

People, Process, Platform les 3 Piliers du MLOps

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



Tout les oppose mais ils vont devoir collaborer...



Son objectif: obtenir le meilleur modèle



Alibek
Data Scientist

- prépare les données
- expérimente dans un notebook
- trace les résultats des différents essais

Son objectif: garantir les performances du service en production



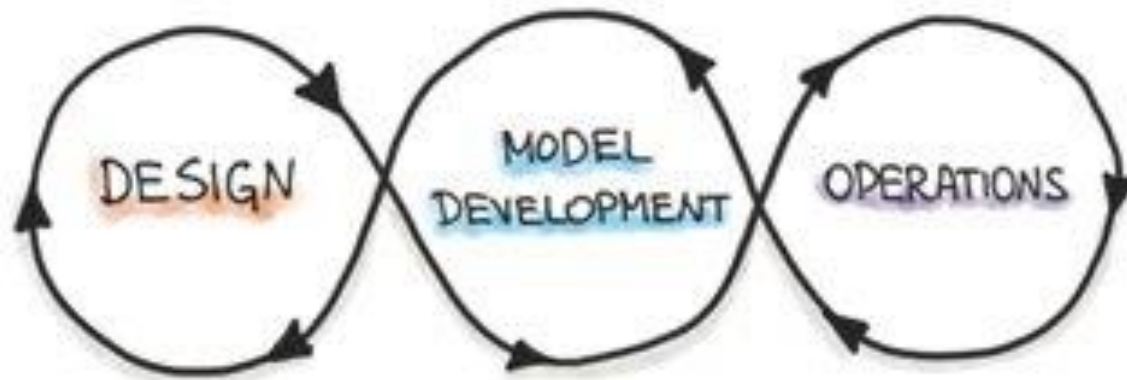
Paul
ML Engineer

- intègre les livrables
- déploie les nouvelles versions
- orchestre et surveille les processus

