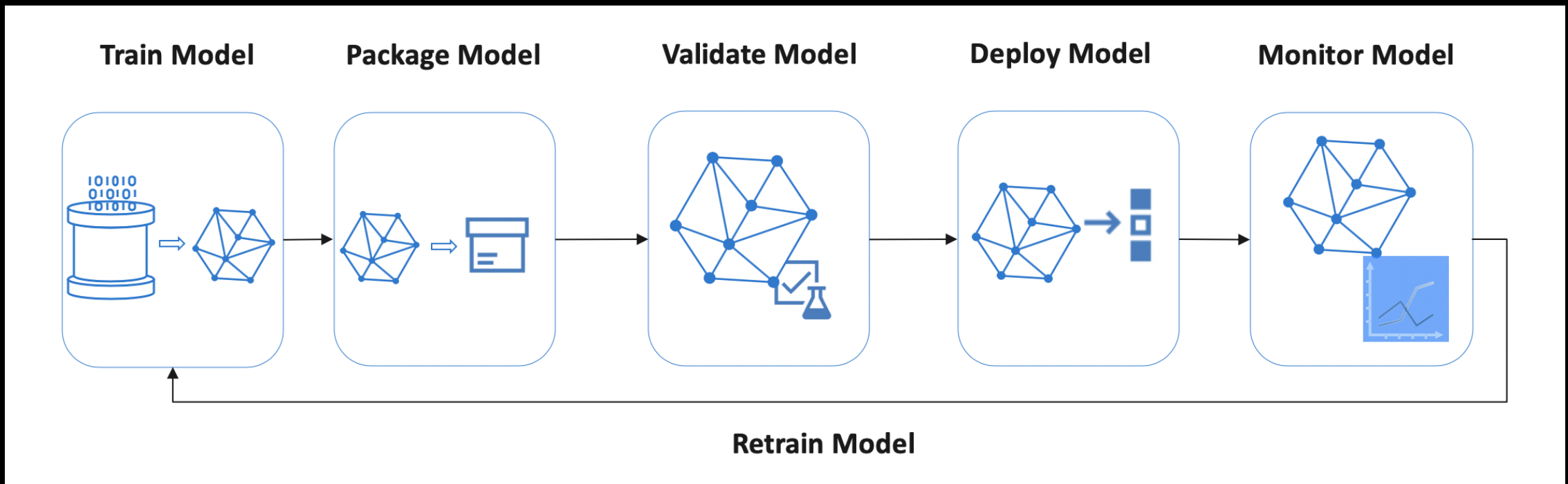


Microsoft Aether

But I don't work at a
big company with
thousands of
ML engineers!

How do we iterate?

Machine Learning Lifecycle





Data Scientist

- Quick iteration
- Versioning
- Reuse
- Great tools
- Ease of management
- Unlimited scale
- Eliminating drift



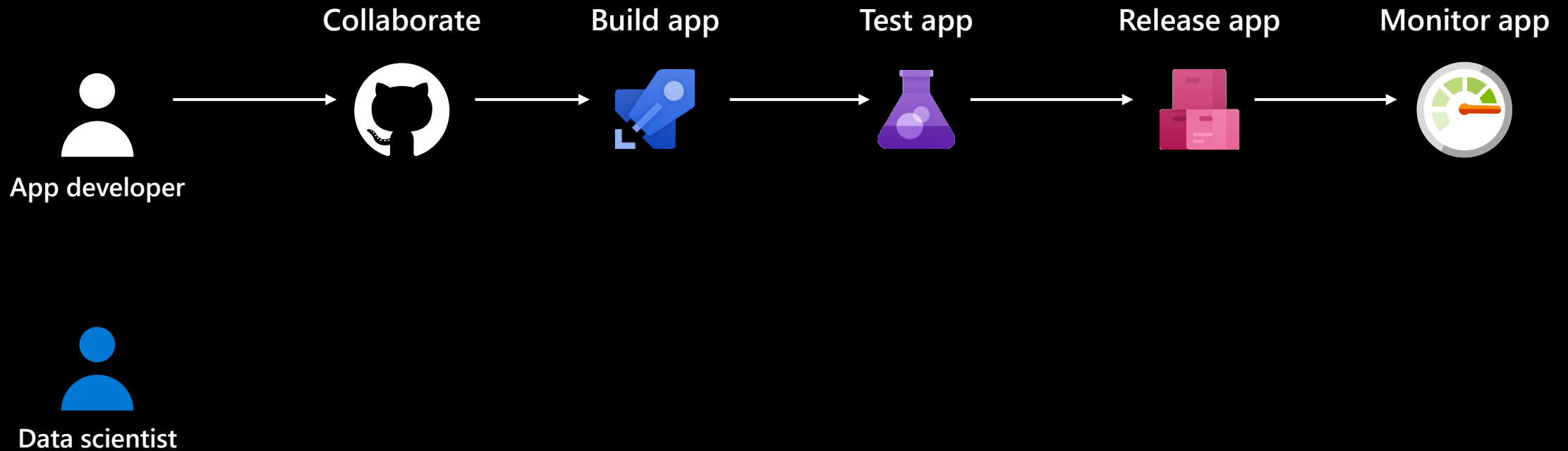
Friends?



