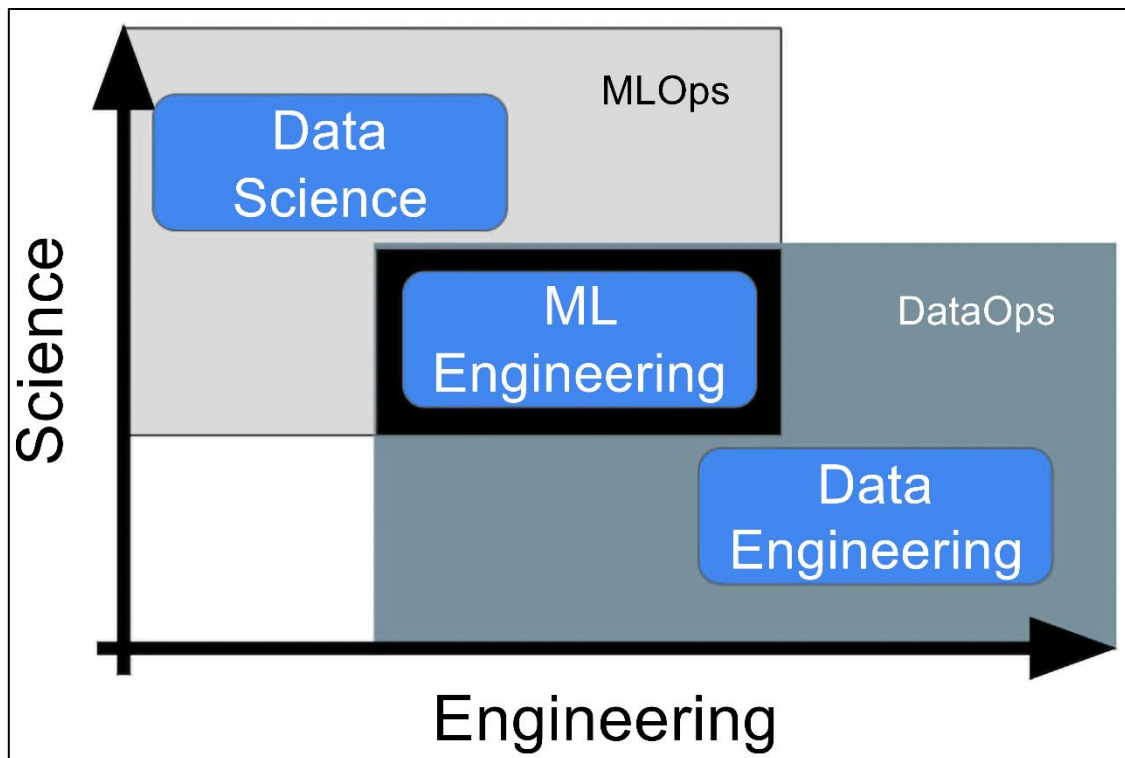
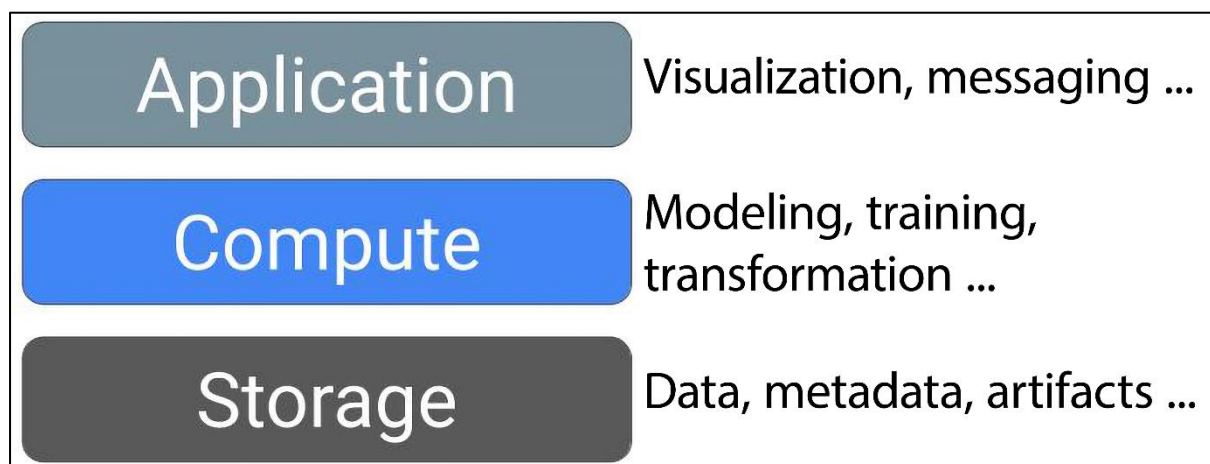
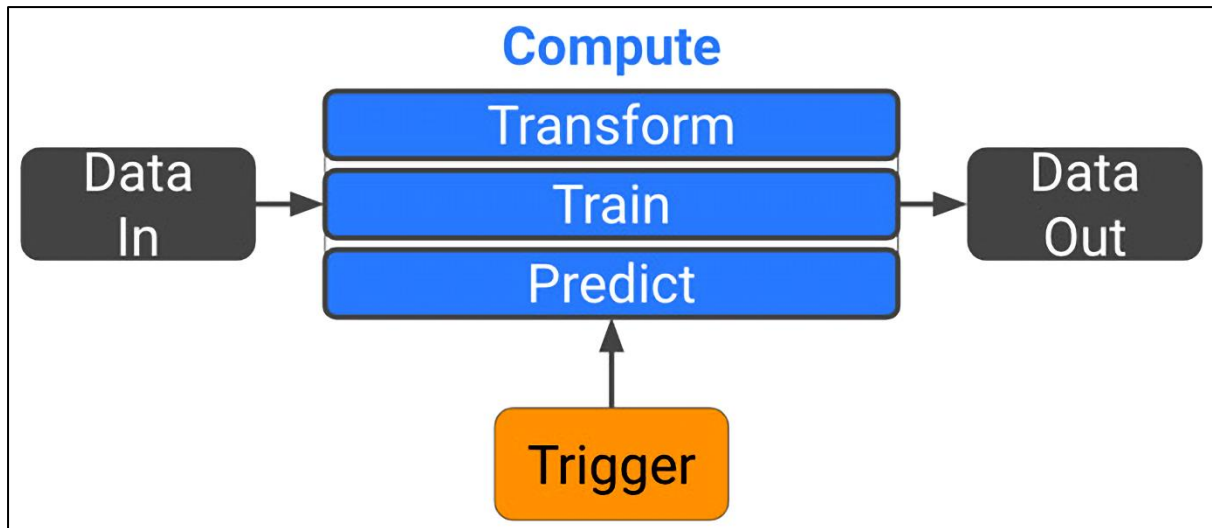
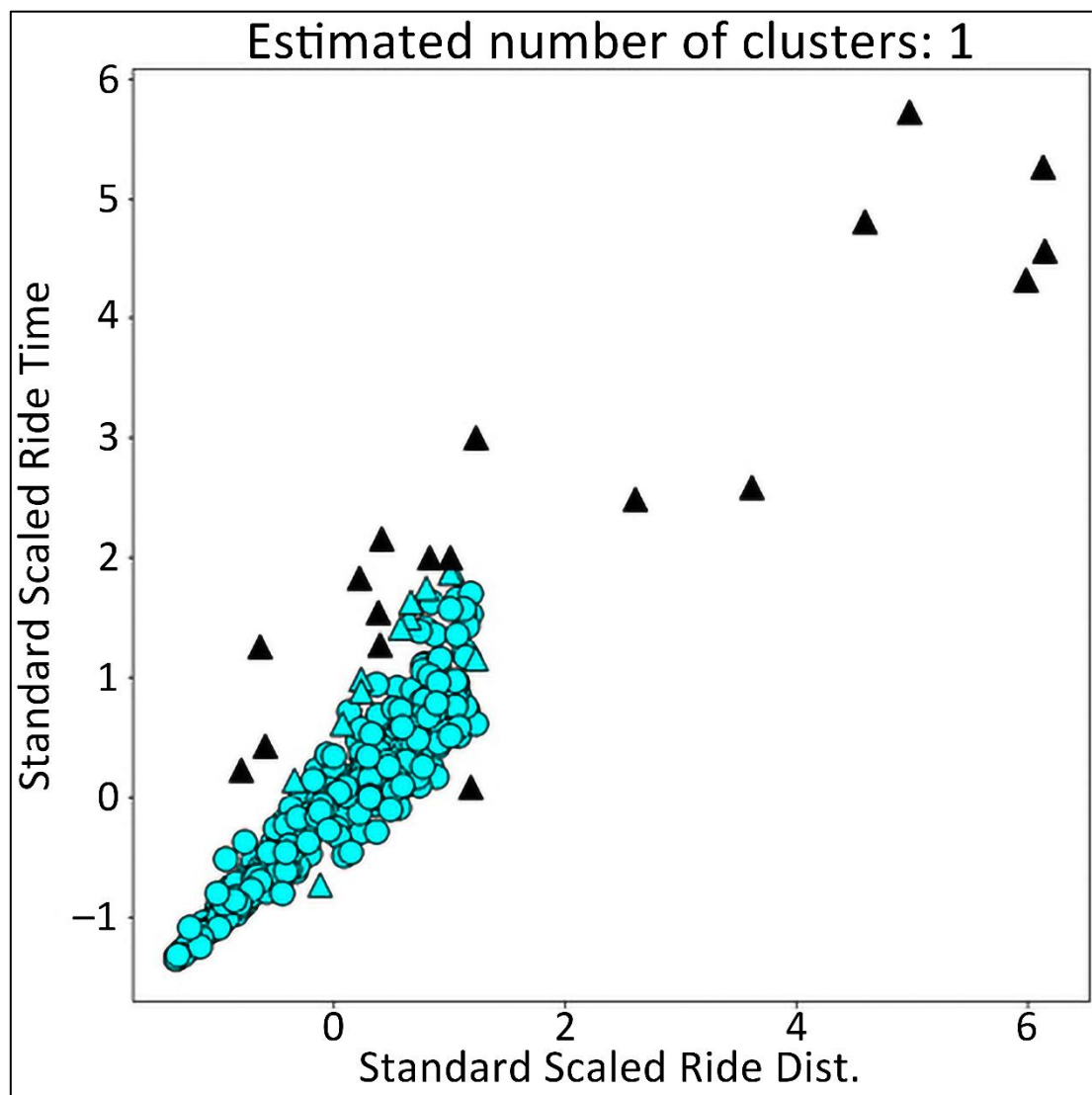
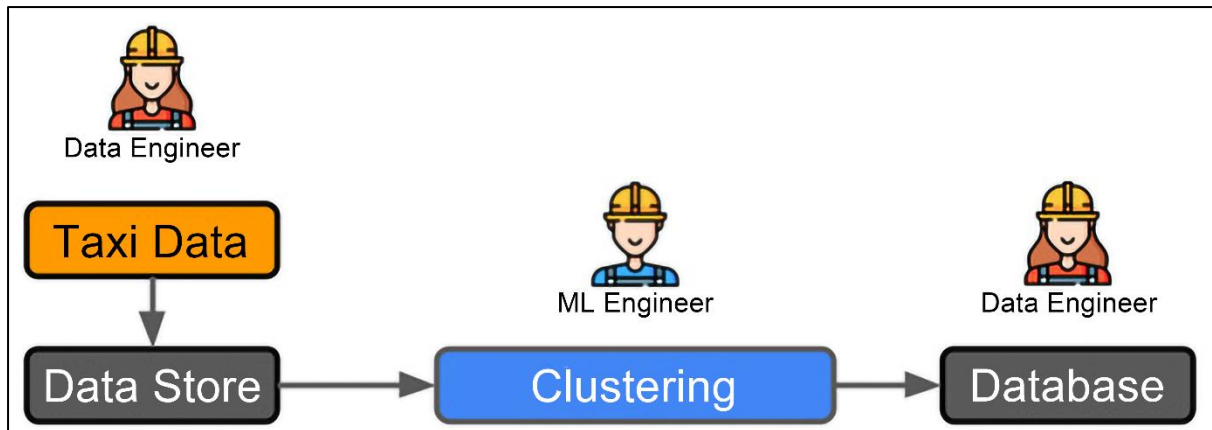
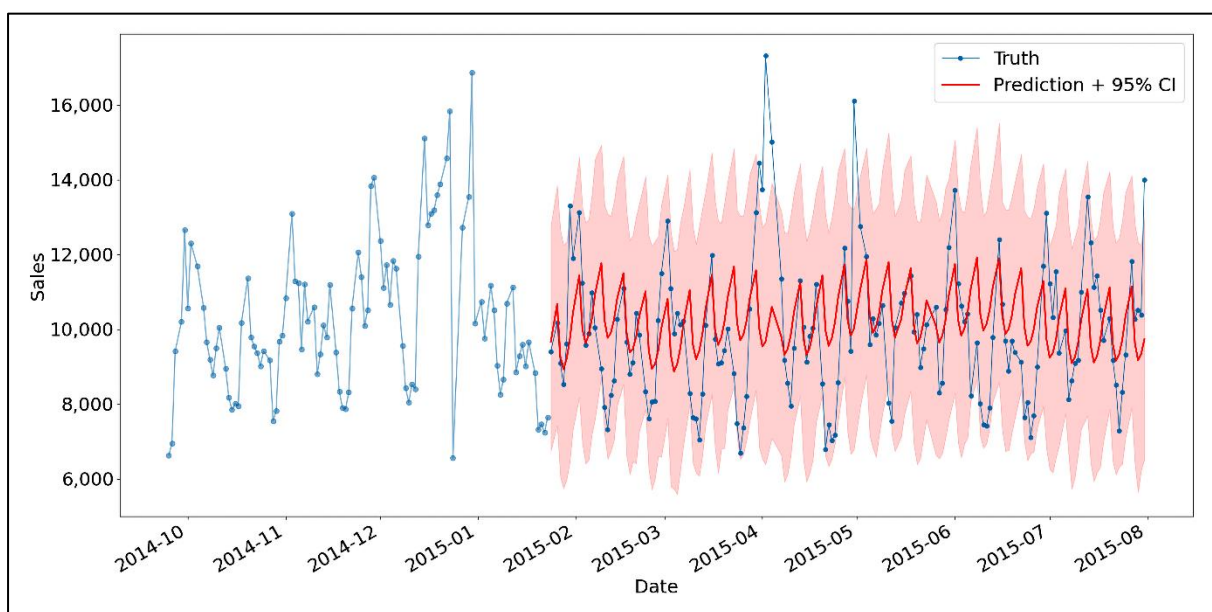
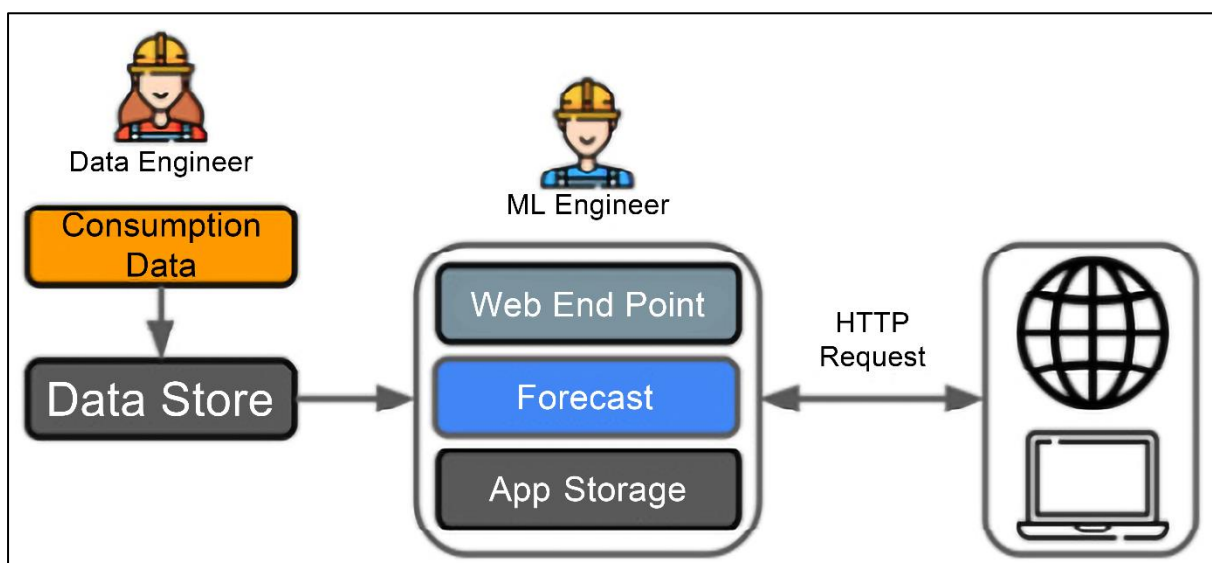
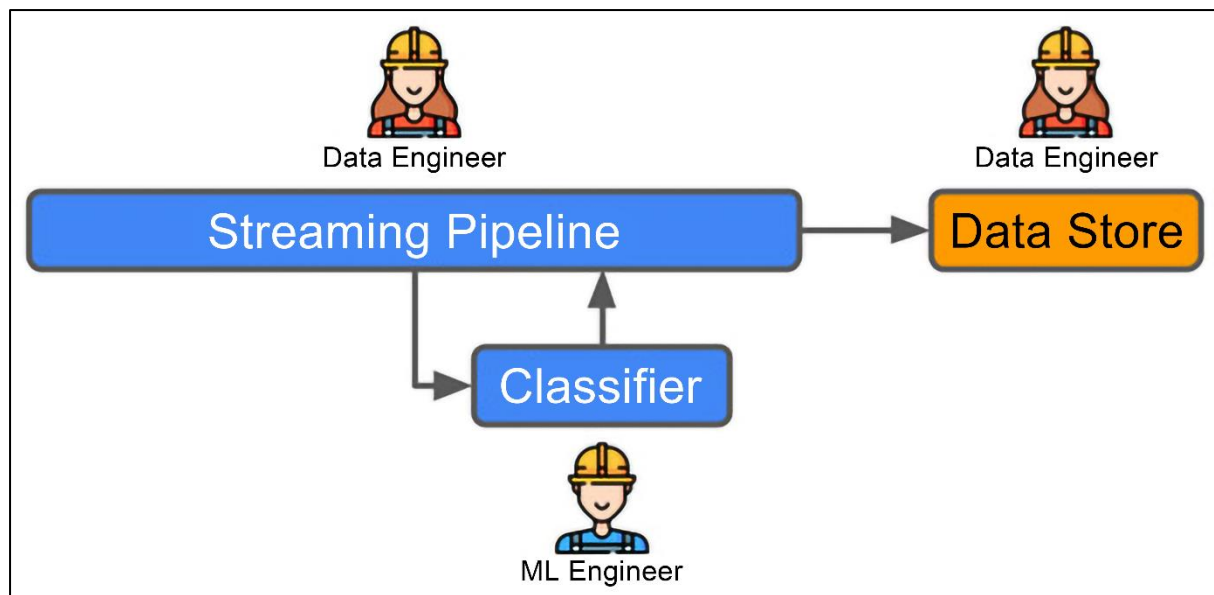


Chapter 1: Introduction to ML Engineering

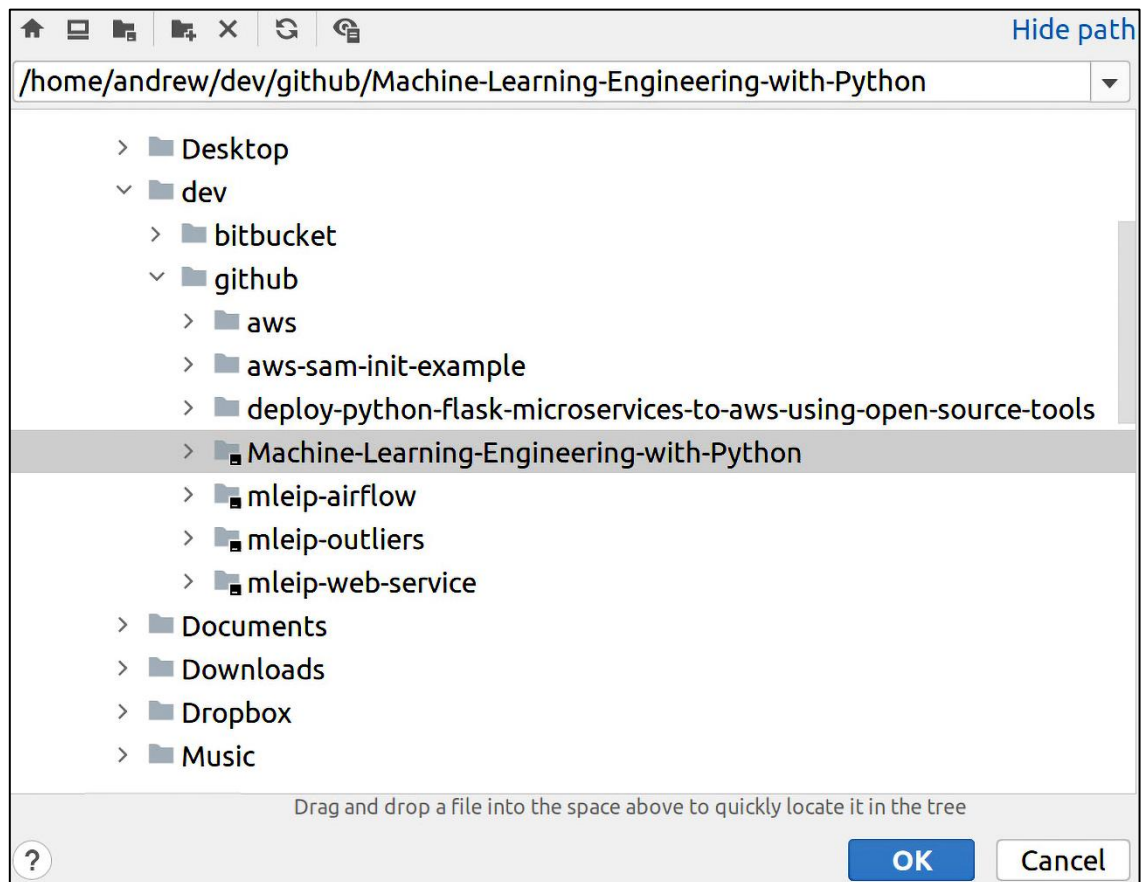








Chapter 2: The Machine Learning Development Process



Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

Owner *

Repository name *



AndyMc629 ▾



Great repository names are short and memorable. Need inspiration? How about **bookish-goggles?**

Description (optional)



Public

Anyone on the internet can see this repository. You choose who can commit.



