



# MLOps : Put the Data Science Under Control

---

*a crowd surrounds three large  
columns, in the background the sun  
rises, Surrealism (DALL-E)*

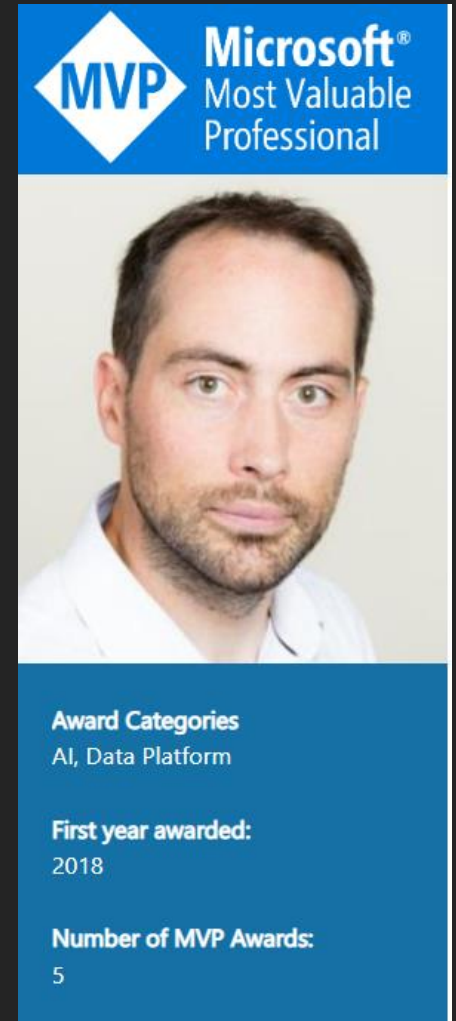


# Speaker : Paul PETON

- Lead Tech in Artificial Intelligence
- Data Platform & AI MVP since 2018
- French speaking podcast



- [GitHub.com/methodidacte/meetups](https://github.com/methodidacte/meetups)



*Everything is against them but they will have to work together...*



## David Data – Data Scientist



**Main goal: to get the best model**

- prepares the data
- experiments in a notebook
- plots the results of the various tests

## Angie Neer – ML Engineer

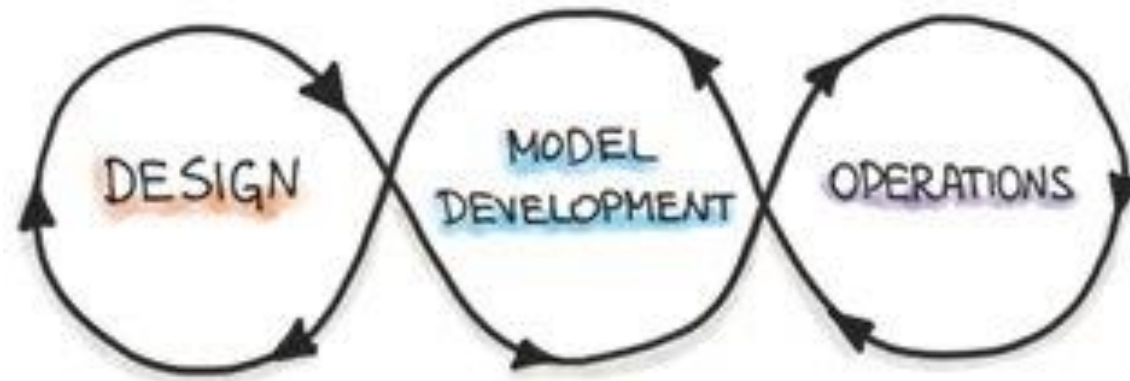


**Main goal: to guarantee the performance of the service in production**

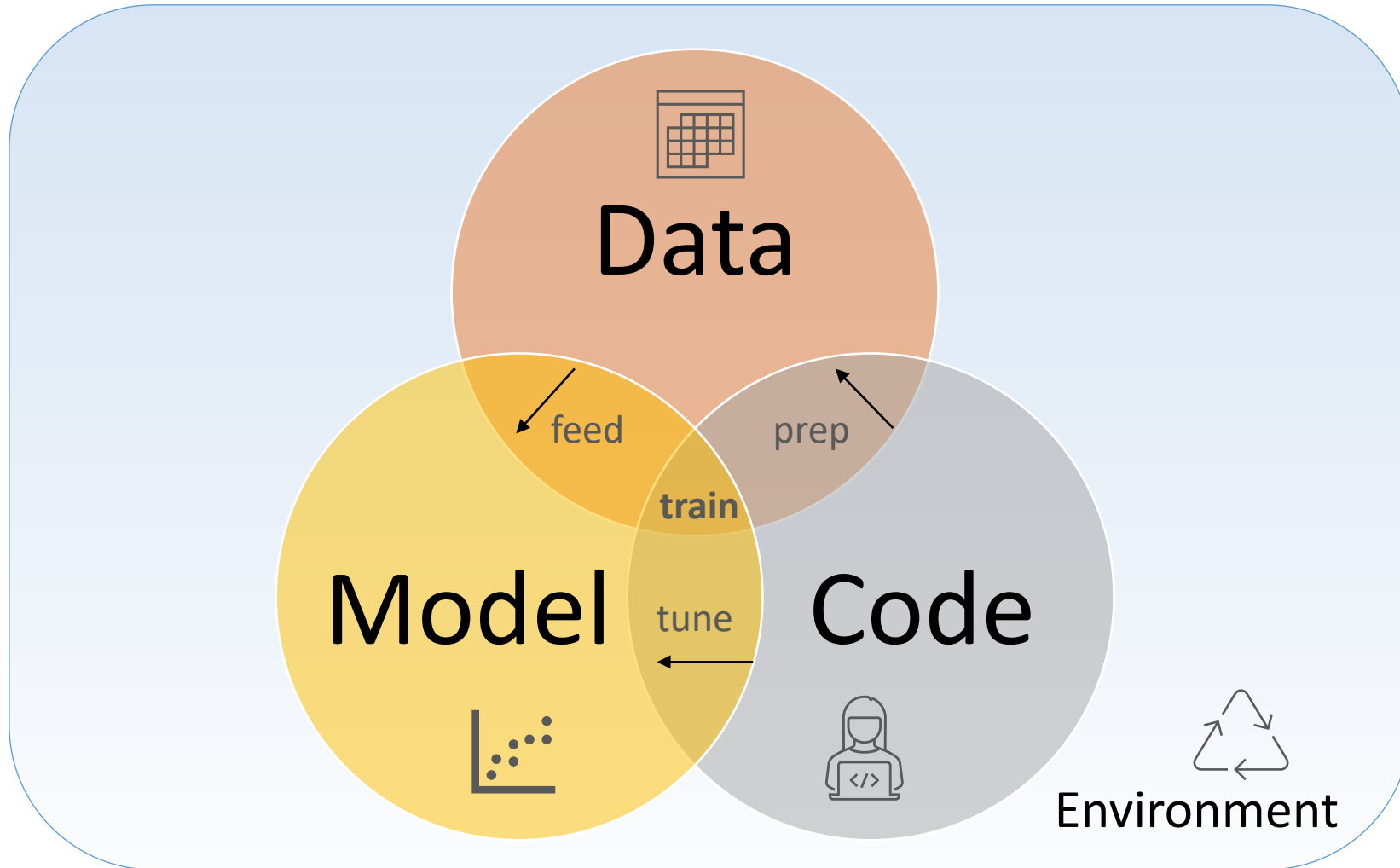
- integrates the deliverables
- deploys new versions
- orchestrates and monitors processes

# MLOps : quick definition

Provide an **end-to-end** Machine Learning development process to **design**, **create**, and **manage** repeatable, testable, and scalable software.