DevOps/SRE

- Quick iteration
- Versioning
- Reuse
- Compliance
- Observability
- Uptime
- Updates

MLOps Workflow



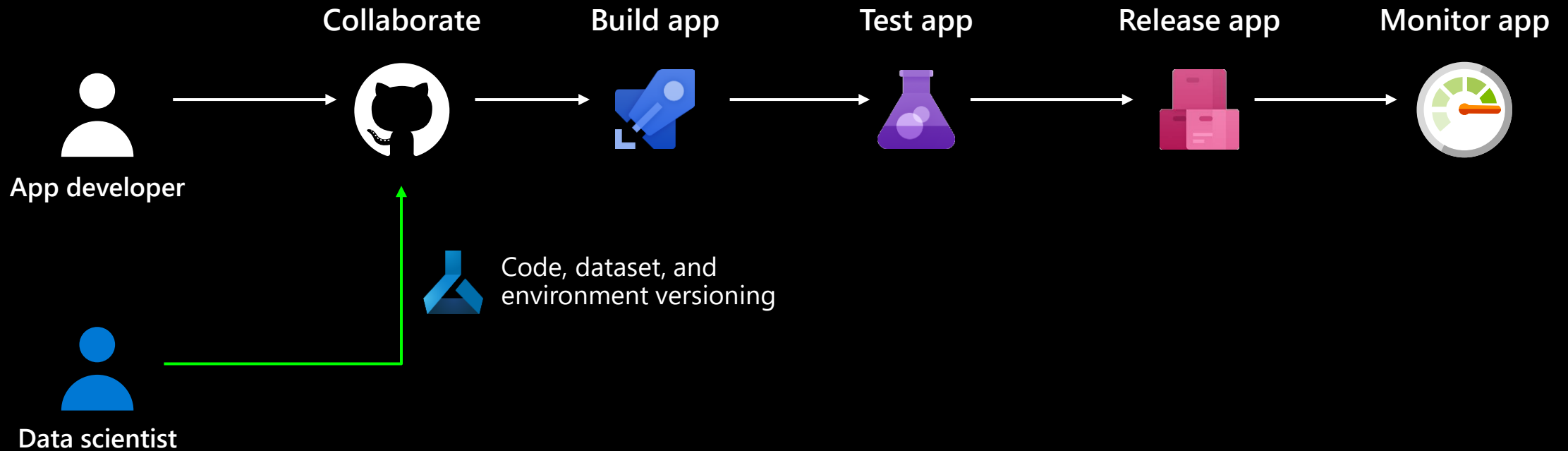
 Model reproducibility

 Model validation

 Model deployment

 Model retraining

MLOps Workflow



Model reproducibility



Model validation

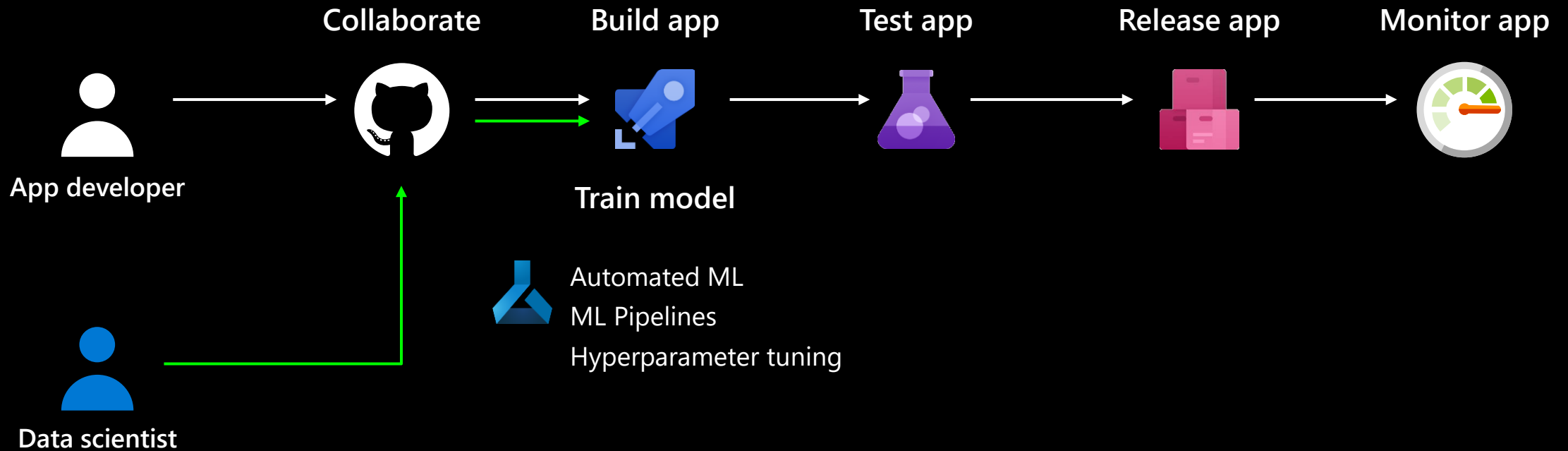


Model deployment



Model retraining

MLOps Workflow



Model reproducibility



Model validation

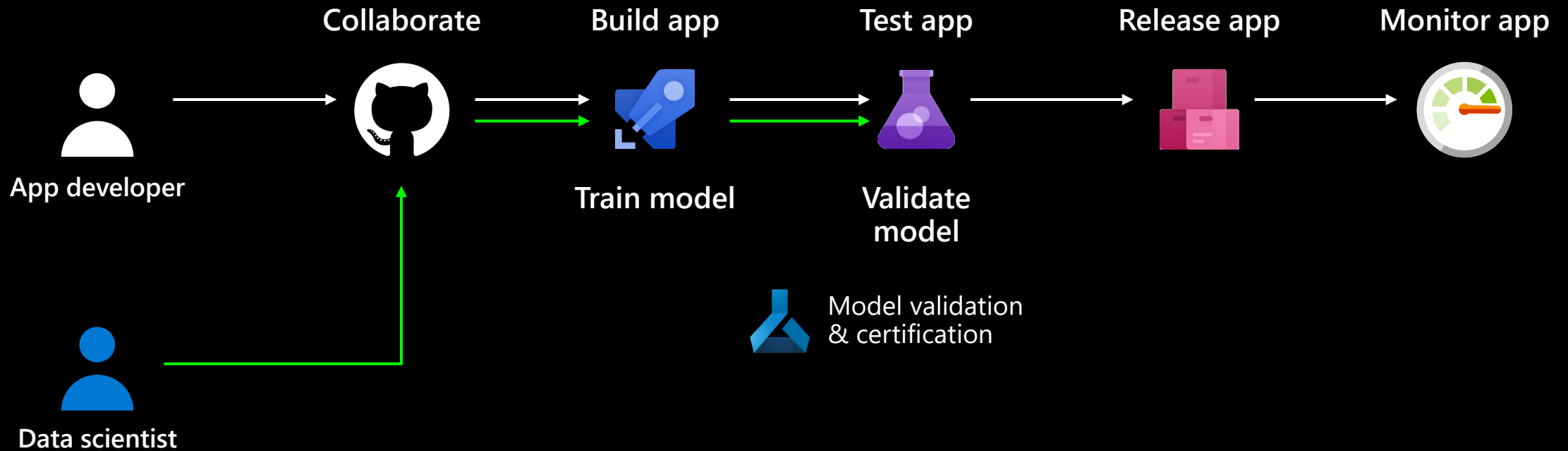


Model deployment



Model retraining

MLOps Workflow



Model reproducibility



Model validation