Private

You choose who can see and commit to this repository.

Initialize this repository with:

Skip this step if you're importing an existing repository.

☐ **Add a README file**

This is where you can write a long description for your project. [Learn more](#).

☐ **Add .gitignore**

Choose which files not to track from a list of templates. [Learn more](#).

Keymap

> Editor

Plugins

< Version Control

Background

Changelists

Commit

Confirmation

File Status Colors

Issue Navigation

Shelf

Git

GitHub

Mercurial

> Subversion

Version Control > GitHub



Andrew McMahon

AndyMc629

github.com

Projects / ml-engineering-in-python

MEIP board

Type

TO DO 4 ISSUES

As a store demand planner, I want to see forecasts for items split by region to anticipate extra orders that need to be made

MEIP-8

Add DBSCAN functionality to DetectionModels in outliers package

MEIP-26

IN PROGRESS 5 ISSUES

As a store demand planner, I want to be able to trigger retraining of models I think are out of date to improve forecast performance

MEIP-9

Build forecasting algorithm: build basic prophet algorithm (use code already developed)

MEIP-6

DONE

See all Done issues

AWS Management Console

AWS services

▼ Recently visited services

 Billing	 Elastic Container Registry	 IAM
 Managed Apache Airflow	 CloudFormation	 AWS Cost Explorer
 EC2	 Secrets Manager	 S3
 Elastic Container Service	 Certificate Manager	 Amazon SageMaker
 VPC	 Elastic Kubernetes Service	 EMR

► All services

Stay connected to your AWS resources on-the-go



AWS Console Mobile App now supports four additional regions. Download the AWS Console Mobile App to your iOS or Android mobile device. [Learn more](#) 

Explore AWS

Introducing AWS Backup Audit Manager

Maintain and demonstrate your data backup and compliance posture at scale. [Learn more](#) » 

[AWS Certification](#)

Discover

Play

Develop

Deploy

As a taxi ride analyst, I want to see anomalous journeys so that I can tailor our offers to customers



Description

Acceptance criteria (scenario):

Given I have access to data and/or visualizations of ML results,

where there are anomalous taxi rides,

then the system accurately identifies and labels these rides and I can see them

GitHub integration for Jira

[Branches](#) [Pull Requests](#) [Commits](#) [Tags](#)

Prep for Prophet

```
df.rename(columns= {'Datetime': 'ds', 'AEP_MW': 'y'}, inplace=True)
```

```
df['ds']=df['ds'].astype('datetime64[ns]')
```

```
df.dtypes
```

```
#Initialize Split Class, we'll split our data 5 times for cv  
ts_splits = TimeSeriesSplit(n_splits=5)
```

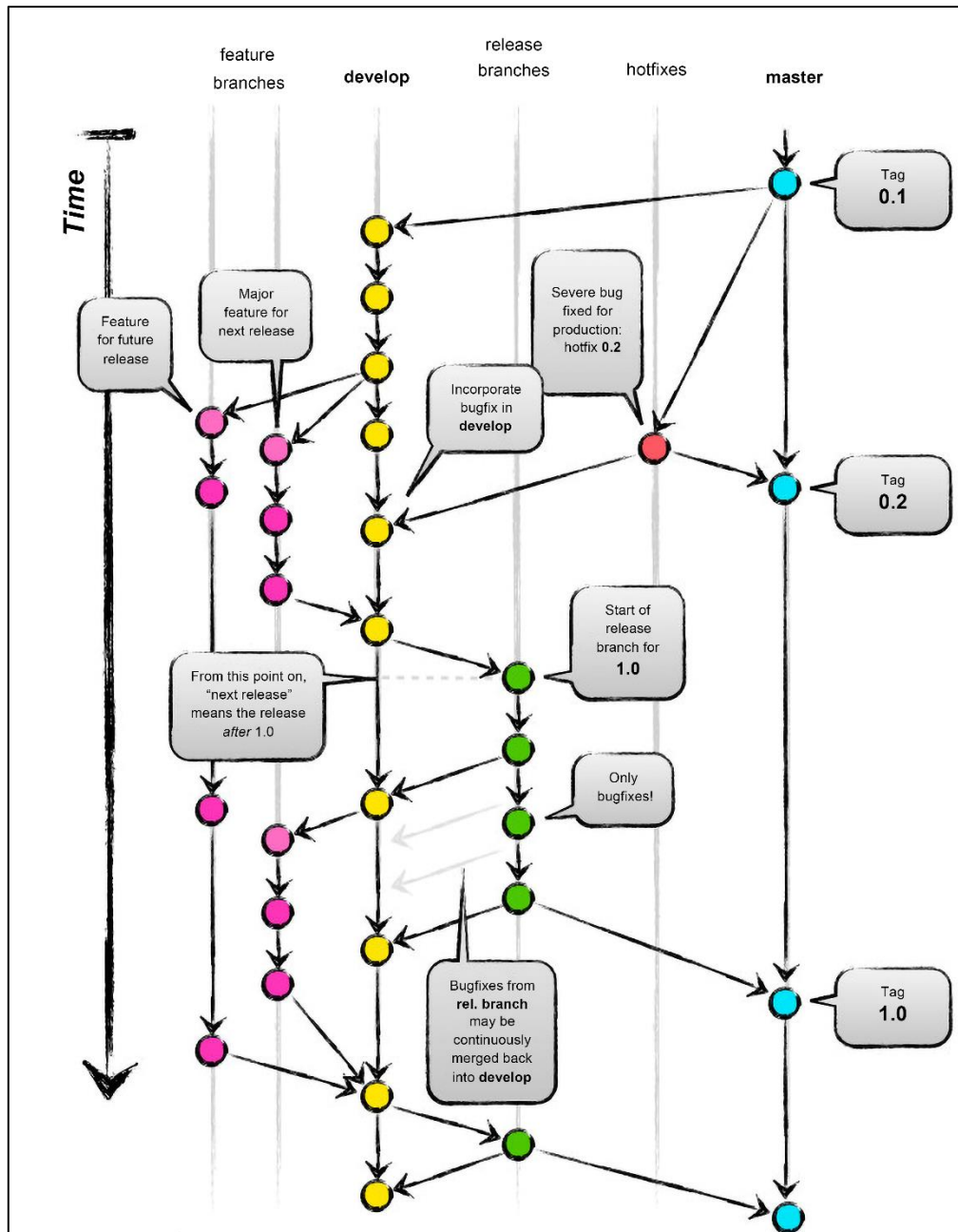
Train and Forecast

```
tmp = time_split_train_test(df.sort_values('ds', ascending=True).iloc[-1000:], ts_splits)
```

```
tmp.head()
```

Plot

```
nrow = 5; ncol = 1;  
fig, axs = plt.subplots(nrows=nrow, ncols=ncol, figsize=(20,30))  
fig.subplots_adjust(hspace=0.4, wspace=0.4)  
for i, ax in enumerate(fig.axes):  
    split_rmse = tmp[(tmp['split']==i) & (tmp['train']==False)]['rmse'].iloc[0]  
  
    ax.set_title('Split '+str(i)+' - RMSE: '+"{:.2f}".format(split_rmse))  
  
    tmp[(tmp['split']==i) & (tmp['train']==True)].plot(x='ds', y='y', ax=ax, color='blue', marker='o')  
    tmp[(tmp['split']==i) & (tmp['train']==False)].plot(x='ds', y='y', ax=ax, color='red', marker='o')  
    tmp[(tmp['split']==i) & (tmp['train']==False)].plot(x='ds', y='yhat', ax=ax, color='orange', marker='^')
```



```
chapter1/bad-git/pipeline.py
... @@ -1,9 +1,14 @@
1 - # EXAMPLE BELOW TAKEN FROM THE SPARK API DOCS, BEFORE BEING UPDATE FOR THE BOOK
1 + # EXAMPLE BELOW ADAPTED FROM THE SPARK DOCS
2 2 # https://spark.apache.org/docs/latest/ml-pipeline.html#pipeline
3 3 from pyspark.ml import Pipeline
4 4 from pyspark.ml.classification import LogisticRegression
5 5 from pyspark.ml.feature import HashingTF, Tokenizer
6 6
7 + import json
8 +
9 + with open("model_config.json") as f:
10 +     model_config = json.load(f)
11 +
12 # Prepare training documents from a list of (id, text, label) tuples.
13 training = spark.createDataFrame([
14     (0, "a b c d e spark", 1.0),
@@ -15,7 +20,7 @@
15 20 # Configure an ML pipeline, which consists of three stages: tokenizer, hashingTF, and lr.
16 21 tokenizer = Tokenizer(inputCol="text", outputCol="words")
17 22 hashingTF = HashingTF(inputCol=tokenizer.getOutputCol(), outputCol="features")
18 - lr = LogisticRegression(maxIter=10, regParam=0.001)
23 + lr = LogisticRegression(maxIter=model_config['maxIter'], regParam=model_config['regParam'])
19 24 pipeline = Pipeline(stages=[tokenizer, hashingTF, lr])
```

mlflow

ExperimentsModels

GitHubDocs

Experiments

Search Experiments

Default

Default

Experiment ID: 0

Artifact Location:
file:///home/andrew/dev/github/Machine-Learning-Engineering-with-Python/chapter1/mlflow/mlruns/0

Notes

None

Search Runs: metrics.rmse < 1 and params.model = "tree" and tags.mlflow.source.type = "LOCAL"

Filter

Search

Clear

Showing 3 matching runs

Compare

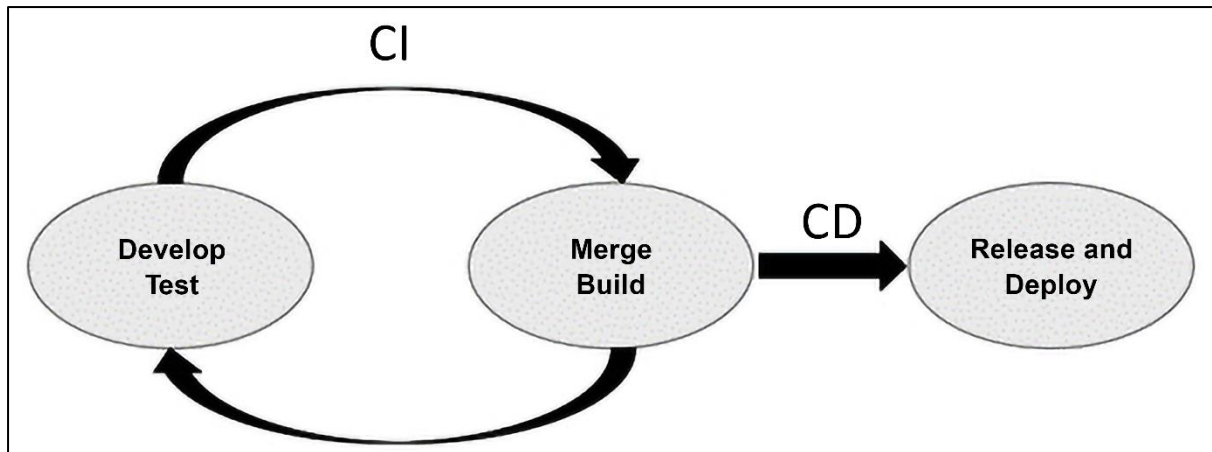
Delete

Download CSV

Columns

	Start Time	Run Name	User	Source	Version	Models	Parameters	Metrics
☐	2021-02-22 13:12:57	-	andrew	mlflow_fore	c6790e	-	param1	foo
☐	2021-02-22 10:23:15	-	andrew	mlflow_fore	c6790e	-	-	rmse
☐	2021-02-17 19:23:53	-	andrew	mlflow_basi	ed8a02	-	20	2.762

Load more



Python package Filter workflow runs

1 workflow run

Event	Status	Branch	Actor
Basic CI run with dummy test	Success	feature/actions	AndyMc629

Python package #12: Commit e293632 pushed by AndyMc629

9 minutes ago 17s

Triggered via push 10 minutes ago

AndyMc629 pushed -o- e29a632 [feature/actions](#)

Status: Success

Total duration: 17s

Artifacts: -

github-actions-basic.yml

on: push

Matrix: build

2 jobs completed

Show all jobs

← Python package

✓ Basic CI run with dummy test #12

🏠 Summary

Jobs

- ✓ build (3.9)
- ✓ build (3.10)

Run details

- 🕒 Usage
- 📄 Workflow file

build (3.10)

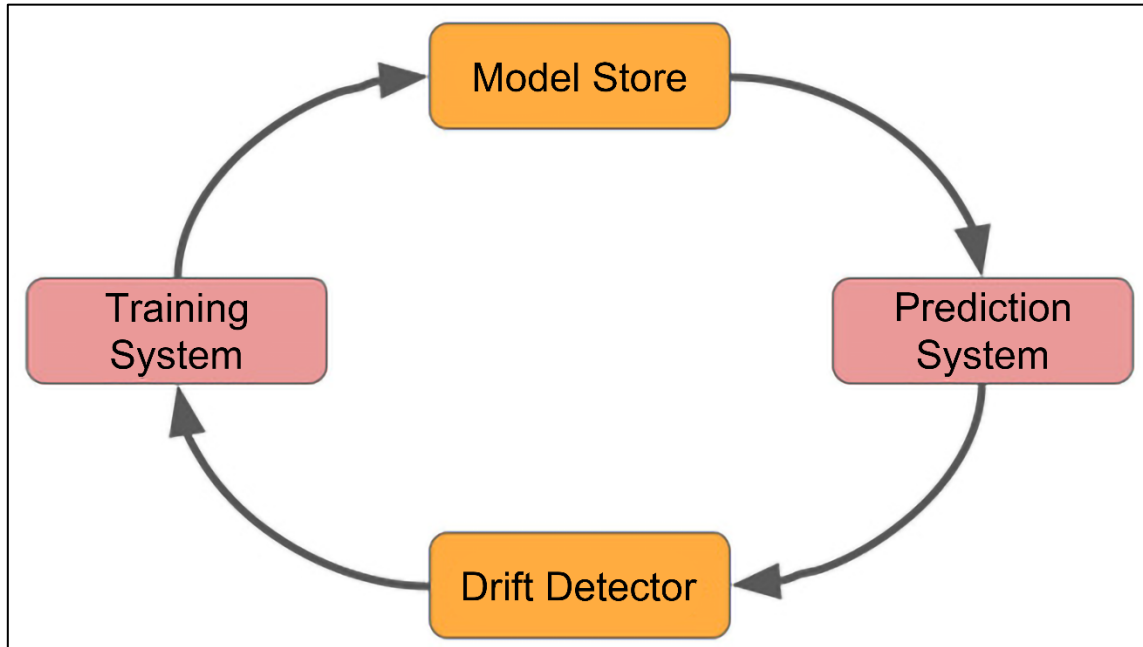
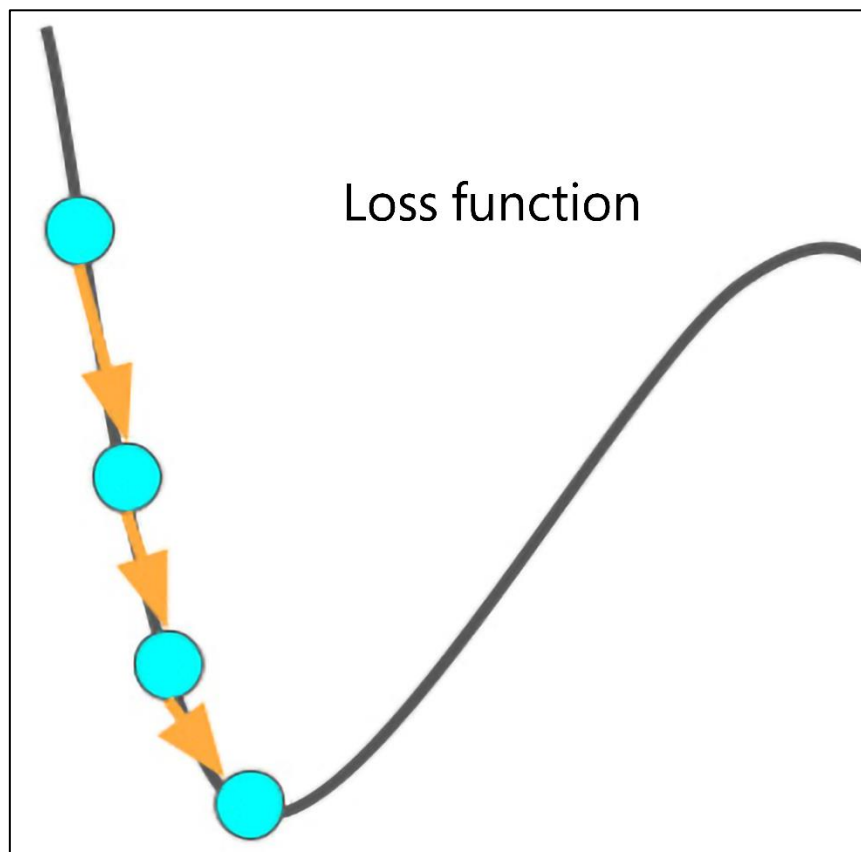
succeeded 11 minutes ago in 8s