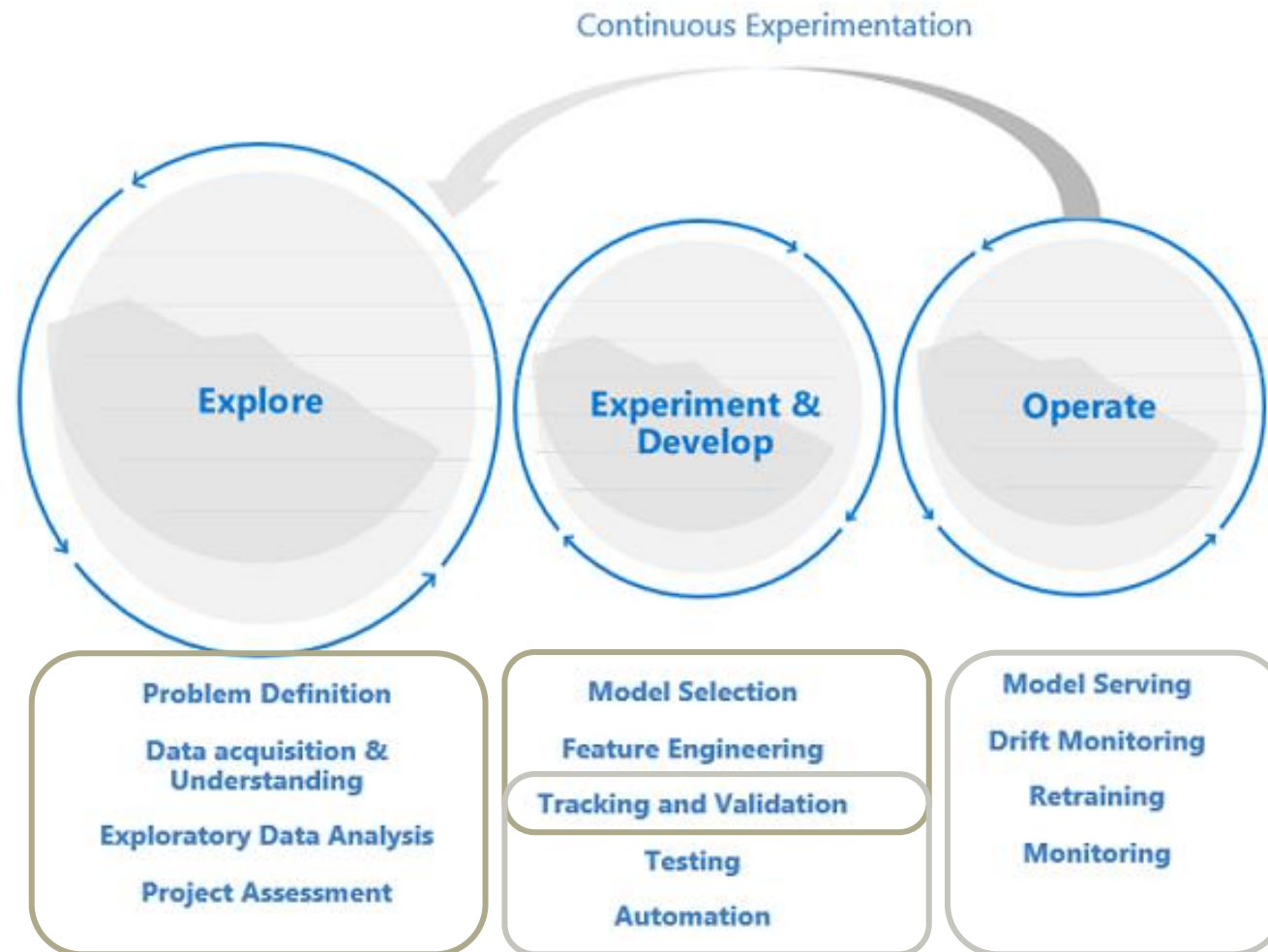


# What we need to put under control



The environment allows the execution of the code (OS, Python, libraries...).

# The three collaborative loops of MLOps



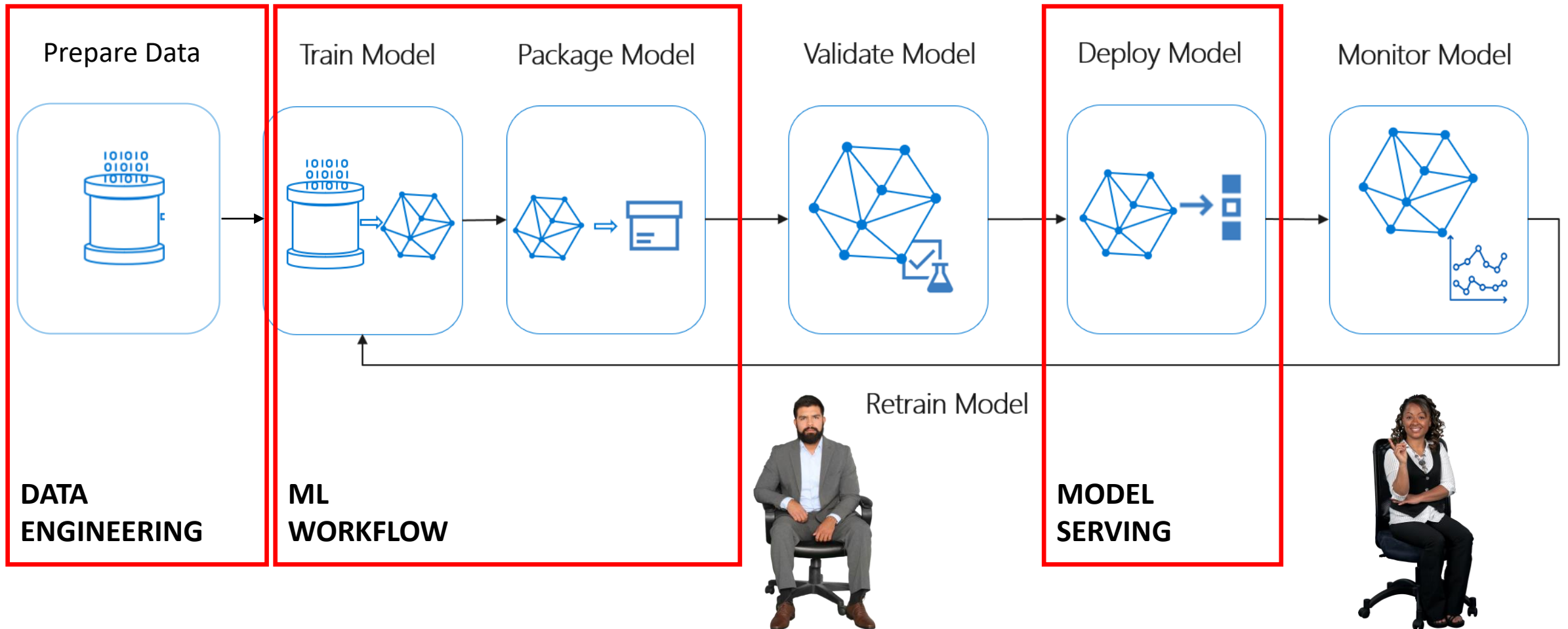


A person wearing a blue suit is riding a green bicycle on a city street. The background shows a stone wall with arched windows and a parked motorcycle. The image has a dark overlay.