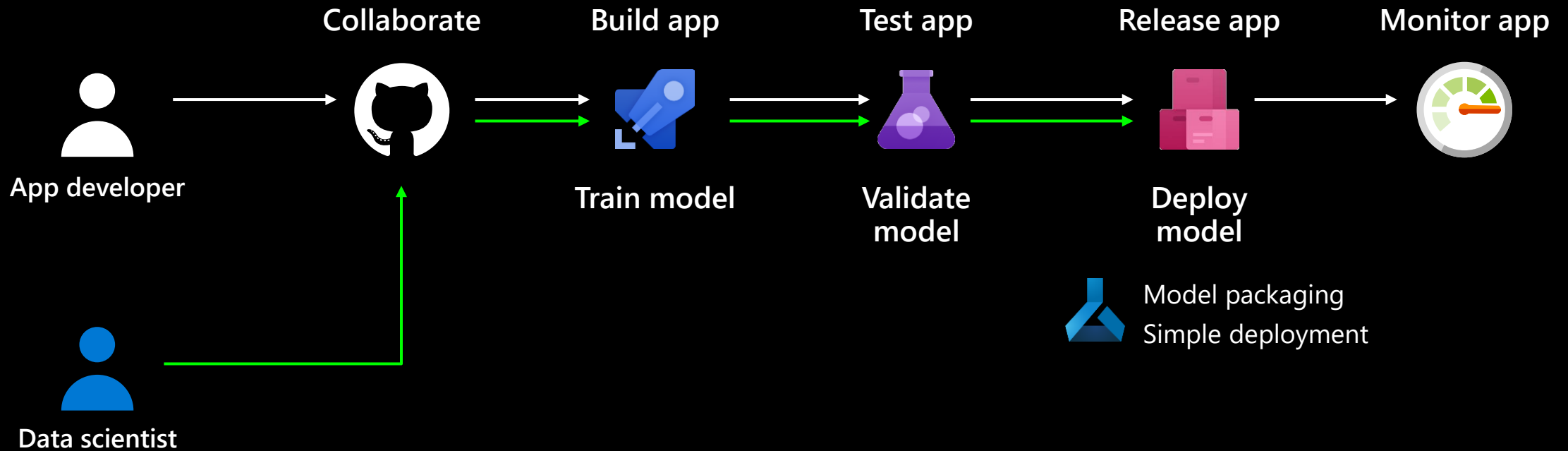


Model deployment



Model retraining

MLOps Workflow



Model reproducibility



Model validation

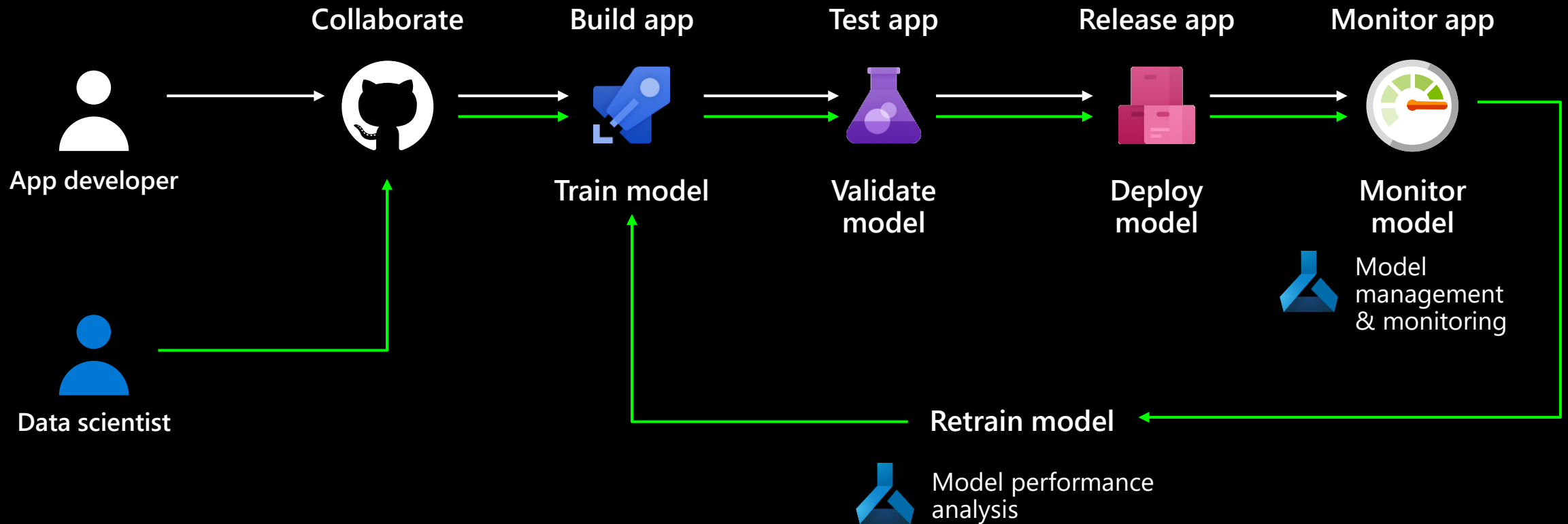


Model deployment



Model retraining

MLOps Workflow



Model reproducibility



Model validation

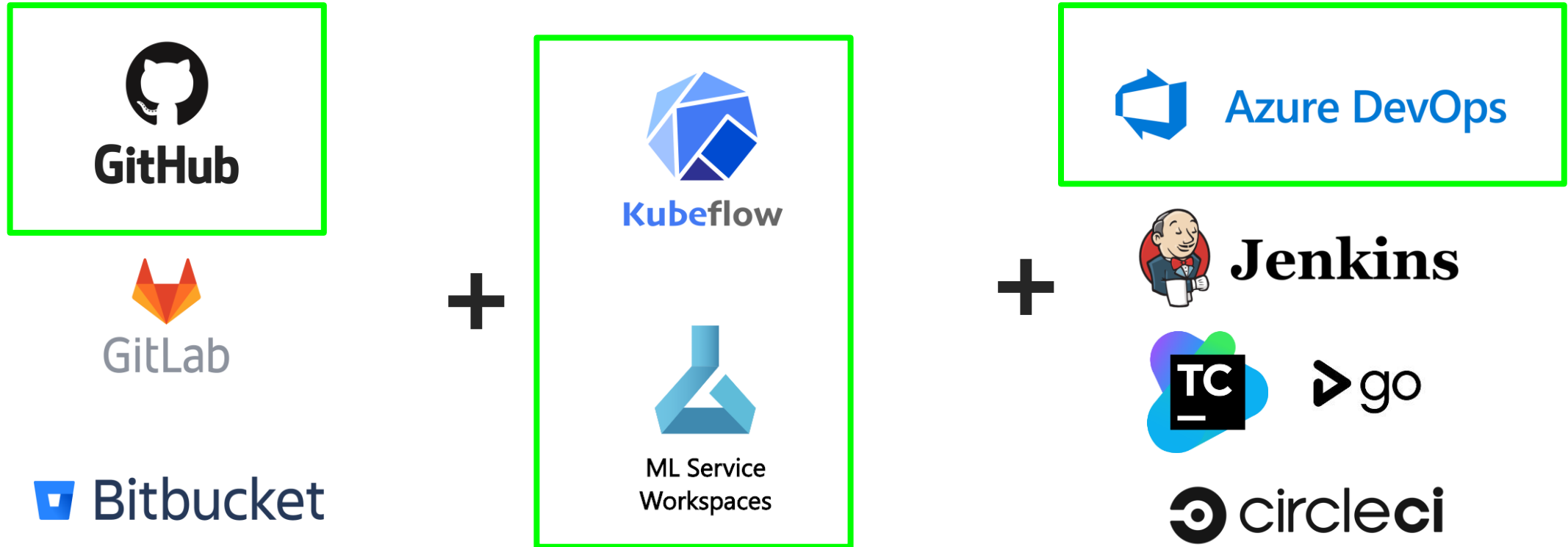


Model deployment



Model retraining

Build Your Own MLOps Platform

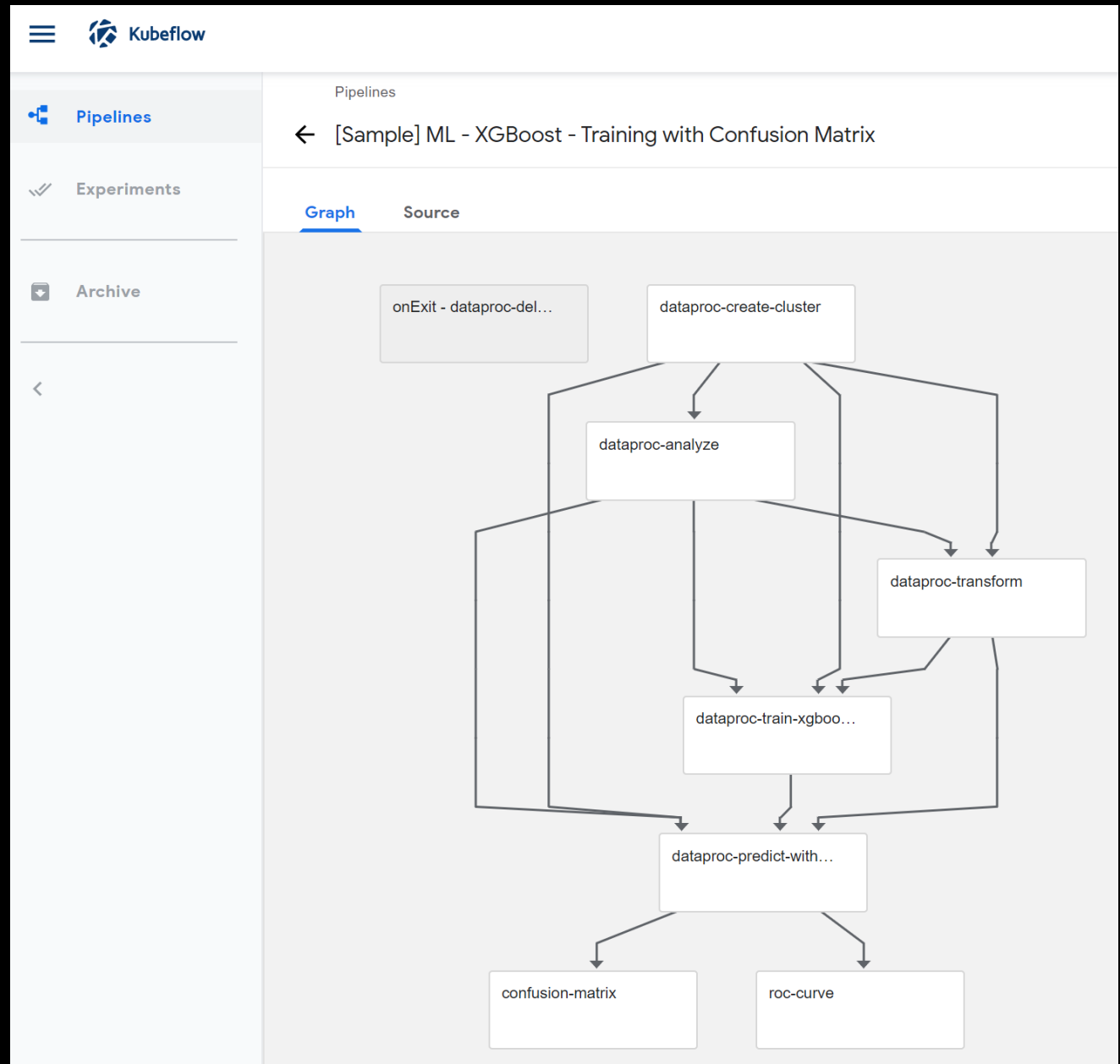


ML Pipeline

A reusable, scaleable ML workflow template

Kubeflow pipeline

A reusable, scalable ML workflow template that runs on containers