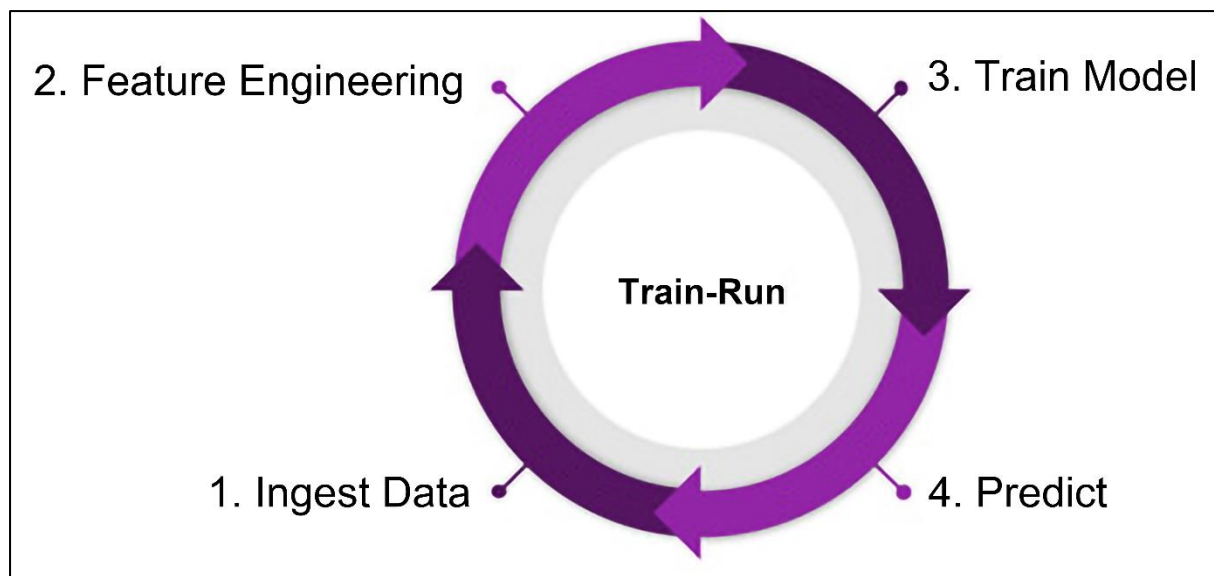
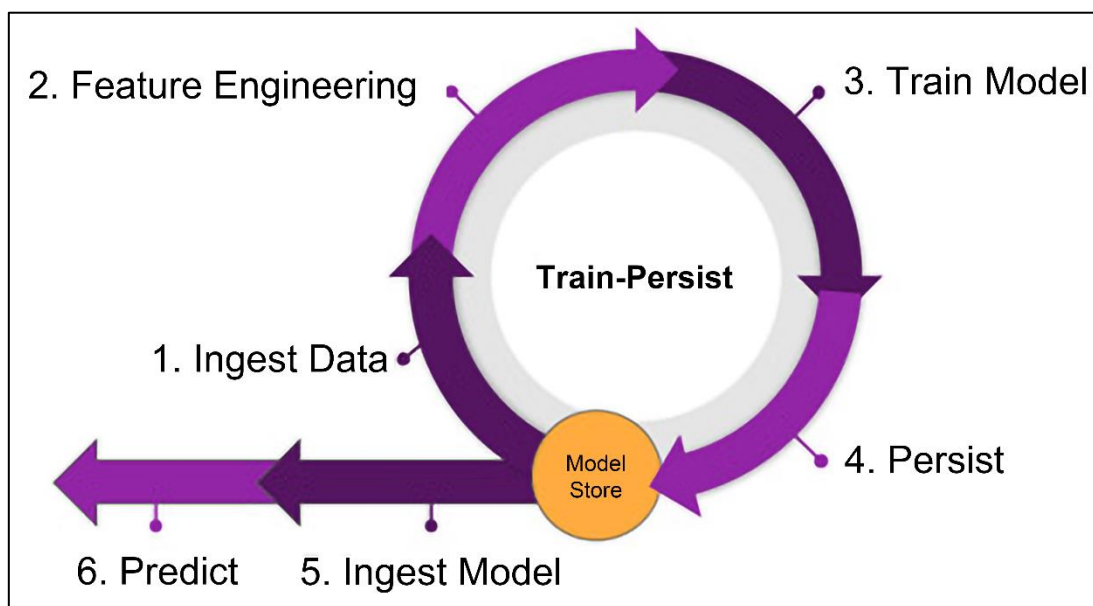
- > ✓ Set up job
- > ✓ Run actions/checkout@v3
- > ✓ Set up Python 3.10
- > ✓ Install dependencies
- > ✓ Lint with flake8
- > ✓ Test with pytest
- > ✓ Post Set up Python 3.10
- > ✓ Post Run actions/checkout@v3
- > ✓ Complete job

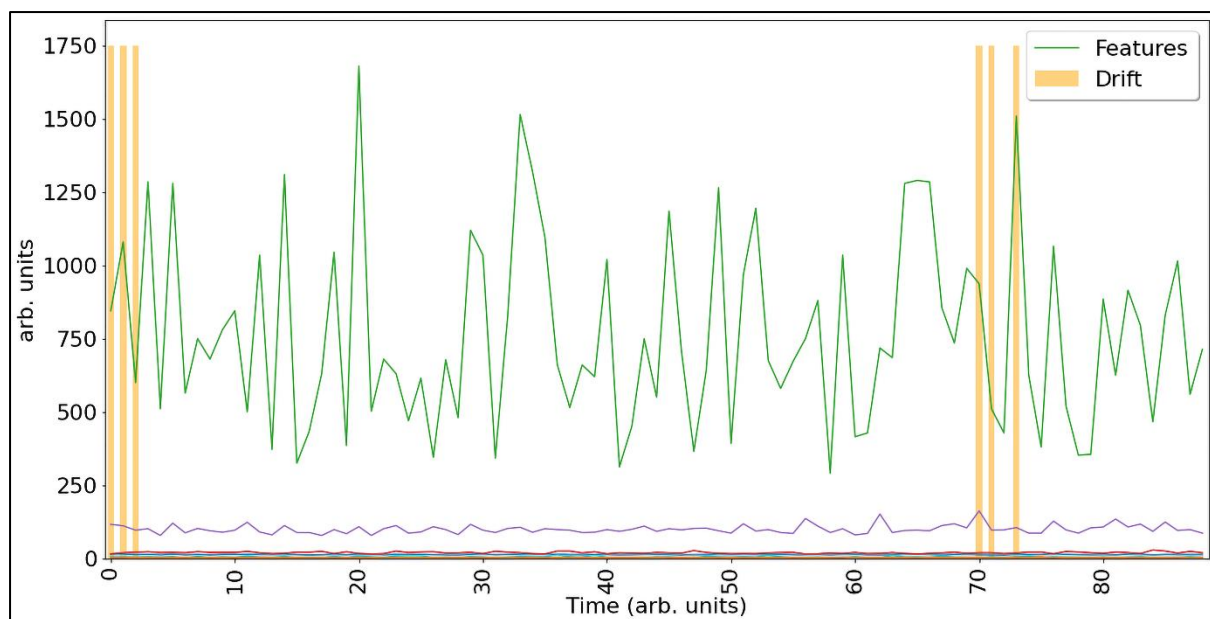
✓ Test with pytest

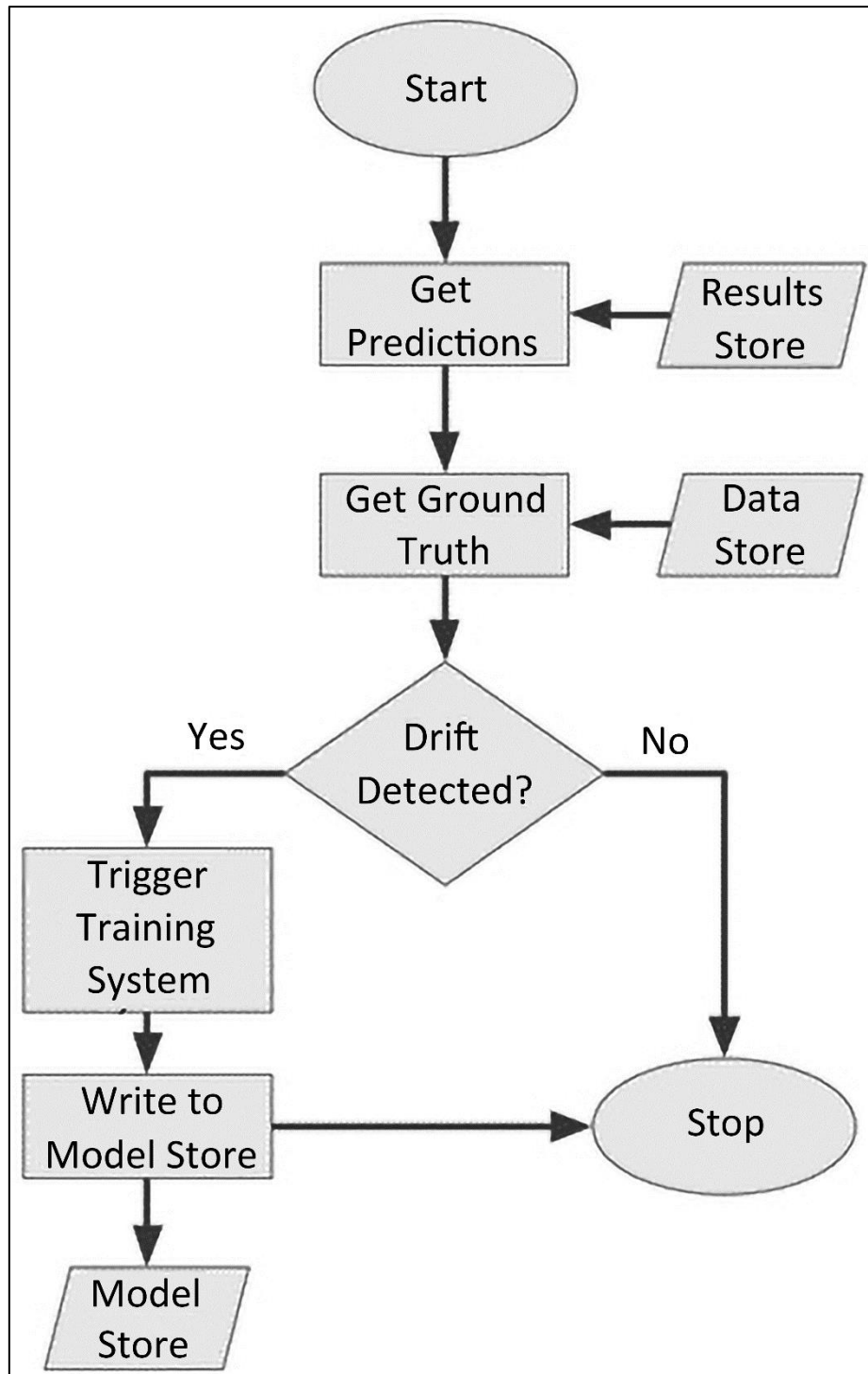
```
1 ▶ Run pytest
11 ===== test session starts =====
12 platform linux -- Python 3.9.16, pytest-7.3.1, pluggy-1.0.0
13 rootdir: /home/runner/work/Machine-Learning-Engineering-with-Python-Second-Edition/Machine-Learning-Engineering-with-Python-Second-Edition/Chapter02
14 plugins: anyio-3.6.2
15 collected 3 items
16
17 test_basic.py . [ 33%]
18 test_model_performance.py .. [100%]
19
20 ===== 3 passed in 1.33s =====
```

Chapter 3: From Model to Model Factory









```

{
  "version": "0.3.1",
  "timestamp": "2023-05-01 14:55:28.221095",
  "metrics": [
    {
      "metric": "DatasetDriftMetric",
      "result": {
        "drift_share": 0.5,
        "number_of_columns": 13,
        "number_of_drifted_columns": 0,
        "share_of_drifted_columns": 0.0,
        "dataset_drift": false
      }
    },
    {
      "metric": "DataDriftTable",
      "result": {
        "number_of_columns": 13,
        "number_of_drifted_columns": 0,
        "share_of_drifted_columns": 0.0,
        "dataset_drift": false,
        "drift_by_columns": {
          "alcalinity_of_ash": {
            "column_name": "alcalinity_of_ash",
            "column_type": "num",
            "stattest_name": "K-S p_value",
            "stattest_threshold": 0.05,
            "drift_score": 0.39530698914758006,
            "drift_detected": false,
            "current": {
              "small_distribution": {
                "x": [
                  14.6,
                  15.99,
                  17.38,
                  18.77,
                  20.16,
                  21.55,
                  22.939999999999998,
                  24.33,
                  25.72,
                  27.11,
                  28.5
                ],
                "y": [
                  0.06466736722981163,
                  0.16975183897825574,

```

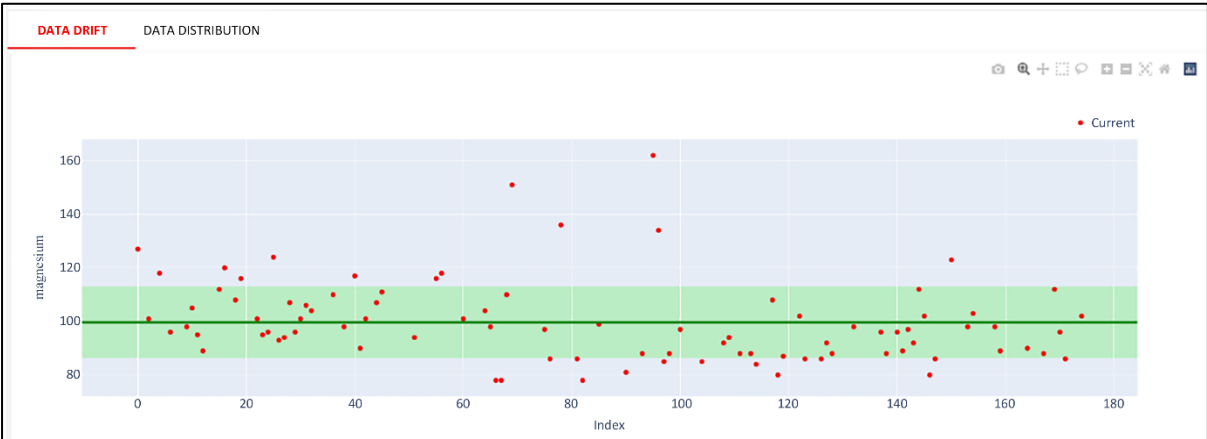
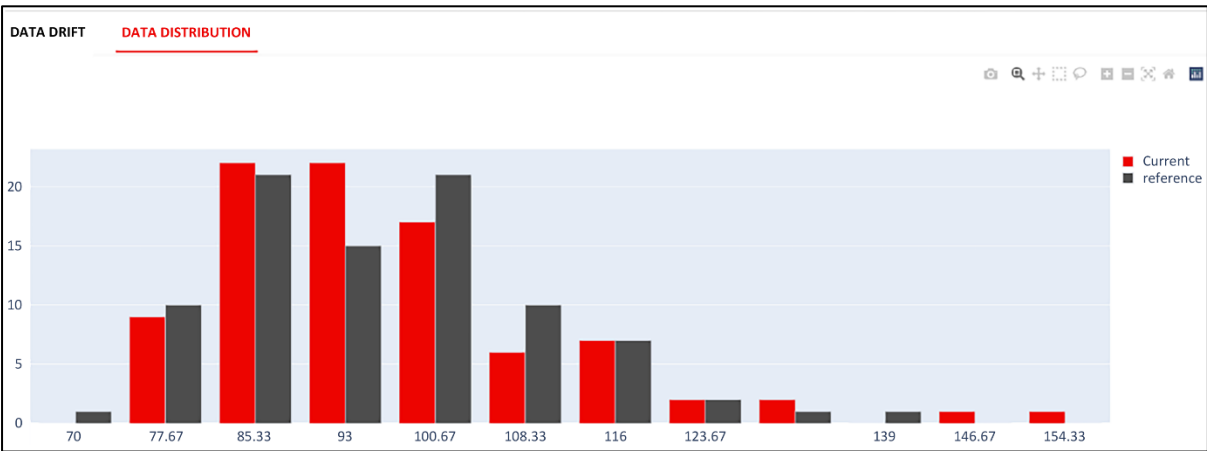


Data Drift Summary

Drift is detected for 0.0% of columns (0 out of 13).

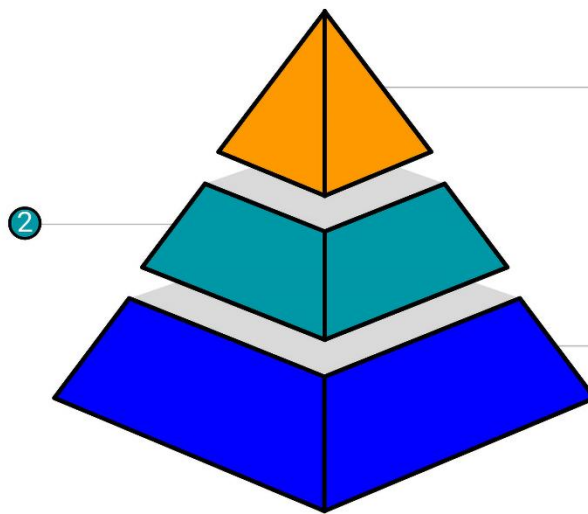
Search

Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test	Drift Score
> magnesium	num			Not Detected	K-S p_value	0.756091
> ash	num			Not Detected	K-S p_value	0.630804
> hse	num			Not Detected	K-S p_value	0.630804
> proline	num			Not Detected	K-S p_value	0.630804
> malic_acid	num			Not Detected	K-S p_value	0.507237
> total_phenols	num			Not Detected	K-S p_value	0.507237



Hyperparameter Optimization

Tune the parameters that tell your ML algorithm how to learn. Use tools that allow for efficient search over hyperparameter spaces.