# Workflow

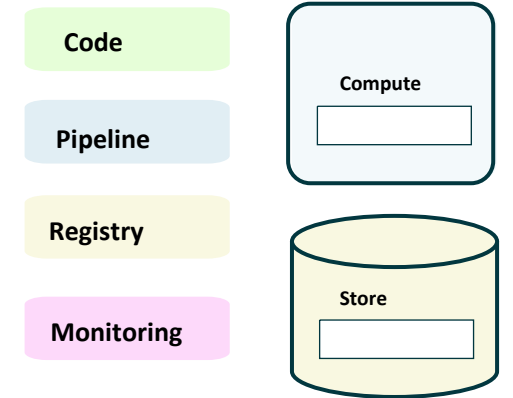
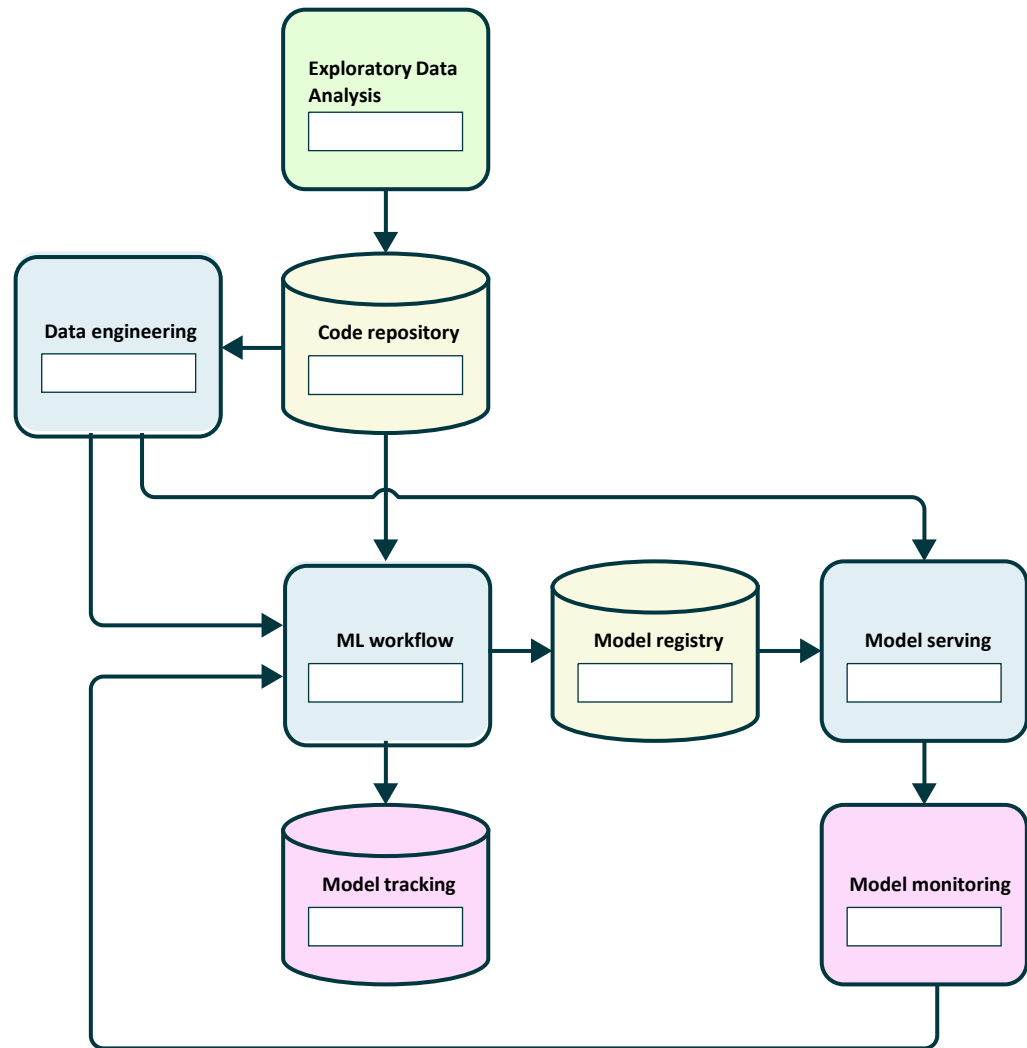
Define the processes, share the roles

# Focus on the three main pipelines





# MLOps end to end process



## MLOps Stack Template

Version 2.0  
January 2023

<https://valohai.com/blog/the-mlops-stack/>  
<https://www.linkedin.com/in/skogstrom/>



Experiment



## Experiment with parameters for a Ridge Regression the Diabetes Dataset

This notebook is for experimenting with different parameters to train a ridge regression model on the Diabetes

```
# Change out of the experimentation directory
%cd ..
```

```
import azureml.core
from azureml.core import Workspace
```

```
# Load the workspace from the saved config file
ws = Workspace.from_config()
```

```
import os, shutil

# Create a folder for the experiment files
training_folder = 'diabetes-training'
```

Diabetes Ridge Regression Parameter Experimentation.ipynb ×

experimentation > Diabetes Ridge Regression Parameter Experimentation.ipynb > ...

```
Open in Notebook Editor
1  [
2  {
3    "cells": [
4      {
5        "cell_type": "markdown",
6        "metadata": {},
7        "source": [
8          "# Experiment with parameters for a Ridge Regression Model on the Diabetes Da
9        ],
10       },
11       {
12         "cell_type": "markdown",
13         "metadata": {},
14         "source": [
15           "This notebook is for experimenting with different parameters to train a ridg
16         ],
17       },
18       {
19         "cell_type": "code",
20         "execution_count": null,
21         "metadata": {},
22         "outputs": [],
23         "source": [
24           "# Change out of the experimentation directory\n",
25           "%cd .."
26         ],
27       }
28     ]
29   }
```

What about notebooks ?