Azure ML

- Prep data
- Train
- Test
- Deploy
- Manage

Preview Microsoft Azure | Machine learning

kubeflow-aml > Home

Welcome!

Home

Author

- Automated ML
- Visual Interface
- Notebooks

Assets

- Datasets
- Experiments
- Models
- Endpoints

Manage

- Compute
- Datstores

Create new ▾

Automated ML
Automatically train and tune a model using a target metric.
[Start now](#)

Visual interface
Drag-and-drop interface from prepping data to deploying models.
[Start now](#)

Notebooks
Code with Python SDK and run sample experiments.
[Start now](#)

My recent resources


Runs

Run Number	Experiment	Updated Time	Status
1	foodprofilev3-ImagePr...	9/22/2019, 3:17:39 PM	NotStarted
1	foodprofilev1-ImagePr...	9/20/2019, 6:27:50 PM	NotStarted

The CI/CD Pipeline

GitHub Repo

<https://github.com/DivineOps/kubeflow-and-mlops>

 Azure Pipelines succeeded

MLOps with Kubeflow, Azure Machine Learning and Azure Pipelines

This repository provides a sample ML CI/CD pipeline using Kubeflow, Azure ML workspaces and Azure Pipelines.









It requires:

- An Azure Account (A trial account works!)
- An Azure DevOps Organization (The free tier works!)