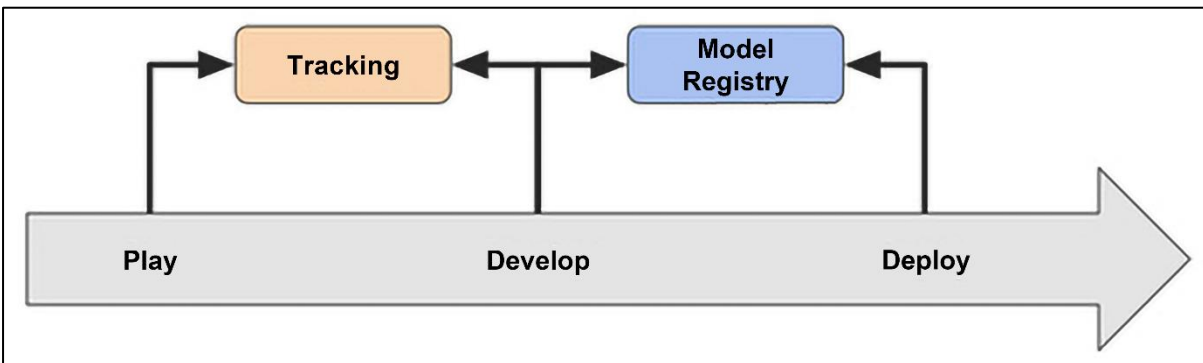


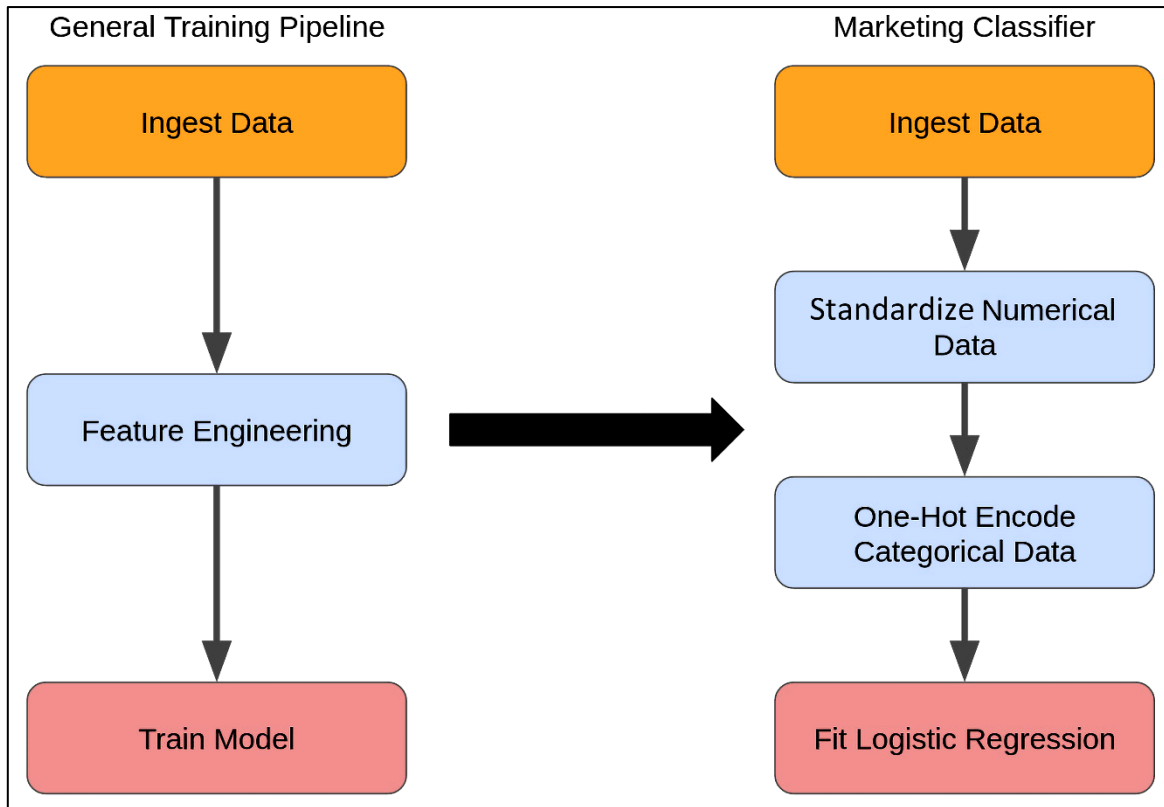
AutoML

Perform automated selection of the best hyperparameters and the best ML algorithm. Provide only a high level steer of this process.

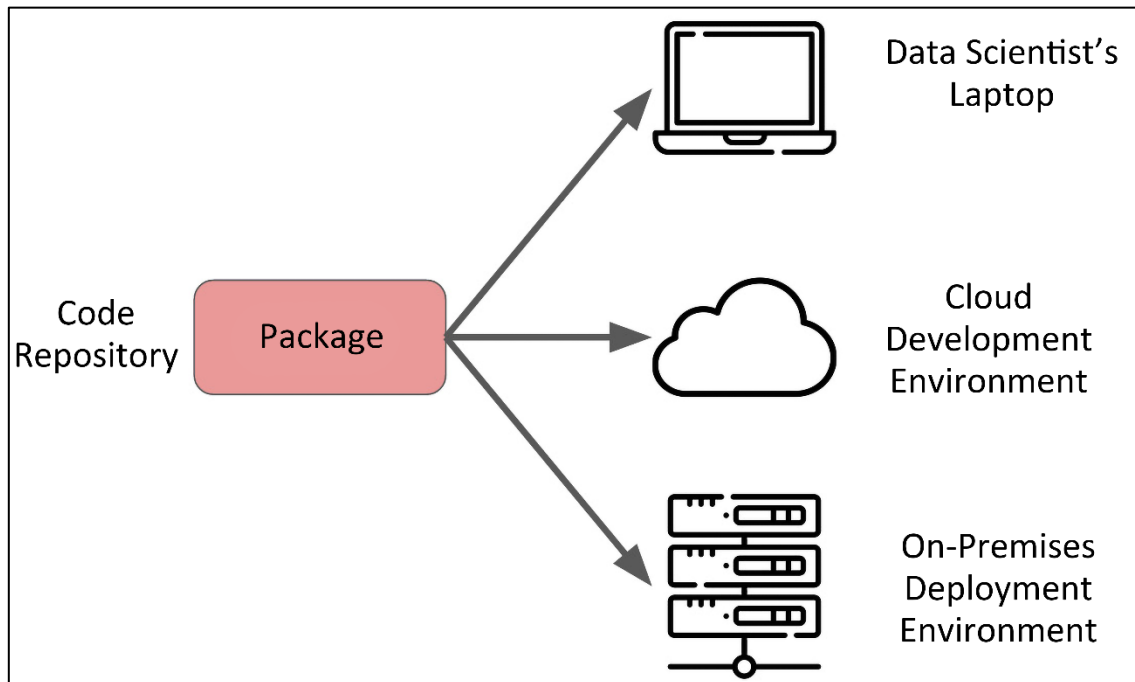
Hand Cranking

Define the hyperparameters for learning explicitly and select one ML algorithm to try at a time.





Chapter 4: Packaging Up



To get started:

```
>>> make create_environment
>>> conda activate mlewp-ed2-ch4-outliers
>>> make update_environment
```

```
[tool.poetry]
name = "my-ml-package"
version = "0.1.0"
description = ""
authors = ["AndyMc629 <andrewpmcmahon629@gmail.com>"]
readme = "README.md"
packages = [{include = "my_ml_package"}]

[tool.poetry.dependencies]
python = "^3.10"

[build-system]
requires = ["poetry-core"]
build-backend = "poetry.core.masonry.api"
```

Using version ^7.2.1 for pytest

Updating dependencies

Resolving dependencies... Downloading <https://files.pythonhosted.org/packages/ed/35/a31aed2993e398f6b09a790a18>
Resolving dependencies... (1.0s)

Writing lock file

Package operations: 4 installs, 0 updates, 0 removals

- Installing exceptiongroup (1.1.0)
- Installing iniconfig (2.0.0)
- Installing pluggy (1.0.0)
- Installing pytest (7.2.1)

```
[tool.poetry]
name = "my-ml-package"
version = "0.1.0"
description = ""
authors = ["AndyMc629 <andrewpmcmahon629@gmail.com>"]
readme = "README.md"
packages = [{include = "my_ml_package"}]

[tool.poetry.dependencies]
python = "^3.10"
pytest = "^7.2.1"

[build-system]
requires = ["poetry-core"]
build-backend = "poetry.core.masonry.api"
```

```
[tool.poetry]
name = "my-ml-package"
version = "0.1.0"
description = ""
authors = ["AndyMc629 <andrewpmcmahon629@gmail.com>"]
readme = "README.md"
packages = [{include = "my_ml_package"}]

[tool.poetry.dependencies]
python = "^3.10"

[tool.poetry.group.dev.dependencies]
pytest = "^7.2.1"
pytest-mock = "*"

[build-system]
requires = ["poetry-core"]
build-backend = "poetry.core.masonry.api"
```

Virtual environment already activated: `/home/andrew/anaconda3/envs/mlwp-ed2-ch4`

Installing dependencies from lock file

Package operations: 1 install, 0 updates, 0 removals

- Installing `pytest-mock` (3.10.0)

Installing the current project: `my-ml-package` (0.1.0)

Building my-ml-package (0.1.0)

- Building sdist
- Built my_ml_package-0.1.0.tar.gz
- Building wheel
- Built my_ml_package-0.1.0-py3-none-any.whl

```
===== test session starts =====
platform linux -- Python 3.8.5, pytest-6.1.1, py-1.9.0, pluggy-0.13.1 -- /home/andrew/anaconda3/envs/mleng/bin/python
cachedir: .pytest_cache
rootdir: /home/andrew/dev/github/Machine-Learning-Engineering-with-Python/chapter4/outlier_package
collected 2 items

outliers/tests/test_create_data.py::test_data_is_numpy PASSED
outliers/tests/test_create_data.py::test_data_is_large PASSED
===== 2 passed in 0.45s =====
```

```
===== test session starts =====
platform linux -- Python 3.8.5, pytest-6.1.1, py-1.9.0, pluggy-0.13.1 -- /home/andrew/anaconda3/envs/mleng/bin/python
cachedir: .pytest_cache
rootdir: /home/andrew/dev/github/Machine-Learning-Engineering-with-Python/chapter4/outlier_package
collected 5 items

outliers/tests/test_create_data.py::test_data_is_numpy PASSED
outliers/tests/test_create_data.py::test_data_is_large PASSED
outliers/tests/test_detectors.py::test_model_creation PASSED
outliers/tests/test_detectors.py::test_model_get_models PASSED
outliers/tests/test_detectors.py::test_model_evaluation PASSED
```

```
[main] INFO    profile include tests: None
[main] INFO    profile exclude tests: None
[main] INFO    cli include tests: None
[main] INFO    cli exclude tests: None
[main] INFO    running on Python 3.10.9
Run started:2023-01-20 21:00:55.255019

Test results:
>> Issue: [B101:assert_used] Use of assert detected. The enclosed code will be removed when compiling to optim
ised byte code.
    Severity: Low    Confidence: High
    CWE: CWE-703 (https://cwe.mitre.org/data/definitions/703.html)
    Location: outliers/tests/test_create_data.py:12:4
    More Info: https://bandit.readthedocs.io/en/1.7.4/plugins/b101\_assert\_used.html
11     def test_data_is_numpy(dummy_data):
12         assert isinstance(dummy_data, numpy.ndarray)
13
```

Code scanned:

Total lines of code: 98

Total lines skipped (#nosec): 0

Run metrics:

Total issues (by severity):

Undefined: 0

Low: 7

Medium: 0

High: 0

Total issues (by confidence):

Undefined: 0

Low: 0

Medium: 0

High: 7

Files skipped (0):

```

=====+
                                     /$$$$$ /$
                                     /$_ $$ /$
    /$$$$$ /$$$$$ | $$ \_ /$$$$$ /$$$$$ /$ /$
    /$_ / |  $$ | $$$ /$_ $$_ | $$_ | $$
    | $$$$ /$$$$$ | $$_ | $$$$$$ | $$_ | $$ | $$
    \_ $$ /$_ $$_ | $$ | $$_ / | $$ /$ | $$ | $$
    /$$$$$/ | $$$$ | $$_ | $$$$$$ | $$$/ | $$$$$$
    |_/ \_ / |_/ \_ / \_ / \_ / \_ / $$_
                                     /$ | $$_
                                     | $$$$/
                                     \_ /

by pyup.io

=====+

REPORT

You are using Safety's free vulnerability database. This data is
outdated, limited, and
licensed for non-commercial use only.
All commercial projects must sign up and get an API key at
https://pyup.io

Safety v2.3.5 is scanning for Vulnerabilities...
Scanning dependencies in your environment:

-> /usr/local/spark/python
-> /home/andrew/anaconda3/envs/mlewp-ed2-ch4-outliers/lib/python3.10/site-
packages

Using non-commercial database
Found and scanned 17 packages
Timestamp 2023-01-20 20:41:05
1 vulnerability found
0 vulnerabilities ignored

=====+
VULNERABILITIES FOUND
=====+

-> Vulnerability found in wheel version 0.37.1
Vulnerability ID: 51499
Affected spec: <0.38.1
ADVISORY: Wheel 0.38.1 includes a fix for CVE-2022-40898: An issue
discovered in Python Packaging Authority (PyPA) Wheel 0.37.1 and earlier...
CVE-2022-40898
For more information, please visit https://pyup.io/v/51499/f17

Scan was completed. 1 vulnerability was found.

=====+
REMEDIATIONS

1 vulnerability was found in 1 package. For detailed remediation & fix
recommendations, upgrade to a commercial license.

=====+

You are using Safety's free vulnerability database. This data is
outdated, limited, and
licensed for non-commercial use only.
All commercial projects must sign up and get an API key at
https://pyup.io

```

- sqlite=3.40.1=h5082296_0
- tk=8.6.12=h1ccaba5_0
- tzdata=2022g=h04d1e81_0
- wheel=0.37.1=pyhd3eb1b0_0
- xz=5.2.10=h5eee18b_1
- zlib=1.2.13=h5eee18b_0
- pip

pip:

- certifi==2022.12.7
- pip==22.3.1
- setuptools==65.6.3
- safety
- wheel>=0.38.1

```
+=====+
                                     /$$$$$ /$
                                   /$ _ $ /$
    /$$$$$ /$$$$$ | $ _ $ /$$$$$ /$$$$$ /$ /$
    /$ _ $ /$ _ $ | $$$ /$ _ $ | $ _ $ | $ | $
    | $$$$ /$$$$$ | $ _ $ | $$$$ | $ | $ | $
    \ $ _ $ /$ _ $ | $ | $ _ $ | $ /$ | $ | $
    /$$$$$ | $$$$ | $ | $$$$ | $$$ | $$$$
    \ $ _ $ /$ _ $ | $ | $ _ $ | $ _ $ /$
                                     /$ | $
                                     | $$$$/
                                     \ $ _ $ /

by pyup.io
+=====+

REPORT

You are using Safety's free vulnerability database. This data is outdated, limited, and
licensed for non-commercial use only.
All commercial projects must sign up and get an API key at https://pyup.io

Safety v2.3.5 is scanning for Vulnerabilities...
Scanning dependencies in your environment:

-> /home/andrew/anaconda3/envs/mlewp-ed2-ch4-outliers/lib/python3.10/site-packages
-> /usr/local/spark/python

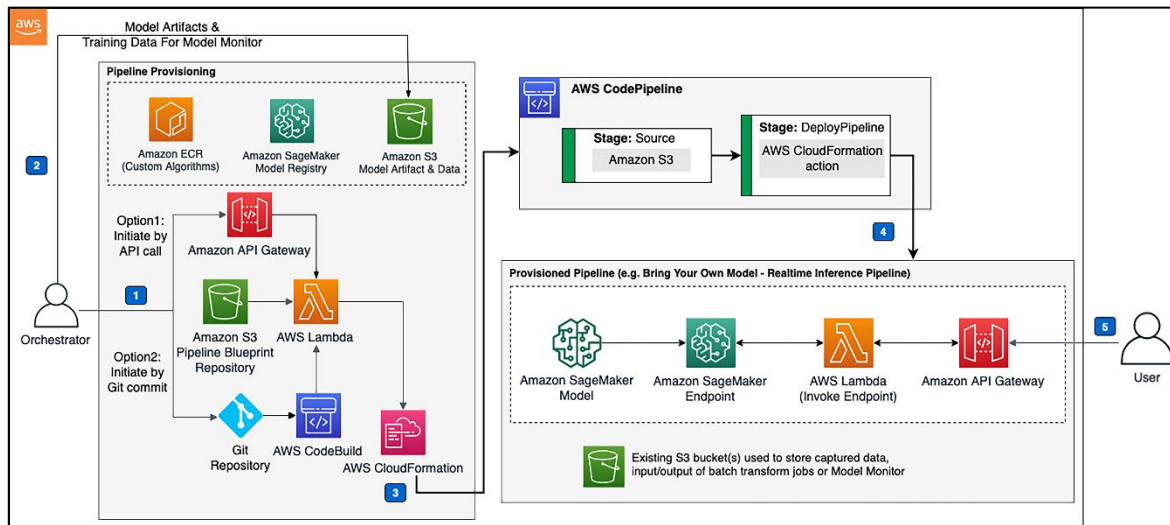
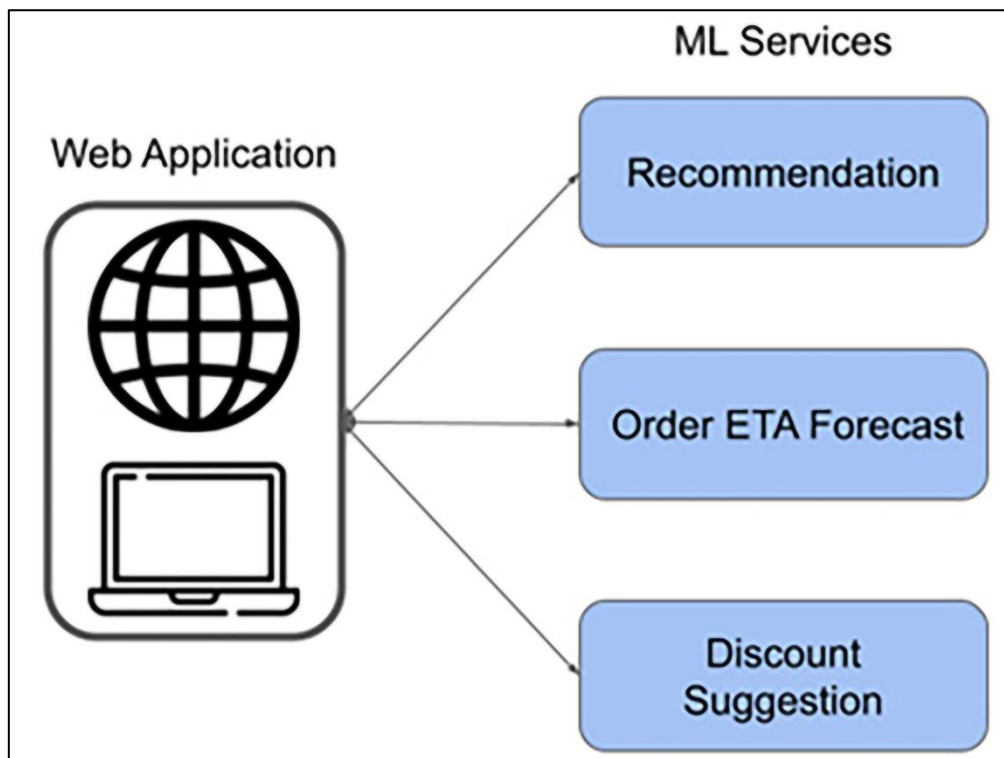
Using non-commercial database
Found and scanned 17 packages
Timestamp 2023-01-20 20:55:44
0 vulnerabilities found
0 vulnerabilities ignored
+=====+

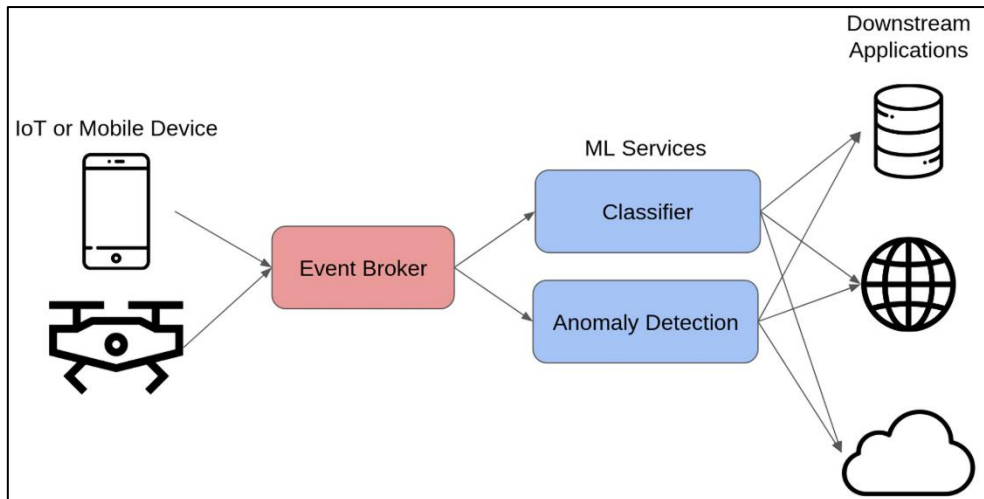
No known security vulnerabilities found.
+=====+

You are using Safety's free vulnerability database. This data is outdated, limited, and
licensed for non-commercial use only.
All commercial projects must sign up and get an API key at https://pyup.io
```

```
2023-01-22 21:31:57,317 | root | ERROR | Unexpected error
Traceback (most recent call last):
  File "/home/andrew/dev/github/Machine-Learning-Engineering-with-Python-Second-Edition",
    df_transformed = feature_transform(list_of_nums)
  File "/home/andrew/dev/github/Machine-Learning-Engineering-with-Python-Second-Edition",
    return df.mean()
AttributeError: 'list' object has no attribute 'mean'
```