- Includes comments, charts, results...
- Can't read the file content in other tool

# Improve your code

---

---

Reproducibility

---

Version control

---

Debug

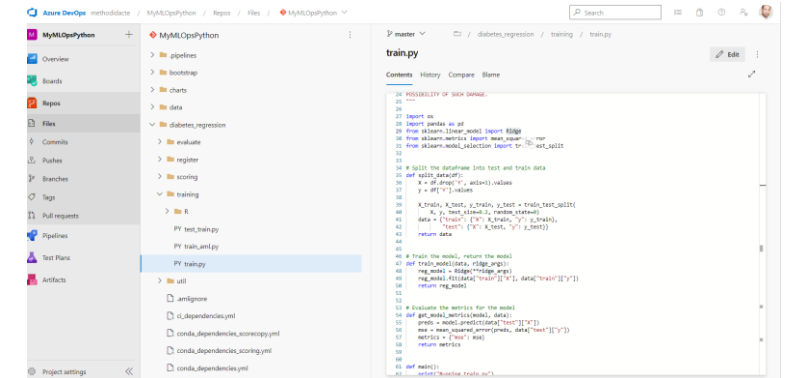
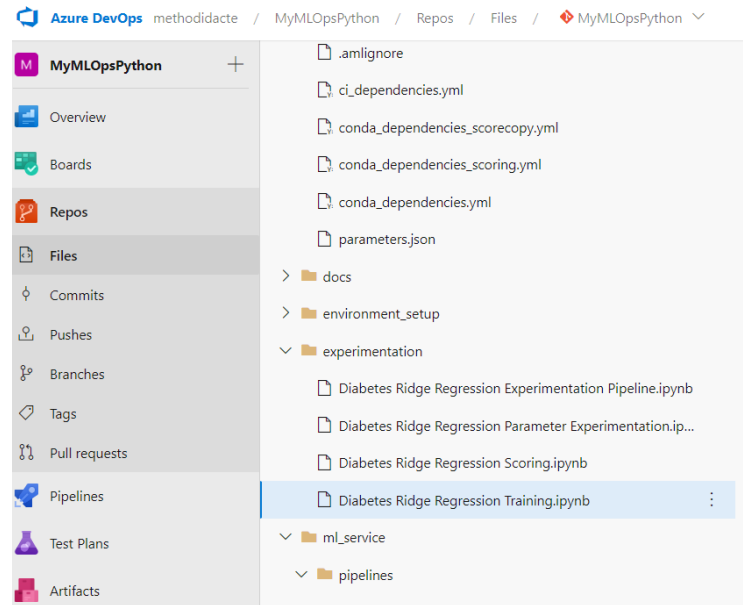
---

Test and reuse

---

CI/CD





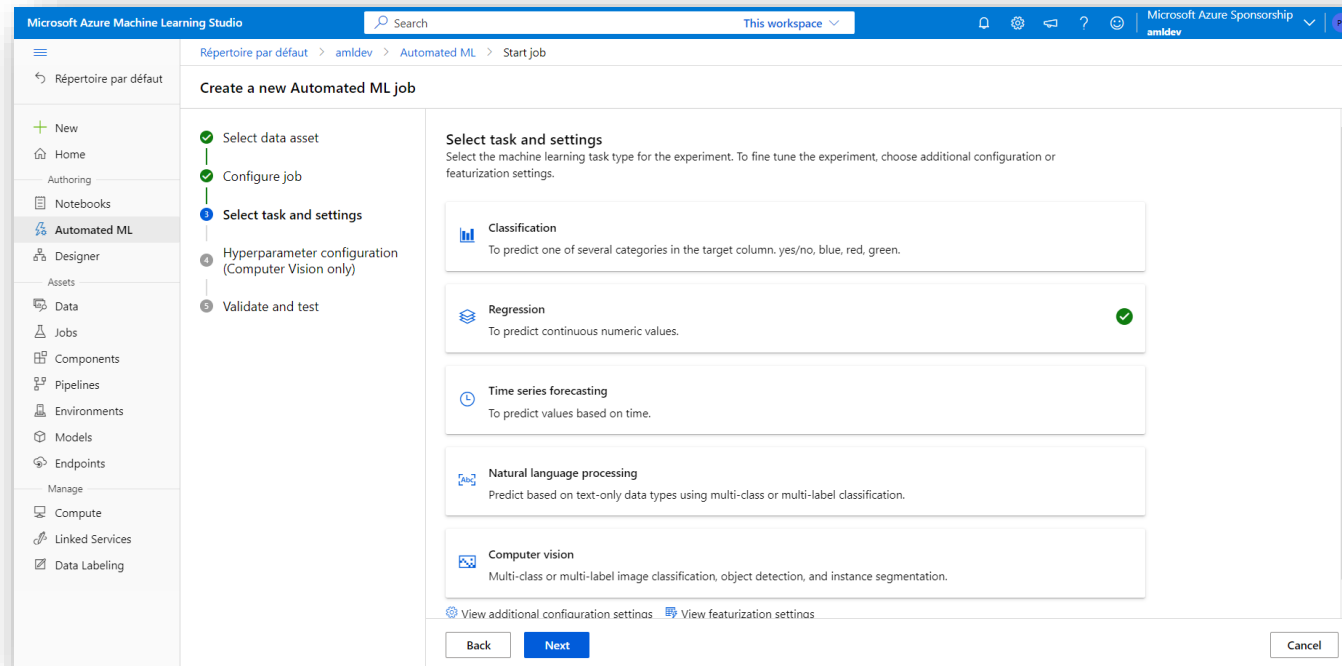
From notebook to  
python script

- convert the code into functions hosted in scripts
- Use the repository structure to reduce complexity and scale our work



# Experiment tracking

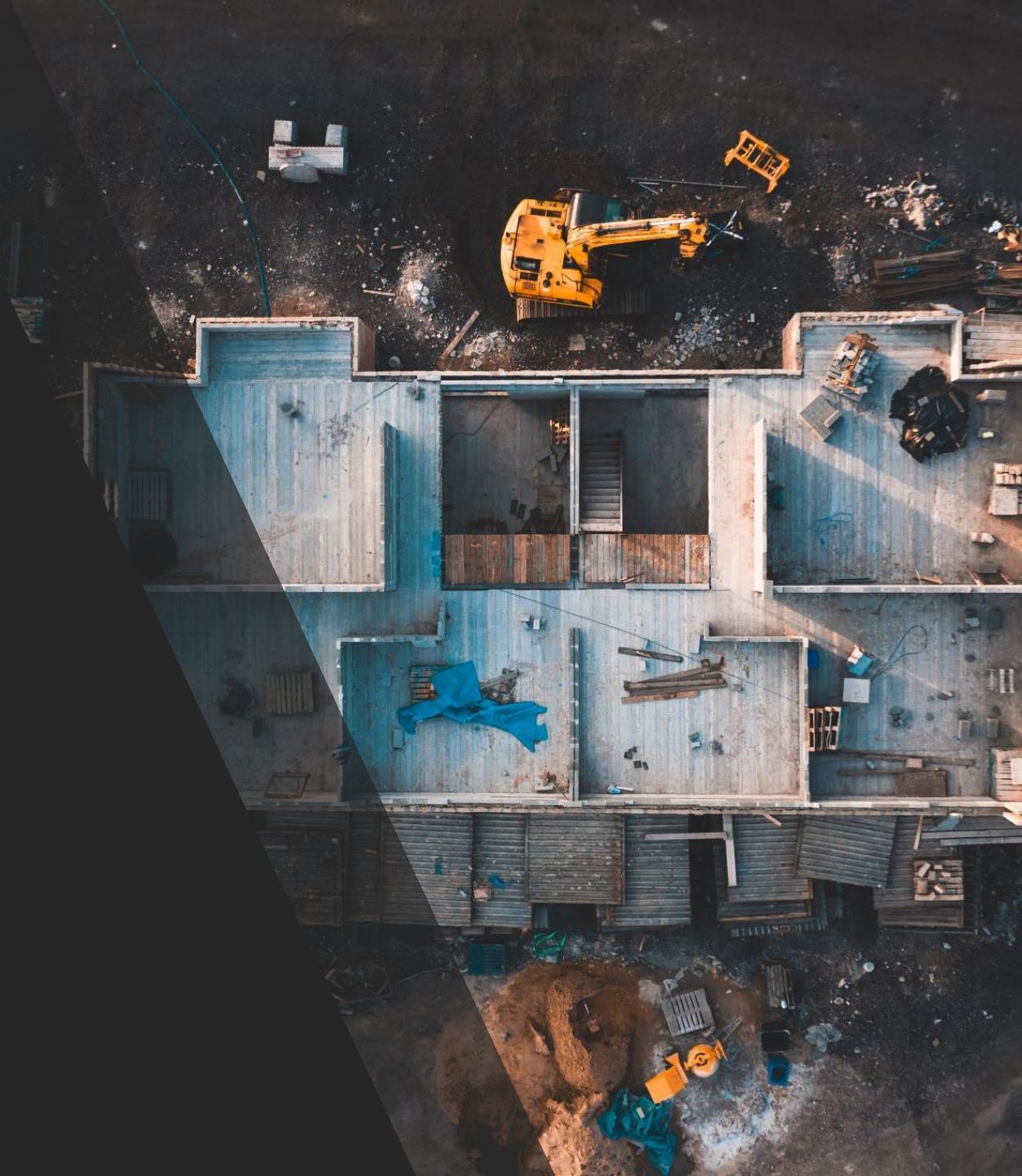
- Use **automated ML** to go quickly to the best model