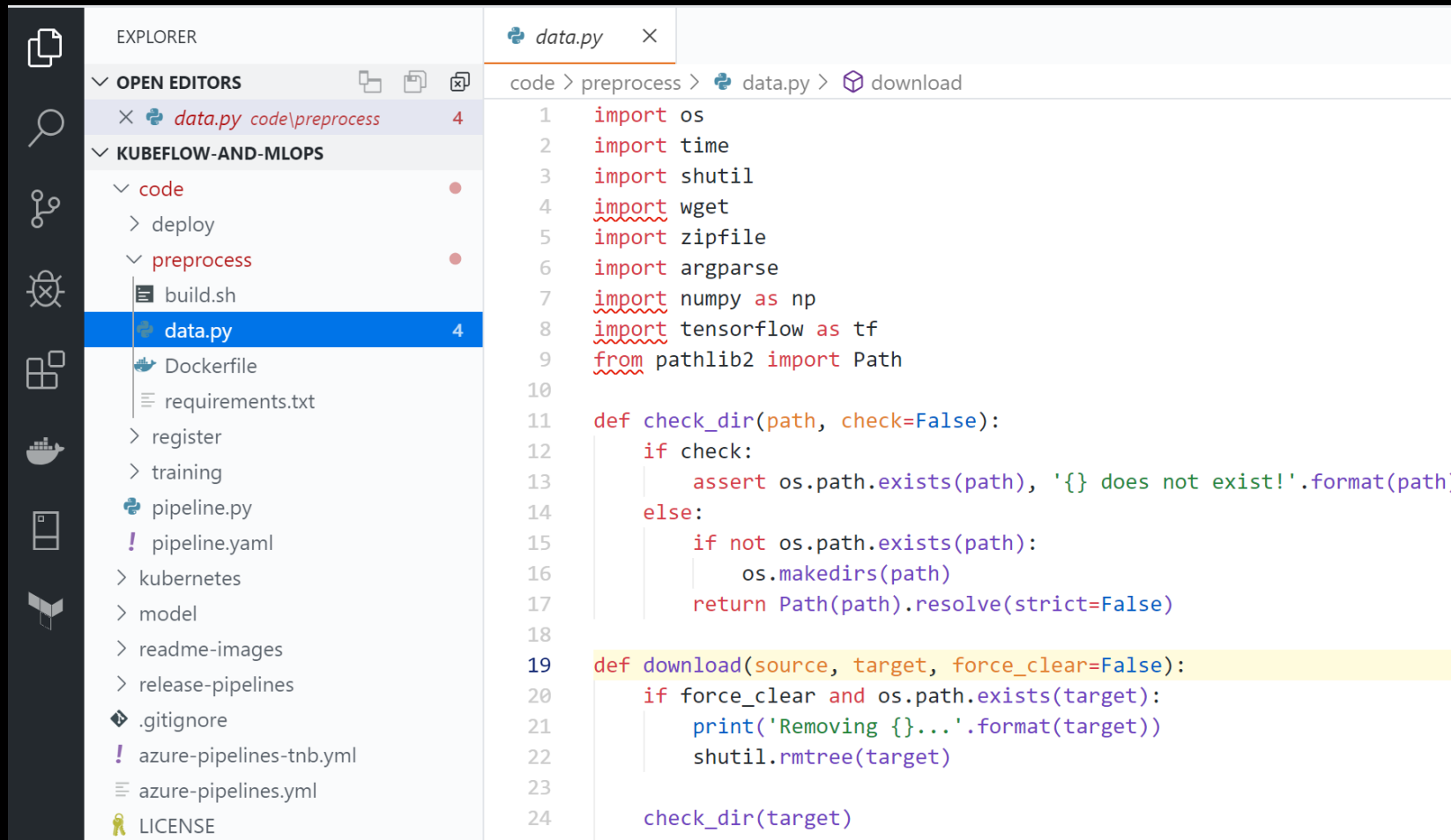
Slides

Link to the slides to the All Things Open version of this talk [Slides](#)

Azure Resources

<input type="checkbox"/>  aks-cluste	Kubernetes service
<input type="checkbox"/>  fooddeplc	Container instances
<input type="checkbox"/>  kf-aks	Kubernetes service
<input type="checkbox"/>  kubeflowa	Container registry
<input type="checkbox"/>  kubeflow-	Machine Learning service workspace
<input type="checkbox"/>  kubeflowa	Storage account
<input type="checkbox"/>  kubeflowa	Key vault
<input type="checkbox"/>  kubeflowa	Application Insights

The Code



```
code > preprocess > data.py > download
1  import os
2  import time
3  import shutil
4  import wget
5  import zipfile
6  import argparse
7  import numpy as np
8  import tensorflow as tf
9  from pathlib2 import Path
10
11 def check_dir(path, check=False):
12     if check:
13         assert os.path.exists(path), '{} does not exist!'.format(path)
14     else:
15         if not os.path.exists(path):
16             os.makedirs(path)
17         return Path(path).resolve(strict=False)
18
19 def download(source, target, force_clear=False):
20     if force_clear and os.path.exists(target):
21         print('Removing {}'.format(target))
22         shutil.rmtree(target)
23
24     check_dir(target)
```

Build Pipeline

<https://github.com/DivineOps/kubeflow-and-mlops>

```
stages:
- stage: ContainerConfig
  displayName: 'Configure and Register Containers'
  jobs:
  - job: Containers
    pool:
      name: Hosted Ubuntu 1604
    steps:
    - task: AzureCLI@1
      inputs:
        azureSubscription: $(KF_SUBSCRIPTION_ID)
        scriptLocation: 'inlineScript'
        inlineScript: |
          az acr login -n $(KF_ACR)
          cd code
          cd preprocess
          docker build -t $(KF_ACR).azurecr.io/preprocess:$BUILD_SOURCEVERSION .
          docker push $(KF_ACR).azurecr.io/preprocess:$BUILD_SOURCEVERSION
        displayName: 'Build & Push Preprocess Image'

    - task: AzureCLI@1