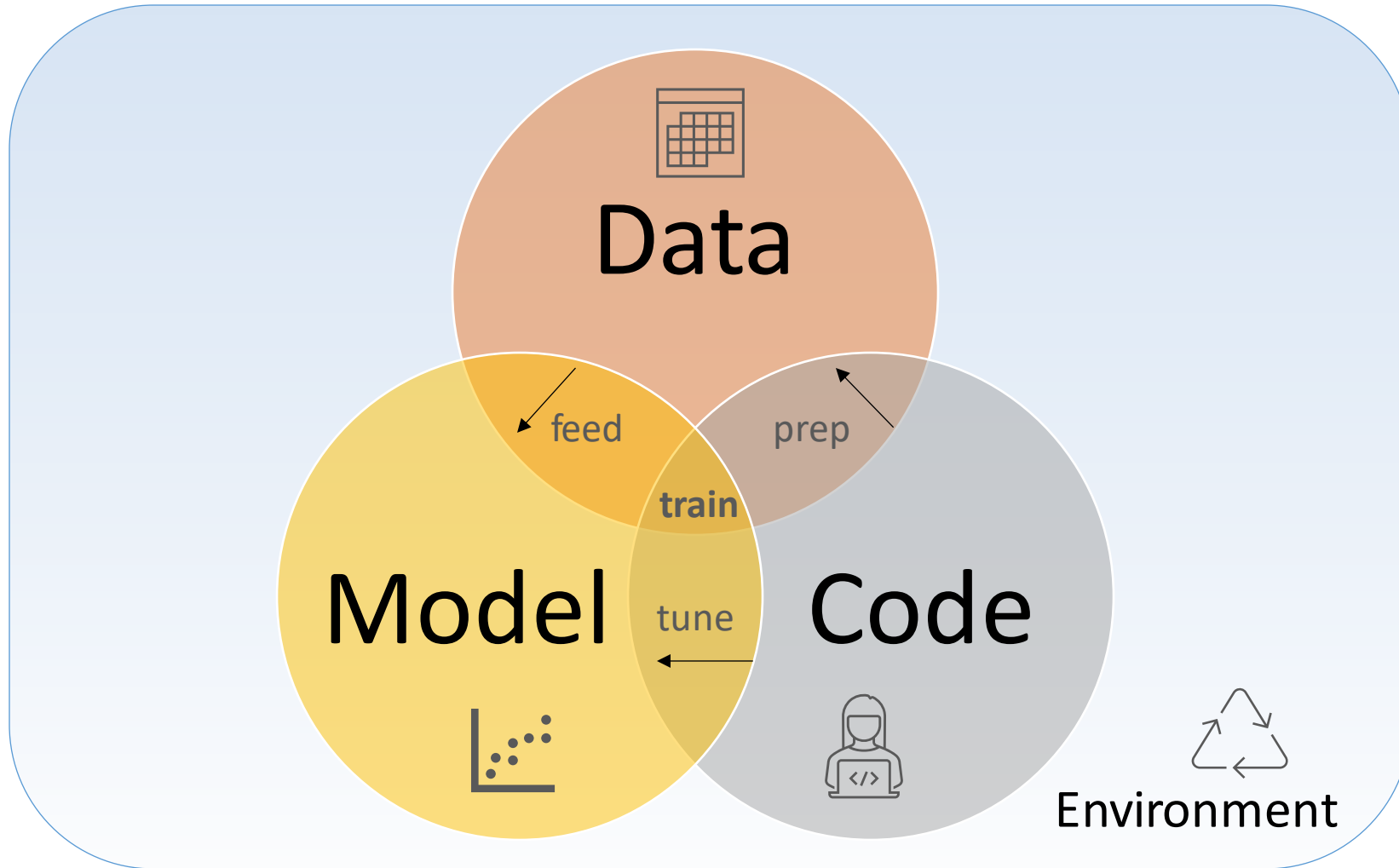


MLOps : une première définition

Fournir un processus de développement du Machine Learning de **bout en bout** pour **concevoir, créer et gérer** des **logiciels reproductibles, testables et évolutifs**.

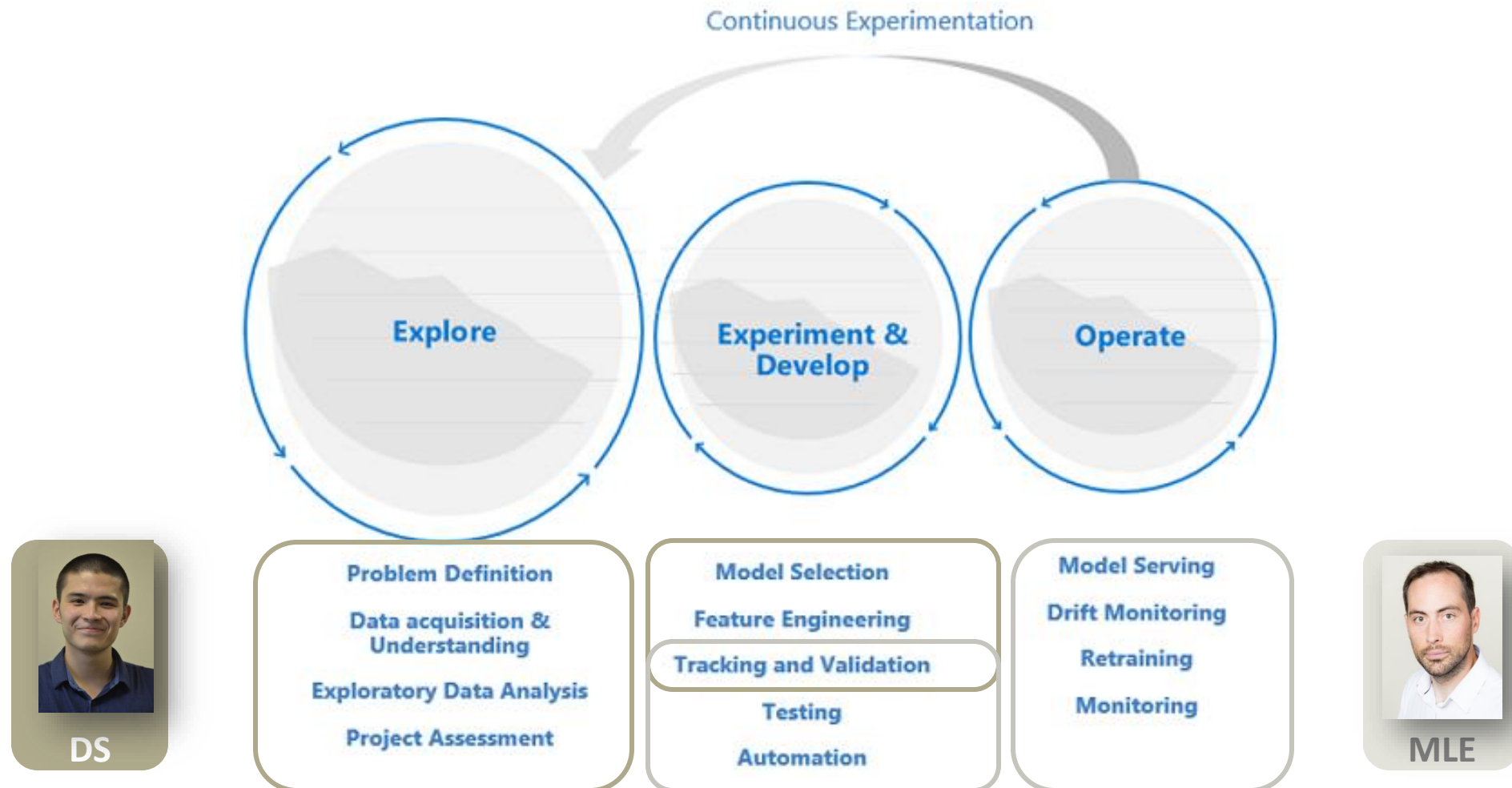


Le triptyque à mettre sous contrôle



L'environnement permet l'exécution du code (OS, Python, librairies...).