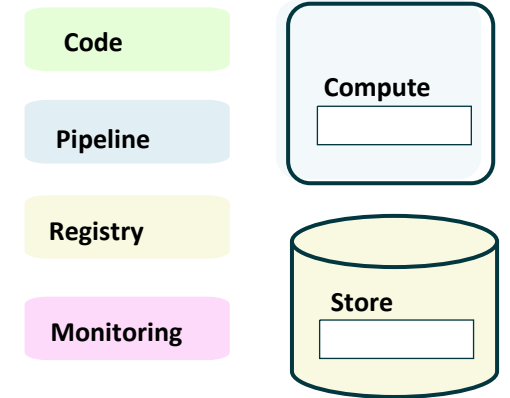
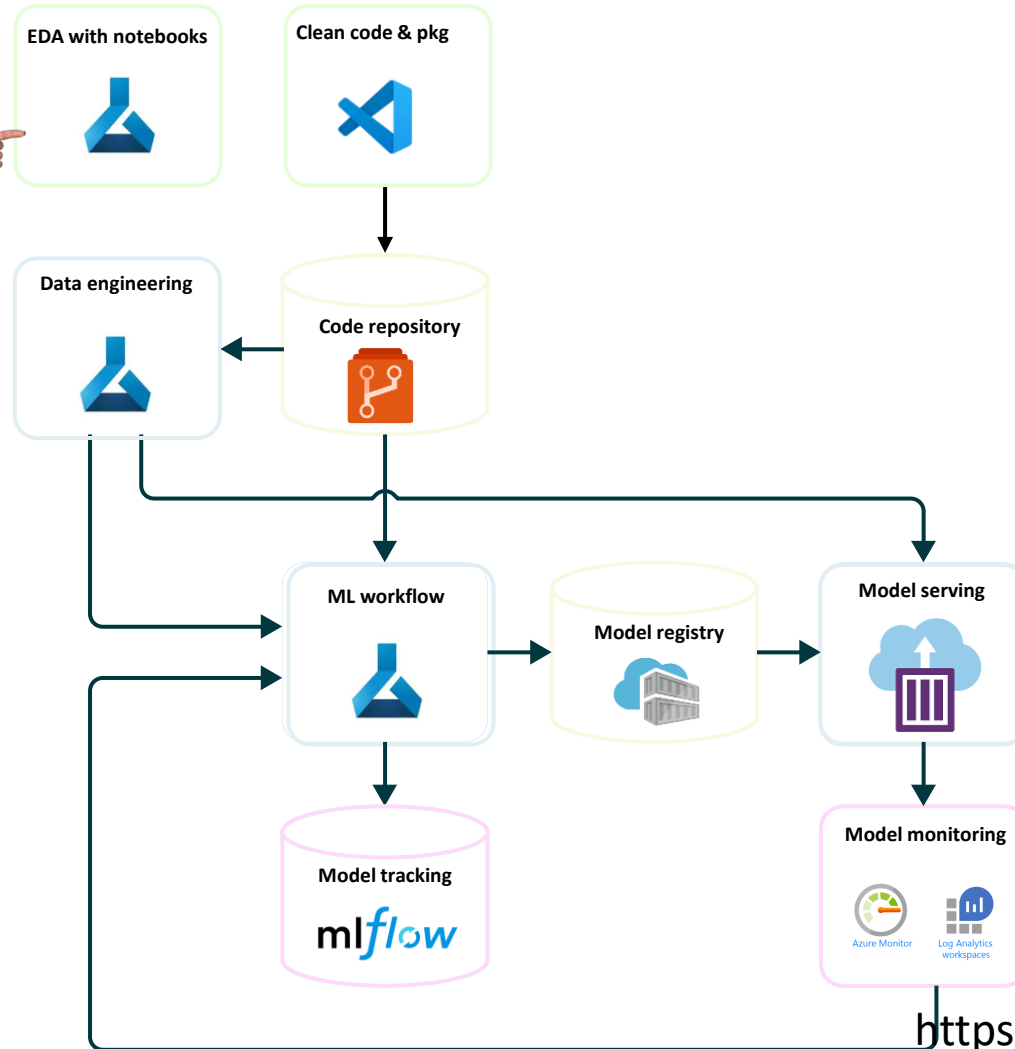
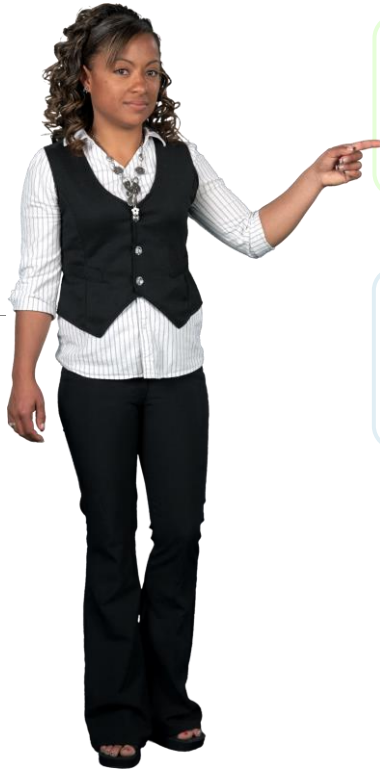
- Follow the performance with the evaluation score