      inputs:
        azureSubscription: $(KF_SUBSCRIPTION_ID)
        scriptLocation: 'inlineScript'
        inlineScript: |
          cd code
          cd training
          docker build -t $(KF_ACR).azurecr.io/training:$BUILD_SOURCEVERSION .
          docker push $(KF_ACR).azurecr.io/training:$BUILD_SOURCEVERSION
        displayName: 'Build & Push Training Image'

    - task: AzureCLI@1
      inputs:
        azureSubscription: $(KF_SUBSCRIPTION_ID)
        scriptLocation: 'inlineScript'
        inlineScript: |
          cd code
          cd register
          docker build -t $(KF_ACR).azurecr.io/register:$BUILD_SOURCEVERSION .
          docker push $(KF_ACR).azurecr.io/register:$BUILD_SOURCEVERSION
        displayName: 'Build & Push Register Image'
```

```
- stage: KubeflowTrigger

variables:
  KF_BATCH: 32
  KF_EPOCHS: 5
  KF_LEARNING_RATE: 0.0001
  KF_MODEL_NAME: tacosandburritos
  KF_PERSISTENT_VOLUME_NAME: azure
  KF_PERSISTENT_VOLUME_PATH: /mnt/azure

dependsOn: ContainerConfig
displayName: 'Trigger Kubeflow Pipeline'
jobs:
- job: Kubeflow
  pool:
    name: Hosted Ubuntu 1604
  steps:
  - task: AzureCLI@1
    env:
      KF_MAPPED_SERVICE_PRINCIPAL_PASSWORD: $(KF_SERVICE_PRINCIPAL_PASSWORD)
    inputs:
      azureSubscription: $(KF_SUBSCRIPTION_ID)
      scriptLocation: 'inlineScript'
      inlineScript: |
        az aks get-credentials -g $(KF_RESOURCE_GROUP) -n $(KF_AKS_CLUSTER)
        kubectrl port-forward --namespace kubeflow svc/ml-pipeline 8888:8888 &
        kubepid=$!

        sudo apt-get install python3-setuptools
        pip3 install wheel
        pip3 install kfp

        touch script.py
        echo "import kfp" >> script.py
        echo "client = kfp.Client(host='localhost:8888')" >> script.py
        echo "client.run_pipeline('${KF_EXPERIMENT_ID}', 'Run ${BUILD_BUILDID}', params={'imagetag': '${BUILD_SOURCEVERSION}}',"

        cat script.py

        python3 script.py

        kill $kubepid
      displayName: 'Trigger Kubeflow Pipeline'
```