Chapter 5: Deployment Patterns and Tools





mlwp2 / New Request

POST http://127.0.0.1:8080/forecast

Params Authorization Headers (8) **Body** Pre-request Script Tests Settings

none form-data x-www-form-urlencoded **raw** binary GraphQL JSON

```
1 {
2   ... "store_number": 10,
3   ... "forecast_start_date": "2023-07-01T00:00:00"
4 }
```

Body Cookies Headers (5) Test Results

Pretty Raw Preview Visualize JSON

```
1 {
2   "result": [
3     0.6043513740449178,
4     0.039012484283529325,
5     0.4893184826596272,
6     0.31851042354330683,
7     0.09421735231755868,
8     0.7291520357848239,
9     0.22508165205615693,
10    0.07317010122711154,
11    0.7551697651805447,
12    0.11153561883299301
13  ],
14   "store_number": 10
15 }
```

IMAGE ID	CREATED AT	REPOSITORY
b4cf53f44287	2023-07-02 22:17:37 +0100 BST	basic-ml-webservice
dad5d3d3ced3	2023-06-29 13:13:54 +0100 BST	inferencefunction
25b59787560e	2023-06-29 12:10:49 +0100 BST	508972911348.dkr.ecr.eu-west-2.amazonaws.com/mlewpstamm
25b59787560e	2023-06-29 12:10:49 +0100 BST	inferencefunction
16cd75dd36ad	2023-06-29 10:20:07 +0100 BST	<none>
ab560c318c87	2023-06-28 14:26:40 +0100 BST	public.ecr.aws/lambda/python
5e71c7186e71	2023-04-10 15:00:27 +0100 BST	electricweegie/custom-kserve-endpoint
5e71c7186e71	2023-04-10 15:00:27 +0100 BST	custom-kserve-endpoint
1465debfde6f	2023-04-03 22:47:45 +0100 BST	gcr.io/k8s-minikube/kicbase
b4faebd75645	2023-03-22 22:17:40 +0000 GMT	vsc-anaconda-1c27a4749abc4e72305ca51f3243da64
b2e19fd3d045	2023-03-22 22:13:08 +0000 GMT	vsc-volume-bootstrap
f60702d4ce22	2023-02-08 22:31:07 +0000 GMT	electricweegie/docker101tutorial
f60702d4ce22	2023-02-08 22:31:07 +0000 GMT	docker101tutorial
9793ee61fc75	2022-11-20 05:16:44 +0000 GMT	alpine/git
2b4661558fb8	2022-11-12 03:39:38 +0000 GMT	alpine
476b7007f4f5	2022-10-25 20:41:01 +0100 BST	kindest/node
df953f7b867a	2020-02-07 01:06:27 +0000 GMT	kindest/node

Amazon ECR
>
Repositories
>
basic-ml-microservice

basic-ml-microservice

View

Images (1)

Find images

Image tag

▼

Artifact type

Pushed at

▼

Size (MB)

▼

Image URI

Digest

Scan status

latest

Image

02 July 2023, 23:23:29 (UTC+01)

70.97

Copy URI

sha256:f1e5a76b1e1e7df...

Complete

Create cluster [Info](#)

An Amazon ECS cluster groups together tasks, and services, and allows for shared capacity and common configurations. All of your tasks, services, and capacity must belong to a cluster.

Cluster configuration

Cluster name

Cluster name here. For example, DevCluster.

There can be a maximum of 255 characters. The valid characters are letters (uppercase and lowercase), numbers, hyphens, and underscores.

► Networking [Info](#)

By default tasks and services run in the default subnets for your default VPC. To use the non-default VPC, specify the VPC and subnets.

► Infrastructure [Info](#)

Your cluster is automatically configured for AWS Fargate (serverless) with two capacity providers. Add Amazon EC2 instances, or external instances using ECS Anywhere.

Serverless

► Monitoring - optional [Info](#)

Container Insights is off by default. When you use Container Insights, there is a cost associated with it.

► Tags - optional [Info](#)

Tags help you to identify and organize your clusters.

▼ Networking [Info](#)

By default tasks and services run in the default subnets for your default VPC. To use the non-default VPC, specify the VPC and subnets.

VPC

Use a VPC with public and private subnets. By default, VPCs are created for your AWS account. To create a new VPC, go to the [VPC Console](#).

vpc-422d8f3b
default

Subnets

Select the subnets where your tasks run. We recommend that you use three subnets for production.

Choose subnets

subnet-23806e68

eu-west-1b 172.31.16.0/20

subnet-80223de6

eu-west-1a 172.31.0.0/20

subnet-f70159ad

eu-west-1c 172.31.32.0/20

Default namespace - optional

Select the namespace to specify a group of services that make up your application. You can overwrite this value at the service level.

Specify a namespace

▼ Infrastructure Info

Serverless

Your cluster is automatically configured for AWS Fargate (serverless) with two capacity providers. Add Amazon EC2 instances, or external instances using ECS Anywhere.

☒ AWS Fargate (serverless)

Pay as you go. Use if you have tiny, batch, or burst workloads or for zero maintenance overhead. The cluster has Fargate and Fargate Spot capacity providers by default.

☐ Amazon EC2 instances

Manual configurations. Use for large workloads with consistent resource demands.

☐ External instances using ECS Anywhere

Manual configurations. Use to add data center compute.

✔ Cluster mlew2-ecs-cluster has been created successfully.

Amazon Elastic Container Service > Clusters

Clusters (1) Info

↻ Create cluster

🔍 Search clusters

< 1 > ⚙

Cluster	Services	Tasks	Registered container instances	CloudWatch monitoring	Capacity provide
mlew2-ecs-cluster	0	No tasks running	0	✔ Default	No default found

Container - 1 Info

Essential container Remove

Container details

Specify a name, container image, and whether the container should be marked as essential. Each task definition must have at least one essential container.

Name

Image URI

Essential container

mlmicro

508972911348.dkr.ecr.eu-west-1.amazonaws.com/

Yes ▼

Private registry Info

Store credentials in Secrets Manager, and then use the credentials to reference images in private registries.

☐ Private registry authentication

Port mappings Info

Add port mappings to allow the container to access ports on the host to send or receive traffic. Any changes to port mappings configuration impacts the associated service connect settings.

Container port

Protocol

Port name

App protocol

5000

TCP ▼

mlmicro-5000-tcp

HTTP ▼

Remove

Add more port mappings

▼ Environment

Specify the infrastructure requirements for the task definition.

App environment [Info](#)

Specify the infrastructure for the task definition.

Add an option ▼

AWS Fargate (serverless) ✕

Operating system/Architecture [Info](#)

Linux/X86_64 ▼

Task size [Info](#)

Specify the amount of CPU and memory to reserve for your task.

CPU

1 vCPU ▼

Memory

3 GB ▼

▼ Container size - optional [Info](#)

For each container associated with the task, specify the container-level CPU and memory.

Add

▼ Task roles, network mode - conditional

Task role [Info](#)

A task IAM role allows containers in the task to make API requests to AWS services. You can create a task IAM role from the [IAM console](#).

Task execution role [Info](#)

A task execution IAM role is used by the container agent to make AWS API requests on your behalf. If you don't already have a task execution IAM role created, we can create one for you.

Network mode [Info](#)

The network mode that's used for your tasks. When the AWS Fargate (serverless) launch type is selected, you must use the awsvpc network mode. If you select the Amazon EC2 instance launch type, you can use different network modes in Linux or Windows. On Linux, you can choose between bridge, awsvpc, host, or none. On Windows, you can choose between default or awsvpc.

Task definition successfully created

basic-ml-microservice-tasks:1 has been successfully created. You can use this task definition to deploy a service or run a task.

Deploy

×

Amazon Elastic Container Service > Task definitions > basic-ml-microservice-tasks > Revision 1 > Containers

basic-ml-microservice-tasks:1

Deploy Actions Create new revision

Overview [Info](#)

ARN arn:aws:ecs:eu-west-1:5089729 11348:task-definition/basic-ml-microservice-tasks:1	Status ACTIVE	Time created 03/07/2023, 14:14:18 UTC	App environment FARGATE
Task role -	Task execution role ecsTaskExecutionRole	Operating system/Architecture Linux/X86_64	Network mode awsvpc

basic-ml-microservice-tasks:1

Deploy ▲

Actions ▼

Create new revision ▼

Create service

Run task

Overview [Info](#)

ARN

arn:aws:ecs:eu-west-1:5089729
11348:task-definition/basic-ml-
microservice-tasks:1

Status

✔ ACTIVE

Time created

03/07/2023, 14:14:18 UTC

App environment

FARGATE

Task role

-

Task execution role

ecsTaskExecutionRole [↗](#)

Operating system/Architecture

Linux/X86_64

Network mode

awsvpc

Existing cluster

Select an existing cluster. To create a new cluster, go to [Clusters](#).

mlewp2-ecs-cluster ▼

▼ Compute configuration (advanced)

Compute options [Info](#)

To ensure task distribution across your compute types, use appropriate compute options.



Capacity provider strategy

Specify a launch strategy to
distribute your tasks across one or
more capacity providers.



Launch type

Launch tasks directly without the use
of a capacity provider strategy.

Launch type [Info](#)

Select either managed capacity (Fargate), or custom capacity (EC2 or user-managed, External instances). External instances are registered to your cluster using the ECS Anywhere capability.

FARGATE ▼

Platform version [Info](#)

Specify the platform version on which to run your service.

LATEST ▼

Task definition

Select an existing task definition. To create a new task definition, go to [Task definitions](#) .

☐ Specify the revision manually

Manually input the revision instead of choosing from the 100 most recent revisions for the selected task definition family.

Family

basic-ml-microservice-tasks ▼

Revision

1 (LATEST) ▼

Service name

Assign a unique name for this service.

basic-ml-microservice-service

Service type [Info](#)

Specify the service type that the service scheduler will follow.

☒ Replica

Place and maintain a desired number of tasks across your cluster.

☐ Daemon

Place and maintain one copy of your task on each container instance.

Desired tasks

Specify the number of tasks to launch.

1

▼ Deployment options

Deployment type [Info](#)

Select a deployment controller type for the service.

Rolling update ▼

Min running tasks % [Info](#)

Specify the minimum percent of running tasks allowed during a service deployment.

100

values in %

Max running tasks % [Info](#)

Specify the maximum percent of running tasks allowed during a service deployment.

200

values in %

▼ Deployment failure detection [Info](#)

☒ Use the Amazon ECS deployment circuit breaker

If the service can't reach a steady state because a task failed to launch, the deployment fails.

☒ Rollback on failures

If the current deployment fails, the service is rolled back to the last completed deployment state.

☐ Use CloudWatch alarm(s)

If the CloudWatch alarm or alarms that you specify transition to the ALARM state, the deployment fails.

▼ Networking

VPC [Info](#)

Choose the Virtual Private Cloud to use.

vpc-422d8f3b
default

Subnets

Choose the subnets within the VPC that the task scheduler should consider for placement.

Choose subnets

subnet-23806e68 ✕
eu-west-1b 172.31.16.0/20

subnet-80223de6 ✕
eu-west-1a 172.31.0.0/20

subnet-f70159ad ✕
eu-west-1c 172.31.32.0/20

Security group [Info](#)

Choose an existing security group or create a new security group.

☒ Use an existing security group

☐ Create a new security group

Security group name

Choose an existing security group.

Choose security groups

sg-e52346b4 ✕
default

Public IP [Info](#)

Choose whether to auto-assign a public IP to the task's elastic network interface (ENI).

☒ Turned on

▼ Load balancing - *optional*

Load balancer type [Info](#)

Configure a load balancer to distribute incoming traffic across the tasks running in your service.

Application Load Balancer ▼

Application Load Balancer

Specify whether to create a new load balancer or choose an existing one.

- ☒ Create a new load balancer
- ☐ Use an existing load balancer

Load balancer name

Assign a unique name for the load balancer.

basic-ml-microservice-lb

Choose container to load balance

mlmicro 5000:5000 ▼

Listener [Info](#)

Specify the port and protocol that the load balancer will listen for connection requests on.

- ☒ Create new listener
- ☐ Use an existing listener
- You need to select an existing load balancer.

Port

80

Protocol

HTTP ▼

Target group [Info](#)

Specify whether to create a new target group or choose an existing one that the load balancer will use to route requests to the tasks in your service.

☒ Create new target group

☐ Use an existing target group

You need to select an existing load balancer.

Target group name

Protocol

HTTP ▼

Health check path [Info](#)

Health check protocol

HTTP ▼

Health check grace period [Info](#)

seconds

[EC2](#) > [Load balancers](#) > basic-ml-microservice-lb

basic-ml-microservice-lb



Actions ▼

▼ Details

Load balancer type
Application

Status
✓ Active

VPC
[vpc-422d8f3b](#)

IP address type
IPv4

Scheme
Internet-facing

Hosted zone
Z32O12XQLNTSW2

Availability Zones
[subnet-23806e68](#) eu-west-1b (euw1-az2)
[subnet-80223de6](#) eu-west-1a (euw1-az1)
[subnet-f70159ad](#) eu-west-1c (euw1-az3)

Date created
July 3, 2023, 22:51 (UTC+01:00)

Load balancer ARN
[arn:aws:elasticloadbalancing:eu-west-1:508972911348:loadbalancer/app/basic-ml-microservice-lb/ba85326f17beaf2c](#)

DNS name
[basic-ml-microservice-lb-821201318.eu-west-1.elb.amazonaws.com](#) (A Record)

HTTP

mlewp2 / ecs-forecast

POST

▼

http://basic-ml-microservice-lb-821201318.eu-west-1.elb.amazonaws.com:80/forecast

Params

Authorization

Headers (8)

Body ●

Pre-request Script

Tests

Settings

● none

● form-data

● x-www-form-urlencoded

● raw

● binary

● GraphQL

JSON ▼

1

2

3

4

5

```
{
  "store_number": 10,
  "forecast_start_date": "2023-07-01T00:00:00"
}
```

Body

Cookies

Headers (5)

Test Results

Pretty

Raw

Preview

Visualize

JSON ▼

≡

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

```
{
  "result": [
    0.9062587832716543,
    0.014864683407240564,
    0.10211568703831342,
    0.17991825928451755,
    0.37969301991243387,
    0.29514363022688117,
    0.3220340608518105,
    0.4222413133102756,
    0.34795750219099175,
    0.5329701549249485
  ],
  "store_number": 10
}
```

▼ How Amazon MWAA works



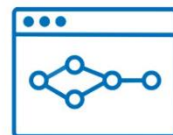
Create an environment

An environment contains your Airflow cluster, including your scheduler, workers, and web server.



Upload your DAGs to Amazon S3

Package and upload your DAG (Directed Acyclic Graph) code to Amazon S3. Amazon MWAA loads the code into Airflow.



Run your DAGs in Airflow

Run your DAGs from the Airflow UI or CLI. Monitor your environment with Amazon CloudWatch.

Name

mlwp2-airflow-dev-env

Use only letters, numbers, dashes, or underscores. Max 80 characters.

Airflow version

2.5.1 (Latest)

Weekly maintenance window start (UTC)

Thursday

02:30

Buckets (9) Info			 Copy content	Empty	Delete	Create bucket
<input type="text" value="Find buckets by name"/>		< 1 > 				
Name	AWS Region	Access	Creation date			
☐ mlwp2-ch5-airflow-example	EU (Ireland) eu-west-1	<u>Bucket and objects not public</u>	July 6, 2023, 07:09:03 (UTC+01:00)			

DAG code in Amazon S3 [Info](#)

Amazon MWAA uses your Amazon S3 bucket to load your DAGs and supporting files. Specify your S3 bucket, and the paths of your DAG folder, plugins.zip, and requirements.txt.



i Create or specify an S3 bucket to store your DAG code. The bucket must have versioning enabled. You can create a new bucket in the [Amazon S3 console](#)

S3 Bucket

The S3 bucket where your source code is stored. Enter an S3 URI or browse and select a bucket.

View

Browse S3

Format: s3://mybucketname

DAGs folder

The S3 bucket folder that contains your DAG code. Enter an S3 URI or browse and select a folder.

View

Browse S3

Format: s3://mybucketname/mydagfolder

Plugins file - *optional*

The S3 bucket ZIP file that contains your DAG plugins. Enter an S3 URI or browse and select a file object and version.

Choose a version

View

Browse S3

Format: s3://mybucketname/myplugins.zip

Requirements file - *optional*

The S3 bucket file that contains your DAG requirements.txt. Enter an S3 URI or browse and select a file object and version.

Choose a version

View

Browse S3

Format: s3://mybucketname/myrequirements.txt

Startup script file - *optional*

The shell script will be executed for customized image. Enter an S3 URI or browse and select a file object and version.

Choose a version

View

Browse S3

Format: s3://mybucketname/startup.sh

Parameters

Parameters are defined in your template and allow you to input custom values when you create or update a stack.

EnvironmentName

An environment name that is prefixed to resource names

PrivateSubnet1CIDR

Please enter the IP range (CIDR notation) for the private subnet in the first Availability Zone

PrivateSubnet2CIDR

Please enter the IP range (CIDR notation) for the private subnet in the second Availability Zone

PublicSubnet1CIDR

Please enter the IP range (CIDR notation) for the public subnet in the first Availability Zone

PublicSubnet2CIDR

Please enter the IP range (CIDR notation) for the public subnet in the second Availability Zone

VpcCIDR

Please enter the IP range (CIDR notation) for this VPC

Web server access

☒ **Private network (Recommended)**

Additional setup required. Your Airflow UI can only be accessed by secure login behind your VPC. Choose this option if your Airflow UI is only accessed within a corporate network. IAM must be used to handle user authentication.

☐ **Public network (No additional setup)**

Your Airflow UI can be accessed by secure login over the Internet. Choose this option if your Airflow UI is accessed outside of a corporate network. IAM must be used to handle user authentication.

 For private network access, the Airflow web server is reached via a VPC endpoint inside your VPC. Connecting to the endpoint requires additional setup. [Learn more about VPC endpoints](#) 

Security group(s)

A VPC security group is required to allow traffic between your environment and your web server.

☒ **Create new security group**

Allow MWAA to create a VPC security group with inbound and outbound rules based on your selection for web server access.

Existing security group(s)

You can choose 1 or more existing security groups to configure the inbound and outbound rules for your environment.

Choose security group



Max 5 security groups

Each Amazon MWAA environment includes the scheduler, web server, and 1 worker. Workers auto-scale up and down according to system load. You can monitor the load on your environment and modify its class at any time.

	DAG capacity*	Scheduler CPU	Worker CPU	Web server CPU
☒ mw1.small	Up to 50	1 vCPU	1 vCPU	0.5 vCPU
☐ mw1.medium	Up to 250	2 vCPU	2 vCPU	1 vCPU
☐ mw1.large	Up to 1000	4 vCPU	4 vCPU	2 vCPU

*under typical usage

Maximum worker count

The maximum number of workers your environment is permitted to scale up to.