Build the platform



# Technical stack components

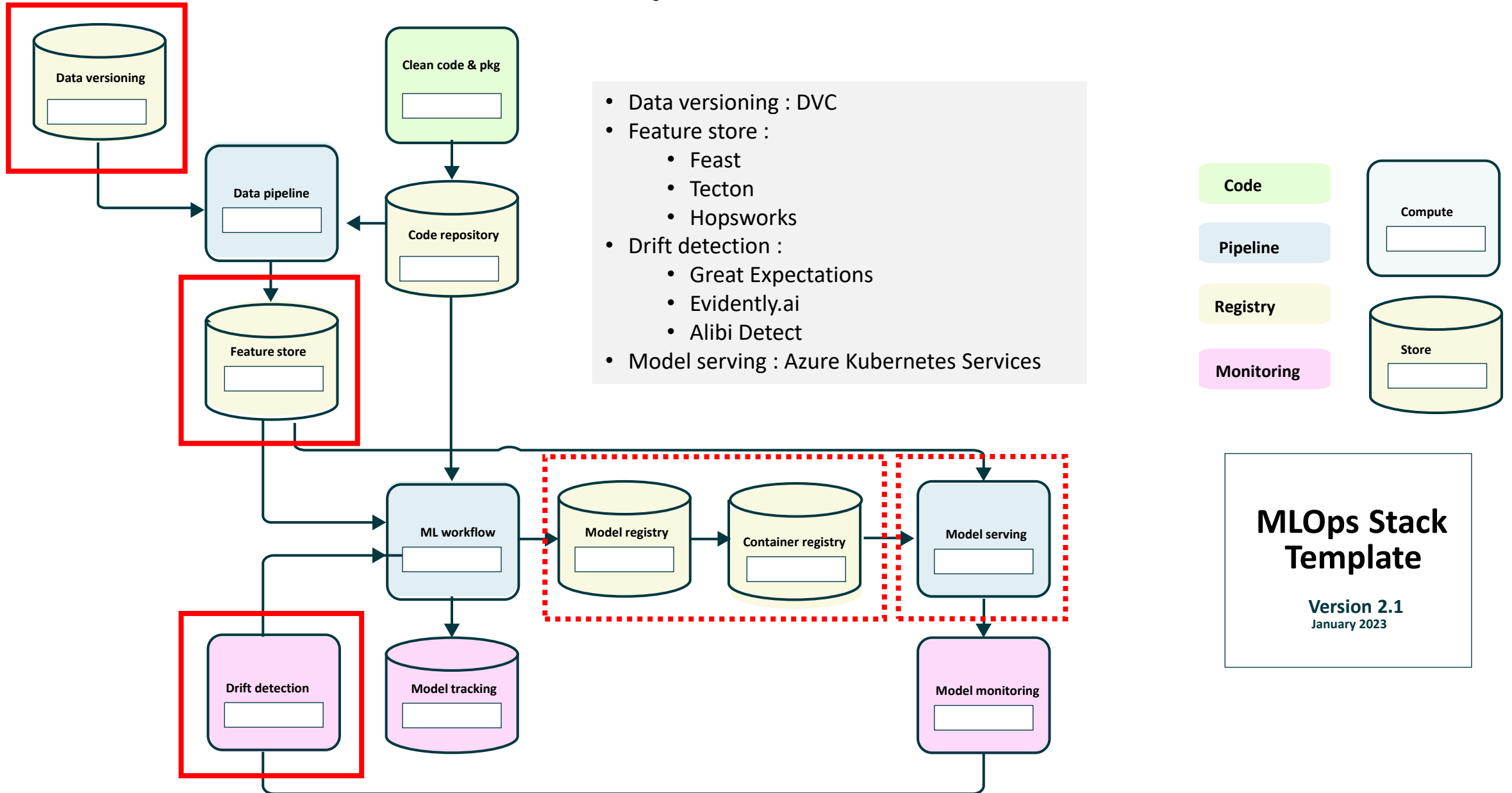


**MLOps Stack Template**

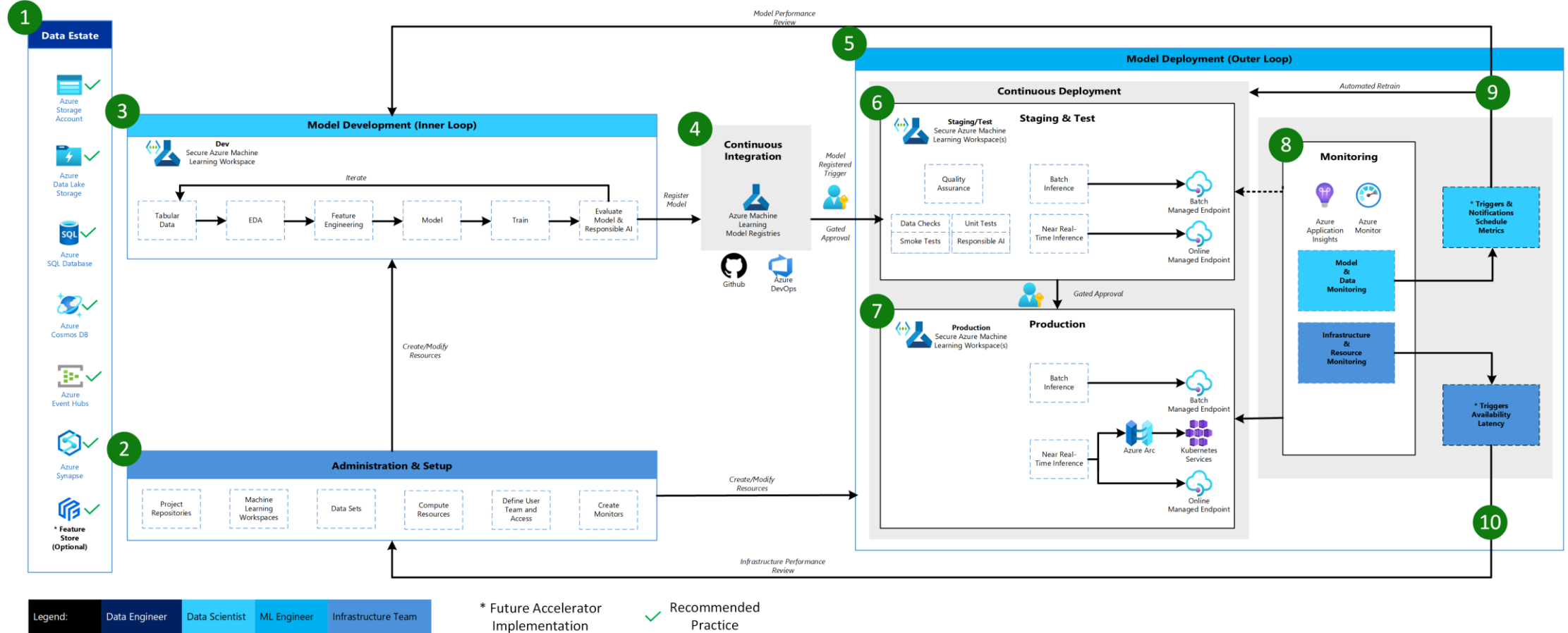
Version 2.0  
January 2023

<https://valohai.com/blog/the-mlops-stack/>  
<https://www.linkedin.com/in/skogstrom/>

# Technical stack components (advanced version)



# MLOps architecture for classical ML on Azure



[GitHub - Azure/mlops-v2: Azure MLOps \(v2\) solution accelerators](https://github.com/Azure/mlops-v2)





# DEMO

Automate the deployments



# Conclusion : *better together !*



Share a culture of **software engineering** (clean code, testing) and DevOps (automation)

Advance by **maturity level** (registers, automation, monitoring)

The platform will become **exponentially more efficient** as new use cases are implemented.