Experiments > Azure Experiment

← ✓ Run 703

Graph Run output Config

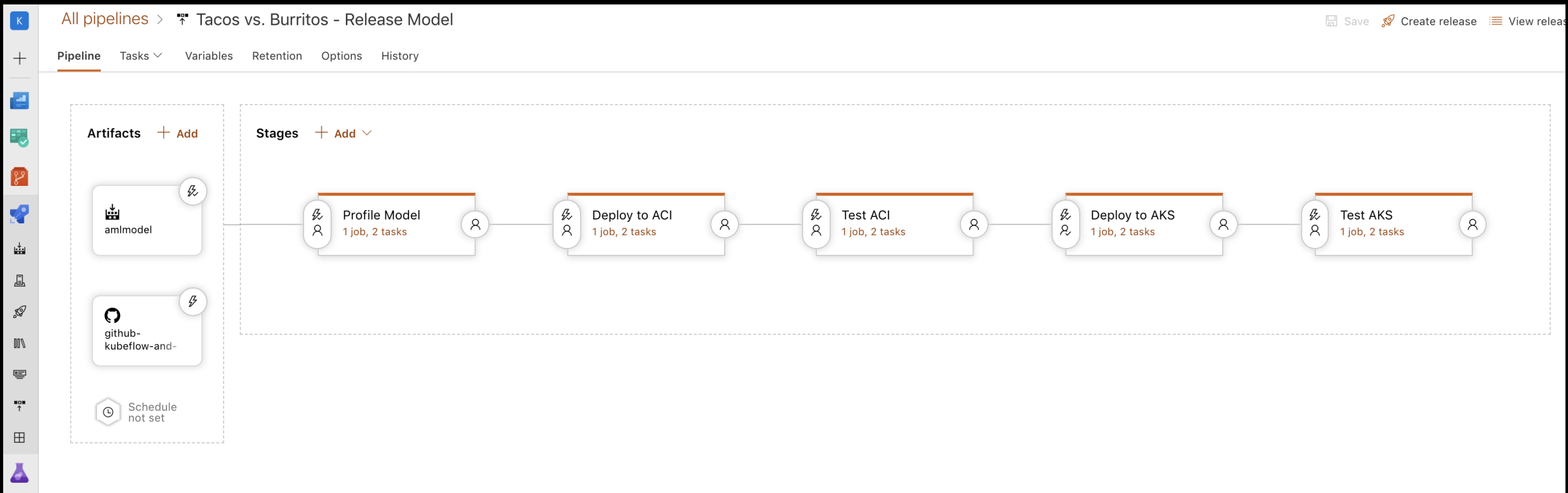
```
graph TD; preprocess[preprocess] --> training[training]; training --> register[register];
```

Kubeflow Pipeline

<https://www.kubeflow.org/docs/azure/azureendtoend/>

Release Pipeline

https://dev.azure.com/sasrose/kubeflow/_release?_a=releases&view=mine&definitionId=3



Even a simple CI/CD pipeline is
better than none!

AI Ethics

Bias is a property of
information



Build AI responsibly!

Thank You!

@DivineOps

Resources

GitHub repo

<https://www.kubeflow.org/docs/azure/azureendtoend/>

Deploy Kubeflow on Azure

<https://www.kubeflow.org/docs/azure/deploy/install-kubeflow/>

Example Kubeflow Azure Pipeline

<https://www.kubeflow.org/docs/azure/azureendtoend/>

Release pipeline

https://dev.azure.com/sasrose/kubeflow/_release