Must be between 1 and 25

Minimum worker count

The minimum number of workers always present in your environment.

Must be less than or equal to maximum workers. Minimum 1 worker

Scheduler count

The number of schedulers your environment will use.

Must be between 2 and 5

Permissions [Info](#)

Execution role

The IAM role used by your environment to access your DAG code, write logs, and perform other actions.



Role name

Use alphanumeric and '+=,._-' characters. Maximum 64 characters.

i Amazon MWAA will create and assume the execution role in IAM named **AmazonMWAA-mlewp2-airflow-dev-env-D0aAPr** on your behalf. This role is configured with permission to retrieve code from your Amazon S3 bucket, use your KMS key, and send data to Amazon CloudWatch. You must add permissions to your execution role if your Airflow DAGs require access to any other AWS services.[Info](#)

Airflow environments

Environments (1)

Edit

Delete

Actions ▾

Create environment

Find environments

< 1 >

	Name ▾	Status ▾	Created date ▾	Airflow version ▾	Airflow UI ▾
	mllep-airflow-dev-env	Creating	Sep 20, 2021 20:32:11 (...)	2.0.2	Open Airflow ...


Airflow
DAGs
Datasets
Security
Browse
Admin
Docs
22:15 UTC


DAGs

All 1
Active 0
Paused 1

Auto-refresh


DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
☐ classification_pipeline airflow <div> <div></div> <div></div> <div></div> <div></div> </div> <div> <div>daily</div> <div>1</div> </div>				2023-07-05, 00:00:00	<div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div>	 	...	

Showing 1-1 of 1 DAGs

Grid

Graph

Calendar

Task Duration

Task Tries

Landing Times

Gantt

Details

<> Code

Audit Log

06/07/2023, 22:22:34

25

All Run Types

All Run States

Clear Filters

deferred

failed

queued

removed

restarting

running

scheduled

shutdown

skipped

success

up_for_reschedule

Auto-refresh

→

Duration

00:00:13

00:00:06

00:00:00

get_data

train_model

persist_model

DAG

classification_pipeline

DAG Details

DAG Runs Summary

Total Runs Displayed

2

Total success

2

First Run Start

2023-07-06, 22:19:44 UTC

Last Run Start

2023-07-06, 22:20:10 UTC

Max Run Duration

00:00:13

Mean Run Duration

00:00:11

AndyMc629 / mleip-airflow Private

Code Issues Pull requests **Actions** Projects Security

Upload DAGS to S3 Upload DAGS to S3 #9

Summary

Jobs

Upload DAGS to Amazon S3

Upload DAGS to Amazon S3
succeeded 1 minute ago in 11s

Set up job

Checkout

Configure AWS credentials from account

Copy files to bucket with the AWS CLI

Post Configure AWS credentials from account

Post Checkout

Complete job

```
Next, you will be prompted to generate a project from the template.
No git tags found in template; using HEAD as ref
🔊 The project template you would like to use
  (Use arrow keys)
» ZenML Starter
```

```
Next, you will be prompted to generate a project from the template.
No git tags found in template; using HEAD as ref
🔊 The project template you would like to use
  ZenML Starter
🔊 Short name for your project
  zenml-mlewp1
🔊 Version of your project
  0.1.0
🔊 The license under which your project will be released
  MIT license
🔊 The name of the person/entity holding the copyright
  Andy McMahon
```

🔧 The generated ZenML project will be populated with some example code featuring one of the scikit-learn classifier models.
Which model class would you like to see being used in the generated code?
Random Forest Classifier

Copying from template version 0.0.0.post28.dev0+88e5d96

```
identical .
create .copier-answers.yml
create run.py
create requirements.txt
conflict __init__.py
Overwrite __init__.py? [Y/n] Y
create pipelines
```

✓ Chapter05 / reaching_zen ●

✓ reaching_zen ●

> .ruff_cache

> .zen ●

> artifacts ●

> materializers ●

> pipelines ●

> steps ●

🐍 __init__.py U

! .copier-answers.yml U

🚢 .dockerignore U

🔑 LICENSE U

≡ requirements.txt U

🐍 run.py U



Log In

Enter your details below

Username

default

Password

Password



Log In

Pipelines

Runs

Stacks

Components

Hi there!

You can find your most important functions here on the dashboard.

1
Number of stacks

2
Number of components

1
Number of pipelines

2
Number of runs

Read our documentation

Read

Change the settings

Manage

Invite new members to your organization

Invite

Configuration

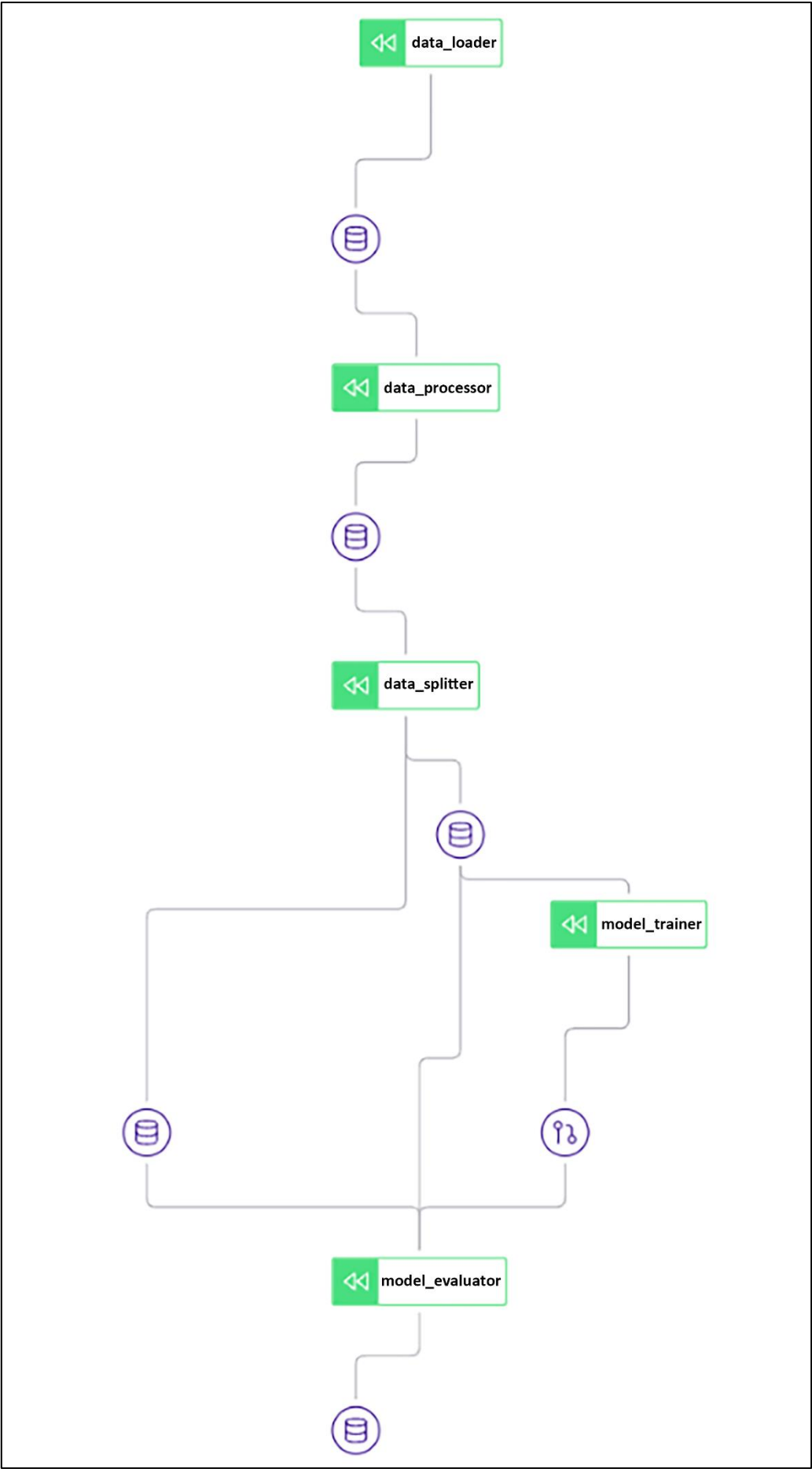
Runs

Q Search

Filter list

RUN ID	RUN NAME	PIPELINE NAME	STATUS	STACK NAME	AUTHOR	CREATED AT
28730a2a	model_training_pipelin_	model_training_pipeline		default	default	15.02.2023 21:12:12
ef3b60f3	model_training_pipelin_	model_training_pipeline		default	default	15.02.2023 21:09:04

1



Pipelines > 462e91e7-3436-492b-9c14-81e0d8b2a6cc

Pipeline CheatsheetDocumentation

ID	NAME	STATUS	AUTHOR	CREATED AT
462e91e7	model_training_pipeline	✓ ✓	default	15.02.2023 21:09:04

ConfigurationRuns

YAML ConfigurationDownload

```
---
version: "0.1"
steps:
  -
    source: "steps.data_loaders.data_loader"
    upstream_steps: []
    inputs: {}
  -
    source: "steps.data_loaders.data_processor"
    upstream_steps:
      - "data_loader"
    inputs:
```

Pipelines

ExperimentsRunsRecurring RunsArtifactsExecutions

Pipelines

+ Upload pipelineRefreshDelete

Filter pipelines

	Pipeline name	Description	Uploaded on ↓
☐	[Tutorial] DSL - Control struct...	source code Shows how to use conditional execution and exit handlers. This pipeline will ra...	17/02/2023, 21:52:42
☐	[Tutorial] Data passing in pyt...	source code Shows how to pass data between python components.	17/02/2023, 21:52:40
☐	[Demo] TFX - Taxi tip predicti...	source code GCP Permission requirements . Example pipeline that does classification with ...	17/02/2023, 21:52:39
☐	[Demo] XGBoost - Iterative ...	source code This sample demonstrates iterative training using a train-eval-check recursive l...	17/02/2023, 21:52:38

Rows per page: 10 < >

← ✓ my_pipeline 2023-02-18 20-53-45

RetryClone runTerminateArchive

GraphRun outputConfig

Simplify Graph

create-dataset

normalize-dataset

train-model

train-model

train-model

iris-training-pipeline-5jxp-4031717868

Input/OutputVisualizationsDetailsVolumes

This step corresponds to execution "train-model".

Input Parameters

n_neighbors9

Input Artifacts

normalized_iris_dataset

[minio://mipipeline/v2/artifact View All](#)
[ts/pipeline/iris-training-pipel](#)
[ine/951ab03c-1519-42b3-9](#)
[a2b-7bfa1f6b9c3b/normaliz](#)
[e-dataset/normalized_iris_d](#)
[ataset](#)
,0,1,2,3,4,Labels
0,-1.7204650534085253,-1.14835550;
1,-1.6972155256597614,-1.39054229;
2,-1.6739
...

Chapter 6: Scaling Up

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	no
33	services	married	secondary	no	4789	yes	yes	cellular	11	may	220	1	339	4	failure	no
35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	185	1	330	1	failure	no
30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	199	4	-1	0	unknown	no
59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	226	1	-1	0	unknown	no
35	management	single	tertiary	no	747	no	no	cellular	23	feb	141	2	176	3	failure	no
36	self-employed	married	tertiary	no	307	yes	no	cellular	14	may	341	1	330	2	other	no
39	technician	married	secondary	no	147	yes	no	cellular	6	may	151	2	-1	0	unknown	no
41	entrepreneur	married	tertiary	no	221	yes	no	unknown	14	may	57	2	-1	0	unknown	no
43	services	married	primary	no	-88	yes	yes	cellular	17	apr	313	1	147	2	failure	no

pdays	previous	poutcome	y	month_as_int
-1	0	unknown	no	10
339	4	failure	no	5
330	1	failure	no	4
-1	0	unknown	no	6
-1	0	unknown	no	5
176	3	failure	no	2
330	2	other	no	5
-1	0	unknown	no	5
-1	0	unknown	no	5
147	2	failure	no	4
-1	0	unknown	no	5
-1	0	unknown	no	4
-1	0	unknown	no	8
-1	0	unknown	yes	4
241	1	failure	no	1
-1	0	unknown	no	8
-1	0	unknown	no	8
152	2	failure	no	4
-1	0	unknown	no	5
152	1	other	no	7

_1	_2	_3	_4	_5	_6	_7	_8	_9	_10	_11	_12	_13	class
14.23	1.71	2.43	15.6	127.0	2.8	3.06	0.28	2.29	5.64	1.04	3.92	1065.0	0
13.2	1.78	2.14	11.2	100.0	2.65	2.76	0.26	1.28	4.38	1.05	3.4	1050.0	0
13.16	2.36	2.67	18.6	101.0	2.8	3.24	0.3	2.81	5.68	1.03	3.17	1185.0	0
14.37	1.95	2.5	16.8	113.0	3.85	3.49	0.24	2.18	7.8	0.86	3.45	1480.0	0
13.24	2.59	2.87	21.0	118.0	2.8	2.69	0.39	1.82	4.32	1.04	2.93	735.0	2
14.2	1.76	2.45	15.2	112.0	3.27	3.39	0.34	1.97	6.75	1.05	2.85	1450.0	0
14.39	1.87	2.45	14.6	96.0	2.5	2.52	0.3	1.98	5.25	1.02	3.58	1290.0	0

Name

mlewp2-cluster

Amazon EMR release [Info](#)

A release contains a set of applications which can be installed on your cluster.

emr-6.11.0 ▼

Application bundle

Spark



Core Hadoop

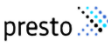


Flink

HBase



Presto



Trino



Custom



Applications included in bundle

Spark 3.3.2 on Hadoop 3.3.3 YARN with and Zeppelin 0.10.1

AWS Glue Data Catalogue settings

Use the AWS Glue Data Catalog to provide an external metastore for your application.

☐ Use for Spark table metadata

Operating system options [Info](#)

☒ Amazon Linux release

☐ Customised Amazon Machine Image (AMI)

☒ Automatically apply the latest Amazon Linux updates

Cluster configuration [Info](#)

Choose a configuration method for the primary, core and task node groups for your cluster.

☒ Instance groups

Choose one instance type per node group

☐ Instance fleets

Choose any combination of instance types within each node group

Instance groups

Primary

Choose EC2 instance type

m5.xlarge

4 vCore 16 GiB memory EBS only storage
On-demand price: USD 0.214 per instance/hour
Lowest spot price: \$0.087 (eu-west-1a) ▼

Actions ▼

☐ Use multiple primary nodes

To improve cluster availability, use three primary nodes with the same configuration and bootstrap actions. You cannot use multiple primary nodes with instance fleets.

► Node configuration - optional

Cluster scaling and provisioning option [Info](#)

The Amazon EMR console only supports EMR-managed scaling. To create a cluster with auto-scaling, use CLI or SDK.

Choose an option



Set cluster size manually

Use this option if you know your workload patterns in advance.



Use EMR-managed scaling

Monitor key workload metrics so that EMR can optimise the cluster size and optimise its resource utilisation.

Provisioning configuration

Set the size of your core and task instance groups. Amazon EMR attempts to provision this capacity when you launch your cluster.

Name	Instance type	Size		Use spot purchasing option
Core	m5.xlarge	<input type="text" value="1"/>	instance(s)	☐
Task - 1	m5.xlarge	<input type="text" value="1"/>	instance(s)	☐

Networking [Info](#)

Virtual private cloud (VPC) [Info](#)

vpc-422d8f3b

Browse

Create VPC [↗](#)

Subnet [Info](#)

subnet-23806e68

Browse

Create subnet [↗](#)

► EC2 security groups (firewall)

Cluster termination [Info](#)

- ☐ Manually terminate cluster
- ☒ Terminate cluster after idle time (recommended)

Idle time

Enter the time until your cluster terminates.

0 days ▼

01:00:00

Choose a time that is greater than 1 minute (00:01:00) and less than 7 days. The time is in hh:mm:ss (24-hour) format.

☐ Terminate cluster after last step has been completed
You cannot edit this selection after you have created your cluster.

☒ Use termination protection
Protect your EC2 instances from accidental termination.

Amazon EMR service role [Info](#)

The service role is an IAM role that Amazon EMR assumes to provision resources and perform service-level actions with other AWS services.

☐ Choose an existing service role
Select a default service role or a custom role with IAM policies attached so that your cluster can interact with other AWS services.

☒ Create a service role
Let Amazon EMR create a new service role so that you can grant and restrict access to resources in other AWS services.

Networking resources

We've already added the resources that you configured in the [Networking](#) section. Choose the VPC, subnet and security groups that the service role can access.

Virtual Private Cloud (VPC)

Choose one or more VPCs ▼

- ✕
vpc-422d8f3b

Subnet

Choose one or more subnets ▼

- ✕
subnet-23806e68

Security group

Choose one or more security groups ▼

ElasticMapReduce-Primary ✕
sg-051deff292498b3c4

ElasticMapReduce-Core ✕
sg-029348bc3ae9c3e02

EC2 instance profile for Amazon EMR

The instance profile assigns a role to every EC2 instance in a cluster. The instance profile must specify a role that can access the resources for your steps and bootstrap actions.

- ☐ Choose an existing instance profile
- Select a default role or a customised instance profile with IAM policies attached so that your cluster can interact with your resources in Amazon S3.

- ☒ Create an instance profile
- Let Amazon EMR create a new instance profile so that you can specify a customised set of resources for it to access in Amazon S3.

S3 bucket access [Info](#)

- ☐ Specific S3 buckets or prefixes in your account [Info](#)
- Choose the buckets or prefixes that you want this instance profile to access.
- ☒ All S3 buckets in this account with read and write access
- Grant the instance profile access to all buckets that have read and write access enabled in your account.

Security configuration and EC2 key pair – optional [Info](#)

Security configuration

Select your cluster encryption, authentication, authorisation and instance metadata service settings.

Amazon EC2 key pair for SSH to the cluster [Info](#)

 You haven't entered an EC2 key. If you're outside a VPN and want to enable SSH or use Hue SQL assistant with this cluster, you must enter an EC2 key.

 Your cluster 'mlewp2-cluster' has been successfully created.