Microsoft Azure Machine Learning Studio

Search

This workspace

Microsoft Azure Sponsorship

amldev

Répertoire par défaut

amldev

Automated ML

Start job

Create a new Automated ML job

1

Select data asset

2

Configure job

3

Select task and settings

4

Hyperparameter configuration  
(Computer Vision only)

5

Validate and test

Configure job

Select from existing experiments or create a new experiment, then select the target column and training compute.

[Learn more on how to configure the experiment.](#)

Data asset

diabetes [\(View data asset\)](#)

Experiment name

Select existing

Create new

New experiment name \*

diabetes\_autoML

Target column \*

Y (Decimal)

Select compute type

Compute cluster

Select Azure ML compute cluster \*

aml-cluster

+ New

Refresh computes

Back

Next

Cancel

Create a new Automated ML job

✓

Select data asset

✓

Configure job

3

Select task and settings

4

Hyperparameter configuration  
(Computer Vision only)

5

Validate and test

Select task and settings

Select the machine learning task type for the experiment. To fine tune the experiment, choose additional configuration or featurization settings.

Classification

To predict one of several categories in the target column. yes/no, blue, red, green.

Regression

To predict continuous numeric values.

✓

Time series forecasting

To predict values based on time.

Natural language processing

Predict based on text-only data types using multi-class or multi-label classification.

Computer vision

Multi-class or multi-label image classification, object detection, and instance segmentation.

View additional configuration settings

View featurization settings

Back

Next

Cancel


## Additional configurations



### Primary metric

Normalized root mean squared error



☒ Explain best model 


☒ Use all supported models 

### Blocked models


FastLinearRegressor, LassoLars, SGD, ExtremeRandomTrees



### Exit criterion


Training job time (hours) 

1

Metric score threshold 

Metric score threshold

### Concurrency

Max concurrent iterations 

4

Save

Cancel



## Featurization



Feature selection identifies the actions performed on the data asset to prepare the data for training. This will not impact the input data needed for inferencing i.e., if columns are excluded from training, the excluded columns will still be required as input for inferencing on the model. [Learn more about Automated ML's featurization.](#)

☒ Enable featurization

☐ Show only selected rows (10 selected)

<input checked="" type="checkbox"/>	Column name	Feature type	Impute with	Data example
<input checked="" type="checkbox"/>	AGE	Auto	Auto	0.0380759064334241, -0.0
<input checked="" type="checkbox"/>	SEX	Auto	Auto	0.0506801187398187, -0.0
<input checked="" type="checkbox"/>	BMI	Numeric	Auto	0.0616962065186885, -0.0
<input checked="" type="checkbox"/>	BP	DateTime	Auto	0.0218723549949558, -0.0
<input checked="" type="checkbox"/>	S1	Categorical	Auto	0.0218723549949558, -0.0
<input checked="" type="checkbox"/>	S2	Categorical hash	Auto	-0.0442234984244464, -0.0
<input checked="" type="checkbox"/>	S3	Text	Auto	-0.0348207628376986, -0.0
<input checked="" type="checkbox"/>	S4	Auto	Auto	-0.0434008456520269, 0.0
<input checked="" type="checkbox"/>	S5	Auto	Auto	-0.00259226199818282, -0.0
<input checked="" type="checkbox"/>	S6	Auto	Auto	0.0199084208763183, -0.0
<input checked="" type="checkbox"/>	S7	Auto	Auto	-0.0176461251598052, -0.0

Save

Cancel

Microsoft Azure Machine Learning Studio

Search

This workspace

Microsoft Azure Sponsorship  
amldev

Répertoire par défaut

amldev

Automated ML

Start job

Create a new Automated ML job

✓

Select data asset

✓

Configure job

✓

Select task and settings

✓

Hyperparameter configuration  
(Computer Vision only)

5

Validate and test

Select the validation and test type

You can choose a validation type and select a test data asset as an optional step. Providing your own validation and test data assets are currently preview features.

Validation type ⓘ

Auto

Auto

k-fold cross validation

Monte Carlo cross validation

Train-validation split

User validation data

Back

Finish

Cancel



Microsoft Azure Machine Learning Studio

Search

This workspace

1

Microsoft Azure Sponsorship  
amldev

Répertoire par défaut

New

Home

Authoring

Notebooks

Automated ML

Designer

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Linked Services

Data Labeling

Répertoire par défaut > amldev > Jobs > diabetes\_autoML > careful\_deer\_lyrygkdd

careful\_deer\_lyrygkdd

☆

✔ Completed

Overview

Data guardrails

Models

Outputs + logs

Child jobs

🔄 Refresh

⏮ Edit and submit (preview)

+ Register model

⊗ Cancel

🗑 Delete

|

▶ Deploy

⬇ Download

🔍 Explain model

# View generated code

📊 Edit columns

↺ Reset view

🔍 Search

Created on

🔍 All filters

✕ Clear all

Showing 1-25 of 45 models

Page size: 25

Algorithm name	Explained	Normalized root mean squared error ↑	Sampling	Created on	Dur...	Hyperparameter
MaxAbsScaler, ElasticNet		0.17039	100.00 %	Feb 26, 2023 9:18 PM	28s	alpha : 0.001 l1_ratio : 1 norm...
StandardScalerWrapper, ElasticNet		0.17039	100.00 %	Feb 26, 2023 9:18 PM	30s	alpha : 0.001 l1_ratio : 0.8436842 ...
StandardScalerWrapper, ElasticNet		0.17040	100.00 %	Feb 26, 2023 9:18 PM	31s	alpha : 0.001 l1_ratio : 0.2705263 ...
MaxAbsScaler, ElasticNet		0.17040	100.00 %	Feb 26, 2023 9:18 PM	29s	alpha : 0.001 l1_ratio : 0.8436842 ...
StandardScalerWrapper, ElasticNet		0.17041	100.00 %	Feb 26, 2023 9:18 PM	29s	alpha : 0.001 l1_ratio : 1 norm...
MaxAbsScaler, ElasticNet		0.17043	100.00 %	Feb 26, 2023 9:18 PM	30s	alpha : 0.001 l1_ratio : 0.6873684 ...
MinMaxScaler, ElasticNet		0.17050	100.00 %	Feb 26, 2023 9:18 PM	29s	alpha : 0.001 l1_ratio : 0.8436842 ...
StandardScalerWrapper, ElasticNet		0.17070	100.00 %	Feb 26, 2023 9:18 PM	29s	alpha : 0.05357894736842105 l1 ...
StandardScalerWrapper, ElasticNet		0.17072	100.00 %	Feb 26, 2023 9:18 PM	29s	alpha : 0.05357894736842105 l1 ...
StandardScalerWrapper, ElasticNet		0.17109	100.00 %	Feb 26, 2023 9:18 PM	31s	alpha : 0.001 l1_ratio : 0.7394736 ...
RobustScaler, ElasticNet		0.17285	100.00 %	Feb 26, 2023 9:18 PM	29s	alpha : 0.001 l1_ratio : 0.2184210 ...

<<

<

Page 1 of 2

>

>>

...

amldev




Jobs

diabetes\_autoML

careful\_deer\_lyrygkdd

bright\_feast\_5xyyk7qk

bright\_feast\_5xyyk7qk



Overview

Model

Explanations (preview)

Metrics

Data transformation (preview)

Test results (preview)


Outputs + logs


Images


Child jobs


Code


Monitoring


 Refresh


 Deploy


 Download


 Explain model

 View generated code

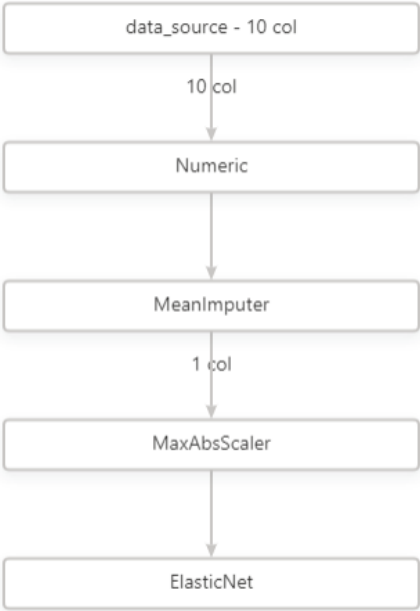
 Test model (preview)

 Register model

 Cancel

 Delete

The following diagram illustrates the data preprocessing, feature engineering, scaling techniques and the machine learning algorithm that Automated ML applied to generate this particular model.



```
graph TD; A["data_source - 10 col"] -- "10 col" --> B["Numeric"]; B --> C["MeanImputer"]; C -- "1 col" --> D["MaxAbsScaler"]; D --> E["ElasticNet"];
```



Microsoft Azure Machine Learning Studio

Search

This workspace

1

Microsoft Azure Sponsorship

amldev

Répertoire par défaut

New

Home

Authoring

Notebooks

Automated ML

Designer

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Linked Services

Data Labeling

amldev > Jobs > diabetes\_autoML > careful\_deer\_lyrygkdd > bright\_feast\_5xyyk7qk

bright\_feast\_5xyyk7qk

Completed

Overview

Model

Explanations (preview)

Metrics

Data transformation (preview)

Test results (preview)

Outputs + logs

Images

Refresh

Deploy

Download

Explain model

View generated code

Test model (preview)

Register model

Cancel

Model summary

Algorithm name

MaxAbsScaler, ElasticNet

Hyperparameters

[View hyperparameters](#)

Normalized root mean squared error

0.17039 [View all other metrics](#)

Sampling

100.00 %

Registered models

No registration yet

Deploy status

No deployment yet

Hyperparameters

Data transformation:

```
1 {
2   "class_name": "MaxAbsScaler",
3   "module": "sklearn.preprocessing",
4   "param_args": [],
5   "param_kwargs": {},
6   "prepared_kwargs": {},
7   "spec_class": "preproc"
8 }
```

Training algorithm:

```
1 {
2   "class_name": "ElasticNet",
3   "module": "sklearn.linear_model",
4   "param_args": [],
5   "param_kwargs": {
6     "alpha": 0.001,
7     "l1_ratio": 1,
8     "normalize": false
9   },
10  "prepared_kwargs": {},
11  "spec_class": "sklearn"
12 }
```

Close

Microsoft Azure Machine Learning Studio

Search

This workspace

1

Microsoft Azure Sponsorship  
amldev

pp

...

>

amldev

>

Jobs

>

diabetes\_autoML

>

careful\_deer\_lyrygkdd

>

bright\_feast\_5xyyk7qk

bright\_feast\_5xyyk7qk

Completed

Overview

Model

Explanations (preview)

Metrics

Data transformation (preview)

Test results (preview)

Outputs + logs

Images

Child jobs

Code

Monitoring

Refresh

Cancel

Create custom chart

View as...

Current view: Local

Edit view

Select metrics

explained\_variance

0.4787709

mean\_absolute\_error

44.36141

mean\_absolute\_perce...

39.66555

median\_absolute\_error

38.45047

normalized\_mean\_abs...

0.1381976

normalized\_median\_a...

0.1197834

normalized\_root\_mea...

0.1703896

normalized\_root\_mea...

0.1629519

r2\_score

0.4716208

root\_mean\_squared\_er...

54.69506

root\_mean\_squared\_lo...

0.4222456

spearman\_correlation

0.6774981

predicted\_true

Predicted vs. True

Average Predicted Value

Ideal

Predicted Value

True Value

residuals

Residuals Histogram

Bin Count

Residuals

25 - 39.41

39.41 - 95.4

95.4 - 151.39

151.39 - 207.38

207.38 - 263.37

263.37 - 319.36

319.36 - 346

-192.52 - -154.02

-154.02 - -115.52

-115.52 - -77.01

-77.01 - -38.5

-38.5 - 0

0 - 38.5

38.5 - 77.01

77.01 - 115.52

115.52 - 154.02

154.02 - 192.52

Microsoft Azure Machine Learning Studio

Search

This workspace

1

Microsoft Azure Sponsorship

amldev

PP

...

>

amldev

>

Jobs

>

diabetes\_autoML

>

careful\_deer\_lyrygkdd

>

bright\_feast\_5xyy7qk

bright\_feast\_5xyy7qk

Completed

Overview

Model

Explanations (preview)

Metrics

Data transformation (preview)

Test results (preview)

Outputs + logs

Images

Child jobs

Code

Monitoring

Refresh

Deploy

Download

Explain model

View generated code

Test model (preview)

Register model

Cancel

Delete

Explanation ID

>>

f73061d6

c423cac2

Metrics Difference Warning: If the dataset is larger than 5k rows, metrics shown in the Model performance tab of the explanation dashboard are computed on subsampled data, which may be different than the model metrics on the full ...

View previous dashboard experience

Explainer: mimim.lightgbm

+ New cohort

Edit cohort

DATA STATISTICS

Regressor

442 datapoints

10 features

DATASET COHORTS

All data

442 datapoints

0 filters

Model performance

Dataset explorer

Aggregate feature importance

Individual feature importance

Explore the top-k important features that impact your overall model predictions (a.k.a. global explanation). Use the slider to show descending feature importances. All cohorts' feature importances are shown side by side and can be toggled off by selecting the cohort in the legend. Click on any of the features in the graph to see a density plot below of how values of the selected feature affect prediction.

Top 4 features by their importance

Aggregate feature importance

BMI

SS

BP

SEX

Sort by cohort

All data

Chart type

Bar

Box

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Linked Services

Data Labeling



← Répertoire par défaut



New



Home



Authoring



Notebooks



Automated ML



Designer



Assets



Data



Jobs



Components



Pipelines



Environments



Models



Endpoints



Manage



Compute



Linked Services



Data Labeling

# Register model from a job output



Select job



Select output



Model settings



Review

## Select output

Specify the corresponding output to register the model.



A named model output has been detected in the job outputs and auto-selected. You can select any other output and/or type if this is not what you intend to select.

Model type \* ⓘ

MLflow

Job output \* ⓘ

mlflow\_log\_model\_1365782944 (azureml\_AutoML\_fd15433d-3ef3-4bc5-aa01...



Back

Next

Cancel