Amazon EMR > EMR on EC2: Clusters > mlewp2-cluster

mlewp2-cluster

Updated less than a minute ago

▼ Summary

Cluster info	Applications	Cluster management	Status and time
Cluster ID j-2MFY5DTL16MYA	Amazon EMR version emr-6.11.0	Log destination in Amazon S3 Logging not configured	Status  Starting
Cluster configuration Instance groups	Installed applications Spark 3.3.2, Zeppelin 0.10.1	Primary node public DNS -	Creation time 1 July 2023 23:14 (UTC+01:00)
Capacity 1 Primary 1 Core 1 Task			Elapsed time 0 seconds

mlewp-ch6-emr-examples [Info](#)

[Objects](#) | [Properties](#) | [Permissions](#) | [Metrics](#) | [Management](#) | [Access Points](#)

Objects (3)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#)

[Upload](#)

Find objects by prefix

< 1 > [Settings](#)

☐	Name	Type	Last modified	Size	Storage class
☐	 bank.csv	csv	July 2, 2023, 07:06:07 (UTC+01:00)	450.7 KB	Standard
☐	 results.parquet/	Folder	-	-	-
☐	 spark_example_emr.py	py	July 2, 2023, 07:38:31 (UTC+01:00)	4.6 KB	Standard

[HTTP](#) mlewp2 / chapter6-serverless

[Save](#)

[POST](#) [https://8qg87m9380.execute-api.eu-west-2.amazonaws.com/Prod/classify_digit/](#)

[Send](#)

[Params](#) [Authorization](#) [Headers \(8\)](#) [Body](#) [Pre-request Script](#) [Tests](#) [Settings](#)

[Cookies](#)

[none](#) [form-data](#) [x-www-form-urlencoded](#) [raw](#) [binary](#) [GraphQL](#) [JSON](#)

[Beautify](#)

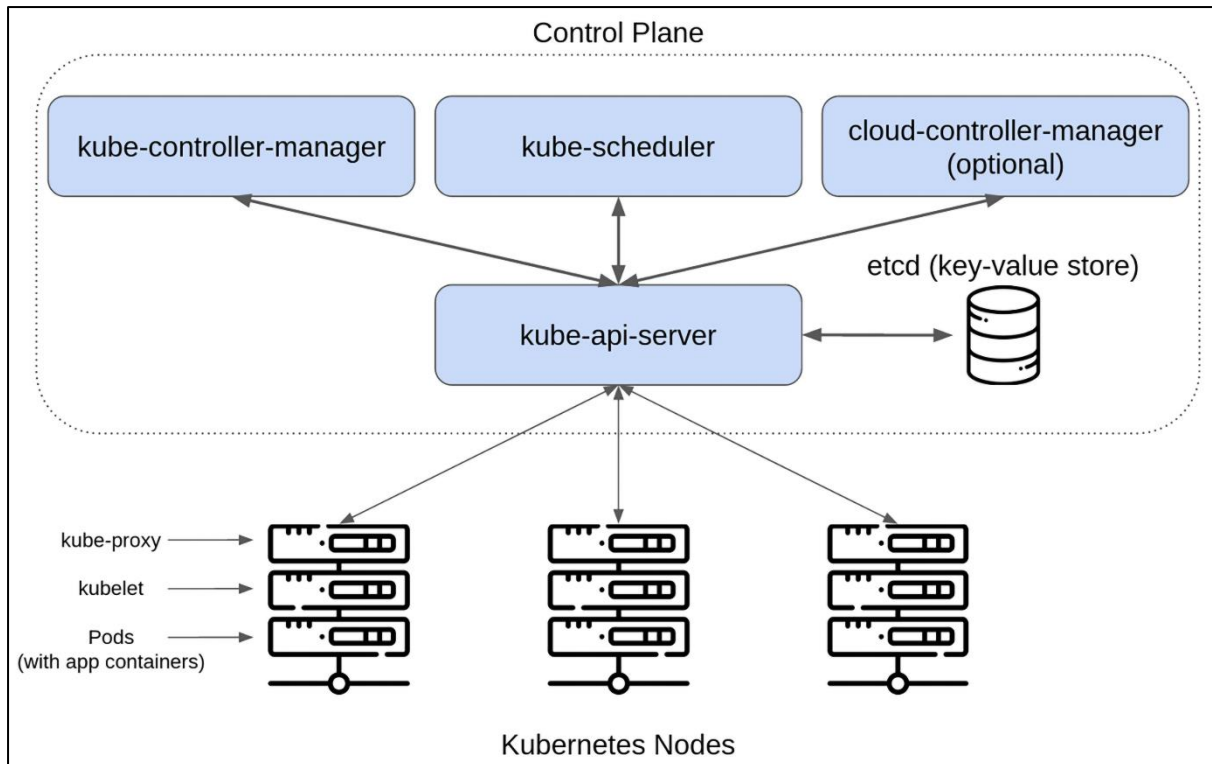
```
1 {"ivBORw0KGgoAAAANSUHEUgAAABwAAACAAAAABXZoBIAAABQGLDQ1BJQ0MgUHJvZmlsZQAAsJxjYGDisswoyGFhYGDIZSspCnJ3UoiIjFJgF8rAyCDLIMQgxsCZmFxc4BgQ4ANUwgCjUcG3a0DVQHBF2TWwmPutx07FLa+827S7Tlpx1TPQigSkktTgbSf4A4IbmQqISBgTEGyFYuLykAsRuAbJEioK0A7CkgdjqEvQLEToKw94DVhAQ5A9kXgGyB5IzEFC07AZCtk4Qkno7EhtoLAmxhwUYmFgQcSiooSa0oAdH0
```

[Body](#) [Cookies](#) [Headers \(11\)](#) [Test Results](#)

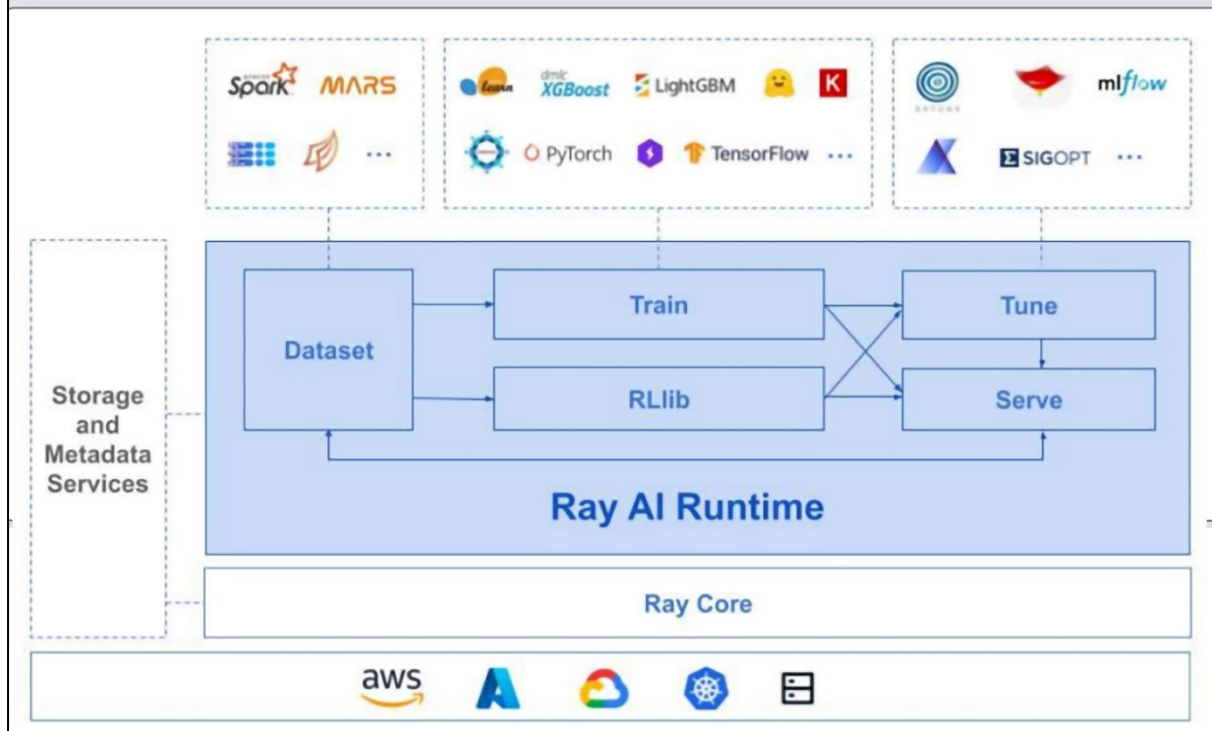
[200 OK](#) [230 ms](#) [529 B](#) [Save as Example](#)

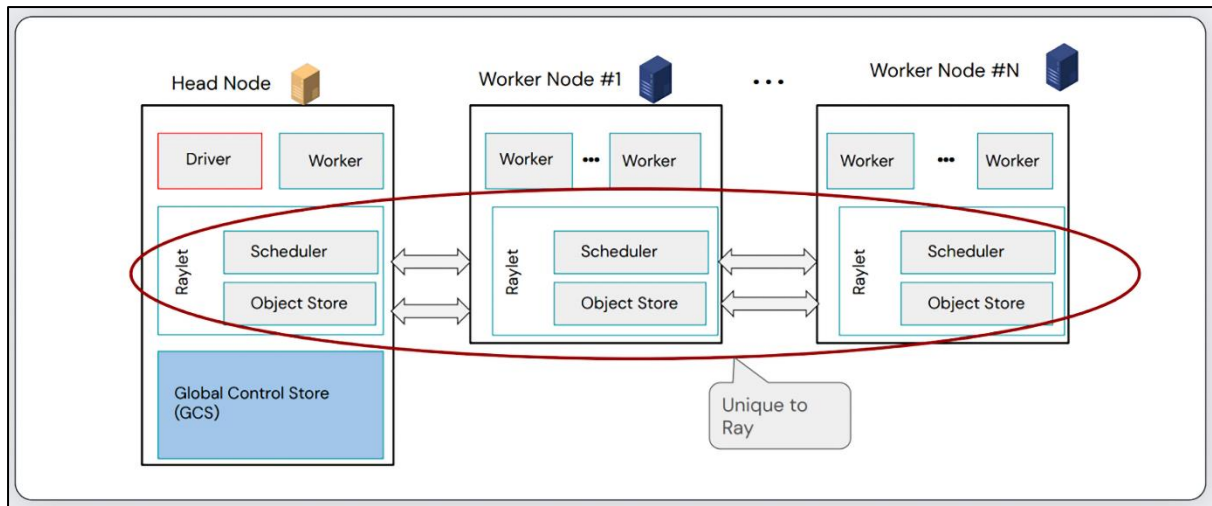
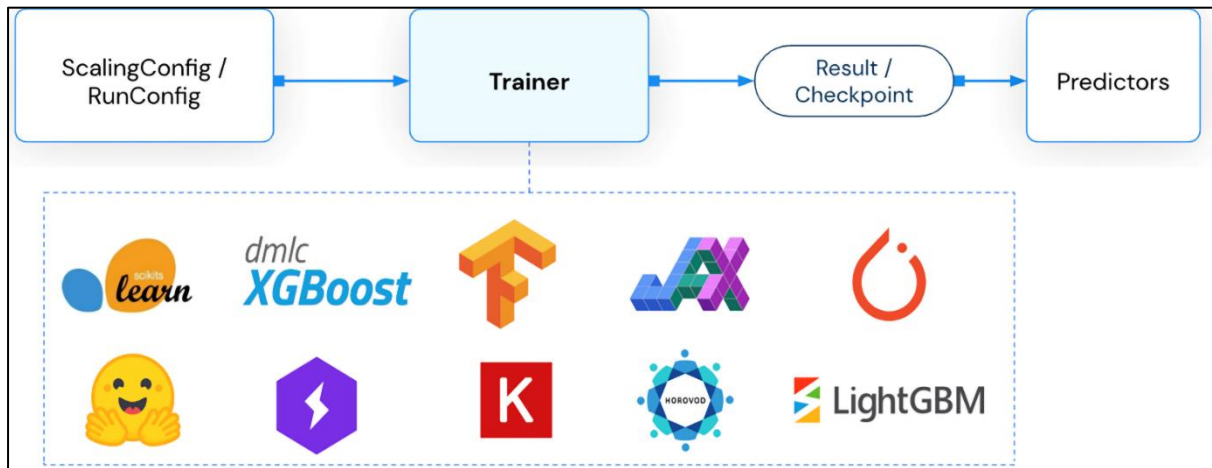
[Pretty](#) [Raw](#) [Preview](#) [Visualize](#) [JSON](#)

```
1 {"predicted_label": 3}
```



Ray AIR in Ray 2.0





Tune Status

Current time: 2023-03-14 20:33:17
Running for: 00:00:07.47
Memory: 5.2/8.0 GiB

System Info

Using FIFO scheduling algorithm.
Resources requested: 0/8 CPUs, 0/0 GPUs, 0.0/2.36 GiB heap, 0.0/1.18 GiB objects

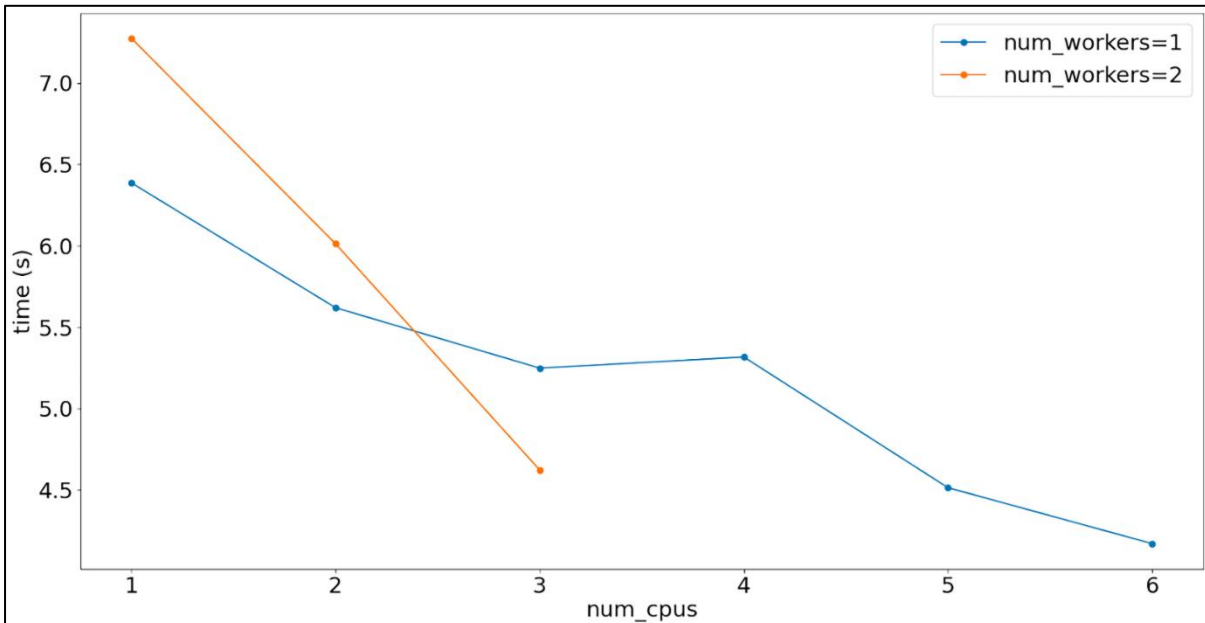
Trial Status

Trial name	status	loc	iter	total time (s)	train-logloss	train-error	valid-logloss
XGBoostTrainer_gecab_00000	TERMINATED	127.0.0.1:1713	21	6.22244	0.0184957	0	0.0897979

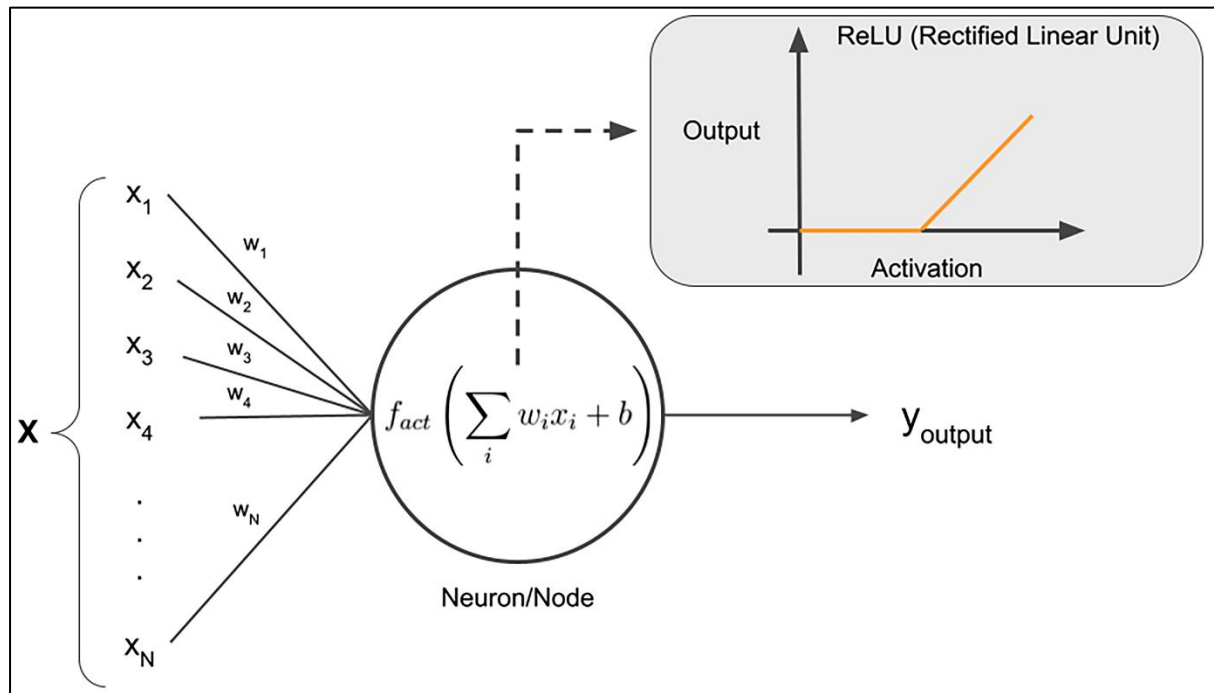
```
(XGBoostTrainer pid=1713) 2023-03-14 20:33:11,212 INFO bulk_executor.py:39 -- Executing DAG InputDataBuffer[Inp
ut] -> AllToAllOperator[aggregate]
(XGBoostTrainer pid=1713) 2023-03-14 20:33:12,242 INFO bulk_executor.py:39 -- Executing DAG InputDataBuffer[Inp
ut] -> TaskPoolMapOperator[StandardScaler]
(XGBoostTrainer pid=1713) 2023-03-14 20:33:12,264 INFO bulk_executor.py:39 -- Executing DAG InputDataBuffer[Inp
ut] -> TaskPoolMapOperator[StandardScaler]
(XGBoostTrainer pid=1713) 2023-03-14 20:33:12,289 INFO bulk_executor.py:39 -- Executing DAG InputDataBuffer[Inp
ut] -> AllToAllOperator[repartition]
(XGBoostTrainer pid=1713) 2023-03-14 20:33:12,325 INFO bulk_executor.py:39 -- Executing DAG InputDataBuffer[Inp
ut] -> AllToAllOperator[repartition]
(XGBoostTrainer pid=1713) 2023-03-14 20:33:15,807 INFO tracker.py:218 -- start listen on 127.0.0.1:49620
(_RemoteRayXGBoostActor pid=1725) [20:33:15] task [xgboost.ray]:5221498736 got new rank 1
(_RemoteRayXGBoostActor pid=1724) [20:33:15] task [xgboost.ray]:5747720096 got new rank 0
(XGBoostTrainer pid=1713) 2023-03-14 20:33:15,821 INFO tracker.py:382 -- @tracker All of 2 nodes getting starte
d
(XGBoostTrainer pid=1713) 2023-03-14 20:33:16,247 INFO tracker.py:388 -- @tracker All nodes finishes job
```

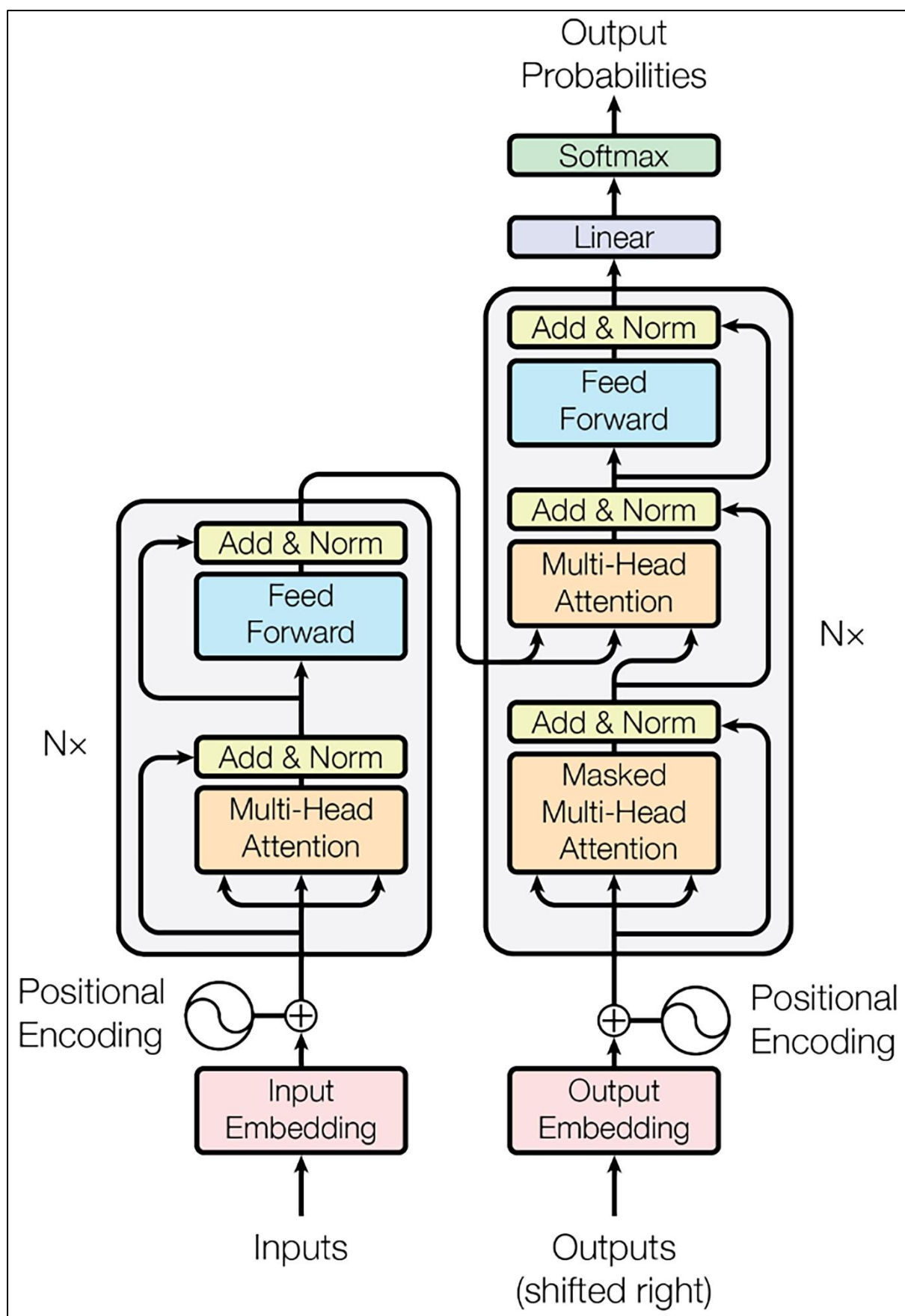
Trial Progress

Trial name	date	done	episodes_total	experiment_id	experiment_tag	hostname	iterations
XGBoostTrainer_gecab_00000	2023-03-14_20-33-17	True	2df66fa1a6b14717bed8b31470d386d4	0	Andrews-MacBook-Pro.local		
2023-03-14 20:33:17,439 INFO tune.py:798 -- Total run time: 8.19 seconds (7.47 seconds for the tuning loop).							



Chapter 7: Deep Learning, Generative AI, and LLMOps





```
1 dataset = fetch_dataset()
2 import pprint
3 pprint.pprint(dataset[100])
```

Found cached dataset amazon_reviews_multi (/Users/apmcm/.cache/huggingface/datasets_views_multi/en/1.0.0/724e94f4b0c6c405ce7e476a6c5ef4f87db30799ad49f765094cf9770e0f

```
{'language': 'en',
 'product_category': 'sports',
 'product_id': 'product_en_0610451',
 'review_body': 'Two nights in the water tide to our dock in the lake..... I\'d '
                'say something liked this.',
 'review_id': 'en_0143676',
 'review_title': 'Let the picture tell you how go this is',
 'reviewer_id': 'reviewer_en_0377453',
 'stars': 1}
```

 copilot_examples.py 2 ●

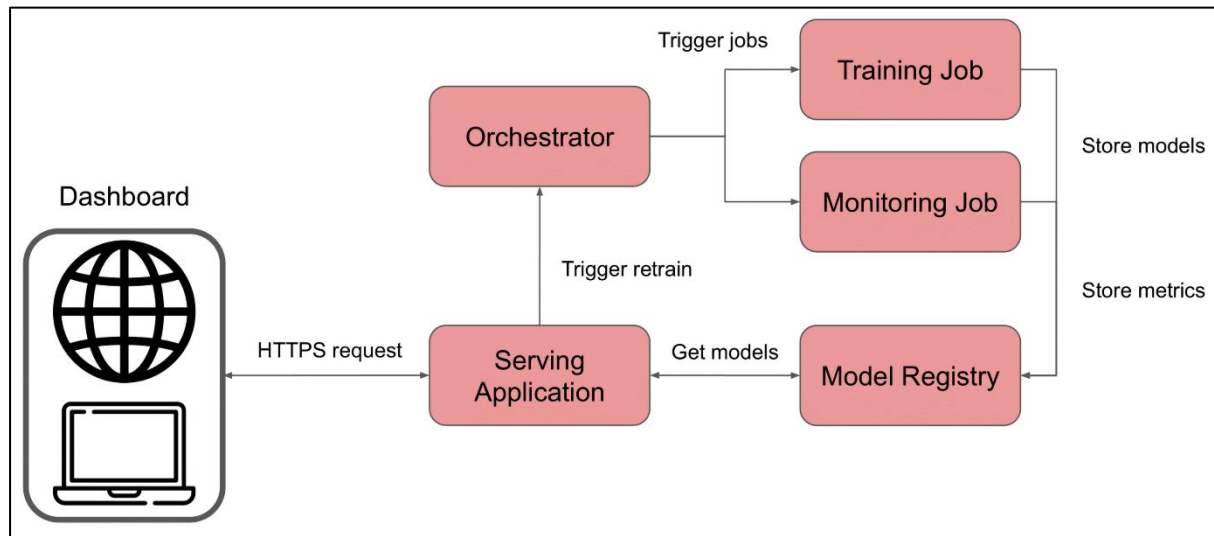
Chapter07 > example_scripts >  copilot_examples.py > ...

```
1 import pandas as pd
2 import numpy as np
3 from sklearn.datasets import load_wine
4
5 | def load_data():|
    data = load_wine()
    df = pd.DataFrame(data.data, columns=data.feature_names)
    df['target'] = data.target
    return df
6 |
```

Function to standardize numerical feature and one hot encode categorical features

```
def standardize(df):|
    df = df.copy()
    for col in df.columns:
        if df[col].dtype == 'object':
            df[col] = df[col].astype('category')
            df[col] = df[col].cat.codes
        else:
            df[col] = (df[col] - df[col].mean()) / df[col].std()
    return df
```

Chapter 8: Building an Example ML Microservice



Chapter 7 - Forecast Example / <http://127.0.0.1:8000/forecast> - local app

POST ▼ <http://127.0.0.1:8000/forecast>

Params Authorization Headers (8) **Body** ● Pre-request Script Tests Settings

☐ none ☐ form-data ☐ x-www-form-urlencoded ☒ raw ☐ binary ☐ GraphQL **JSON** ▼

```
1  [
2    {
3      "store_id": "4",
4      "begin_date": "2023-03-01T00:00:00Z",
5      "end_date": "2023-03-07T00:00:00Z"
6    },
7    {
8      "store_id": "3",
9      "begin_date": "2023-03-01T00:00:00Z",
10     "end_date": "2023-03-07T00:00:00Z"
11   },
12   {
13     "store_id": "10",
14     "begin_date": "2023-03-01T00:00:00Z",
15     "end_date": "2023-03-07T00:00:00Z"
16   }
17 ]
18
```

Body Cookies Headers (4) Test Results

Pretty Raw Preview Visualize JSON

```
1 {
2   "request": {
3     "store_id": "4",
4     "begin_date": "2023-03-01T00:00:00Z",
5     "end_date": "2023-03-07T00:00:00Z"
6   },
7   "forecast": [
8     {
9       "timestamp": "2023-03-01T00:00:00",
10      "value": 20716
11    },
12    {
13      "timestamp": "2023-03-02T00:00:00",
14      "value": 20816
15    },
16    {
17      "timestamp": "2023-03-03T00:00:00",
18      "value": 21228
19    },
20    {
21      "timestamp": "2023-03-04T00:00:00",
22      "value": 21829
23    },
24    {
25
```

Chapter 9: Building an Extract, Transform, Machine Learning Use Case

```
webserver | [2023-05-21 21:55:46 +0100] [4488] [INFO] Starting gunicorn 20.1.0
webserver | [2023-05-21 21:55:46 +0100] [4488] [INFO] Listening at: http://0.0.0.0:8080 (4488)
webserver | [2023-05-21 21:55:46 +0100] [4488] [INFO] Using worker: sync
webserver | [2023-05-21 21:55:46 +0100] [4499] [INFO] Booting worker with pid: 4499
webserver | [2023-05-21 21:55:46 +0100] [4500] [INFO] Booting worker with pid: 4500
webserver | [2023-05-21 21:55:46 +0100] [4501] [INFO] Booting worker with pid: 4501
webserver | [2023-05-21 21:55:46 +0100] [4502] [INFO] Booting worker with pid: 4502
standalone |
standalone | Airflow is ready
standalone | Login with username: admin password: pqrXyuMhWyP5sMTP
standalone | Airflow Standalone is for development purposes only. Do not use this in production!
standalone |
```

 Airflow 22:04 UTC ▾ [→\] Log In](#)

Sign In

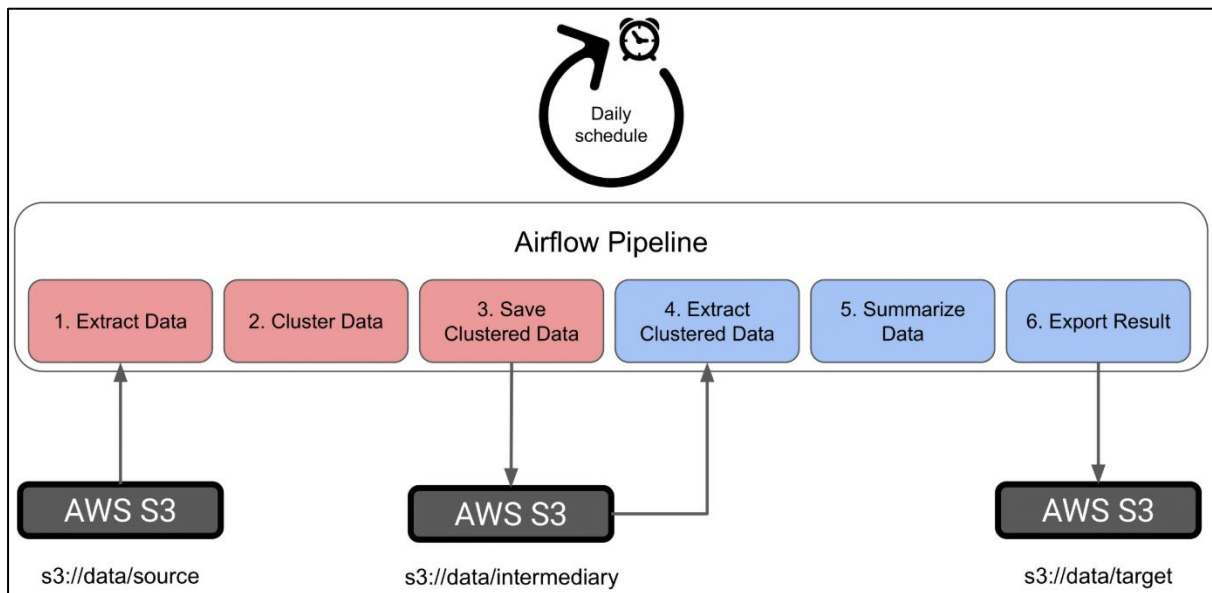
Enter your login and password below:

Username:

Password:

Sign In

<div>  Airflow </div> <div> DAGs Datasets Security Browse Admin Docs </div> <div> 22:08 UTC  </div>						
DAGs						
<div> <div> All 45 Active 0 Paused 45 </div> <div> Filter DAGs by tag </div> <div> Search DAGs </div> <div> ☒ Auto-refresh  </div> </div>						
 DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks
☐ dataset_consumes_1 consumes dataset-scheduled	airflow	   	Dataset		On s3://dag1/output_1.txt	        
☐ dataset_consumes_1_and_2 consumes dataset-scheduled	airflow	   	Dataset		0 of 2 datasets updated	        
☐ dataset_consumes_1_never_scheduled consumes dataset-scheduled	airflow	   	Dataset		0 of 2 datasets updated	        
☐ dataset_consumes_unknown_never_scheduled dataset-scheduled	airflow	   	Dataset		0 of 2 datasets updated	        
☐ dataset_produces_1 dataset-scheduled produces	airflow	   	@daily		2023-05-20, 00:00:00	        
☐ dataset_produces_2 dataset-scheduled produces	airflow	   	None			        



```

{
  "ride_dist": 5.0956885357,
  "ride_time": 0.1182137494,
  "ride_speed": 43.105717902,
  "ride_id": 20230524134,
  "selection_idx": 1,
  "news": "It is expected to be a busy shopping day today as many retailers attempt are offering discounts to try to lure shoppers back into city centre stores after the COVID-19 pandemic and lockdown. Many are expected to make there way into Glasgow city centre today to take advantage of the discounts on offer. There is also an expected surge in activity on online shopping sites.",
  "weather": "The weather is expected to be sunny and dry over the next few days, with temperatures in the mid-teens looking set to entice people out and about.",
  "traffic": "Traffic is expected to be heavy on the M8 motorway near Glasgow today due to an influx of shoppers into the city centre.",
  "label": 0
},
{
  "ride_dist": 7.2791349999,
  "ride_time": 0.4467654754,
  "ride_speed": 16.2929666702,
  "ride_id": 20230524135,
  "selection_idx": 0,
  "news": "Reports are that there has been an accident on the M8 motorway near Glasgow due to icy conditions on the roads. No one has been seriously injured but the road is blocked and traffic is backed up for miles. The police are on the scene and are expected to aid the clearing of the scene in the next few hours.",
  "weather": "The forecast for the West of Scotland over the next few days is for cold and wet weather with icy conditions to remain. Drivers are advised to only make journeys where absolutely necessary, especially since the current cold spell has already led to a number of accidents across the region.",
  "traffic": "There is a traffic jam on the M8 motorway near Glasgow. Expect delays.",
  "label": -1
},

```

```

{
  "ride_dist": 7.1756319208,
  "ride_time": 0.5520666183,
  "ride_speed": 12.9977645499,
  "ride_id": 20230524391,
  "selection_idx": 2,
  "news": "Economic conditions are slowly improving for the West of Scotland after a series of targeted investments in the area. Some high profile companies from across the globe have been lured to the Greater Glasgow Area after a targeted campaign by the Scottish Government to attract more investment in the area. The main pitch laid out to the investors has been centered around Scotland's excellent education system and it's very high quality of life for employees and customers.",
  "weather": "The forecast for the Greater Glasgow Area today remains overcast with a chance of showers and mild winds.",
  "traffic": "Traffic is expected to be normal today in the Greater Glasgow Area.",
  "label": -1,
  "summary": "Economic conditions in the West of Scotland are improving due to targeted investments and the attraction of high profile companies. The weather in the Greater Glasgow Area today will be overcast with a chance of showers and mild winds. Traffic is expected to be normal.",
  "prompt": "\n\nThe following information describes conditions relevant to taxi journeys through a single day in Glasgow, Scotland.\n\nNews: Economic conditions are slowly improving for the West of Scotland after a series of targeted investments in the area. Some high profile companies from across the globe have been lured to the Greater Glasgow Area after a targeted campaign by the Scottish Government to attract more investment in the area. The main pitch laid out to the investors has been centered around Scotland's excellent education system and it's very high quality of life for employees and customers.\n\nWeather: The forecast for the Greater Glasgow Area today remains overcast with a chance of showers and mild winds.\n\nTraffic: Traffic is expected to be normal today in the Greater Glasgow Area.\n\nSummarise the above information in 3 sentences or less.\n"
},

```

Apache Airflow

```
version          | 2.6.1
executor         | SequentialExecutor
task_logging_handler | airflow.utils.log.file_task_handler.FileTaskHandler
sql_alchemy_conn  | sqlite:///Users/apmcm/airflow/airflow.db
dags_folder      | /Users/apmcm/airflow/dags
plugins_folder    | /Users/apmcm/airflow/plugins
base_log_folder   | /Users/apmcm/airflow/logs
remote_base_log_folder |
```

System info

```
OS | Mac OS
architecture | arm
uname | uname_result(system='Darwin', node='Andrews-MacBook-Pro.local',
locale | ('en_US', 'UTF-8')
python_version | 3.10.8 | packaged by conda-forge | (main, Nov 22 2022, 08:25:01) [Clang 14.0.0]
python_location | /opt/homebrew/Caskroom/miniforge/base/envs/mlewp-chapter08/bin/python3
```

Tools info

```
git | git version 2.39.1
ssh | OpenSSH_9.0p1, LibreSSL 3.3.6
kubectl | NOT AVAILABLE
gcloud | NOT AVAILABLE
cloud_sql_proxy | NOT AVAILABLE
mysql | NOT AVAILABLE
sqlite3 | 3.39.5 2022-10-14 20:58:05 554764a6e721fab307c63a4f98cd958c8
psql | NOT AVAILABLE
```

Paths info

```
airflow_home | /Users/apmcm/airflow
system_path   | /opt/homebrew/Caskroom/miniforge/base/envs/mlewp-chapter08/t
               | texes/App/usr/bin:/usr/bin:/bin:/usr/sbin:/sbin:/var/run/cor
               | xd/codex.system/bootstrap/usr/appleinternal/bin
python_path   | /opt/homebrew/Caskroom/miniforge/base/envs/mlewp-chapter08/t
               | rew/Caskroom/miniforge/base/envs/mlewp-chapter08/lib/python3
               | /apmcm/airflow/plugins
airflow_on_path | True
```

Providers info

```
apache-airflow-providers-apache-spark | 4.0.1
apache-airflow-providers-common-sql  | 1.4.0
apache-airflow-providers-ftp          | 3.3.1
apache-airflow-providers-http         | 4.3.0
apache-airflow-providers-imap         | 3.1.1
apache-airflow-providers-sqlite       | 3.3.2
```

Airflow

DAGs

Datasets

Security

Browse

Admin

Docs

20:27 UTC

AU

The scheduler does not appear to be running. Last heartbeat was received 35 seconds ago.

The DAGs list may not update, and new tasks will not be scheduled.

DAG: etml_dag

running

Schedule: @daily

Next Run: 2023-05-24, 00:00:00

Grid

Graph

Calendar

Task Duration

Task Tries

Landing Times

Gantt

Details

<> Code

Audit Log

2023-05-23T00:00:01Z

Runs

25

Run

scheduled__2023-05-23T00:00:00+00:00

Layout

Left > Right

Update

Find Task...

PythonOperator

deferred

failed

queued

removed

restarting

running

scheduled

shutdown

skipped

success

up_for_reschedule

up_for_retry

upstream_failed

no_status

extract_cluster_save

→

extract_summarize

Auto-refresh

mlewp-ch9-etml-data

Info

Objects

Properties

Permissions

Metrics

Management

Access Points

Objects (3)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Refresh

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

< 1 >

	Name	Type	Last modified	Size	Storage class
☐	 clustered_data_20230524.json	json	May 24, 2023, 12:19:31 (UTC+01:00)	318.0 KB	Standard
☐	 clustered_summarized_20230524.json	json	May 24, 2023, 12:23:35 (UTC+01:00)	668.8 KB	Standard
☐	 taxi-rides-20230524.json	json	May 24, 2023, 09:55:15 (UTC+01:00)	350.0 KB	Standard