Les trois boucles collaboratives du MLOps

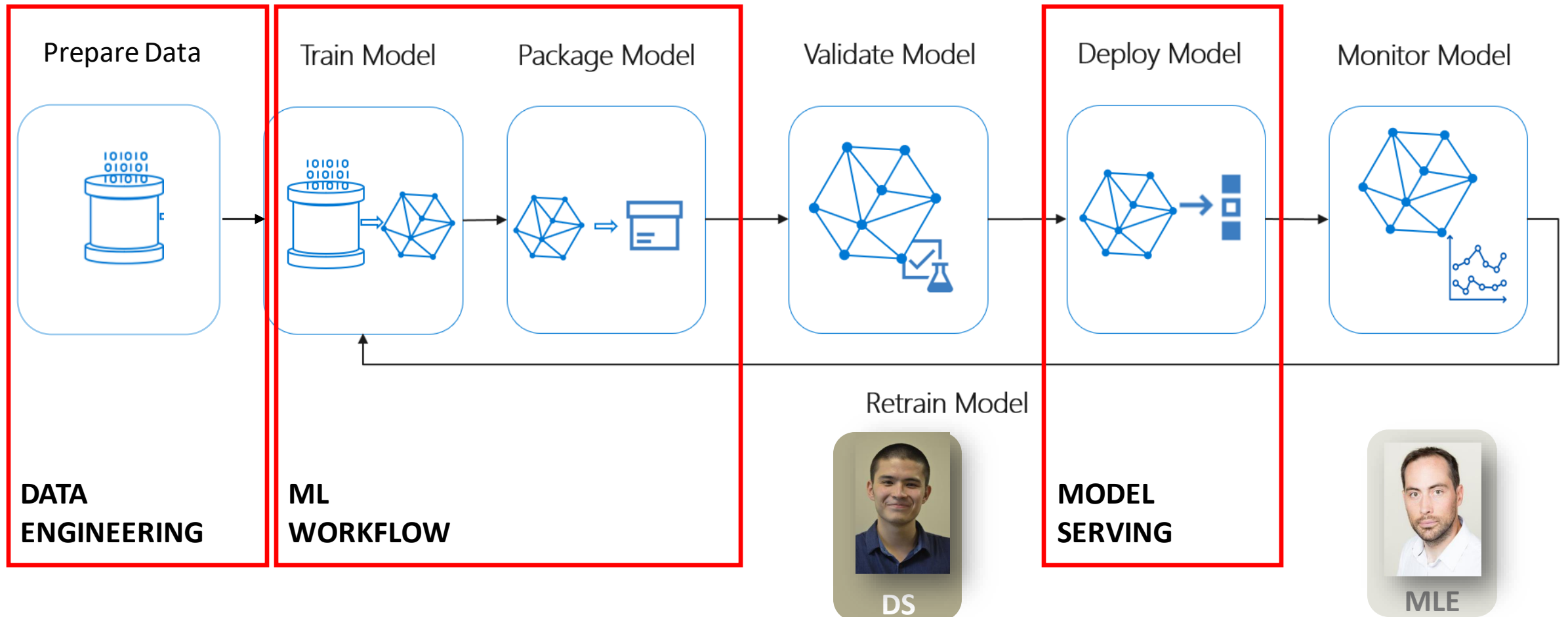


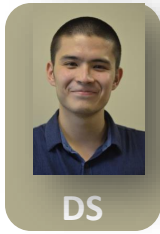
A person wearing a blue suit is riding a green bicycle on a city street. The background shows a stone wall with arched windows and a parked motorcycle. The image has a dark, semi-transparent overlay.

Chaine

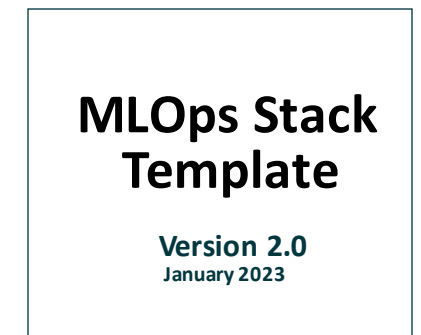
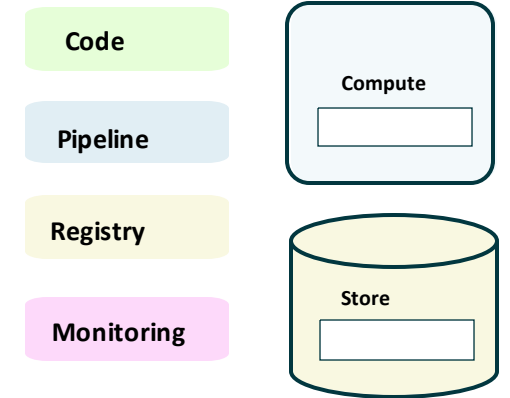
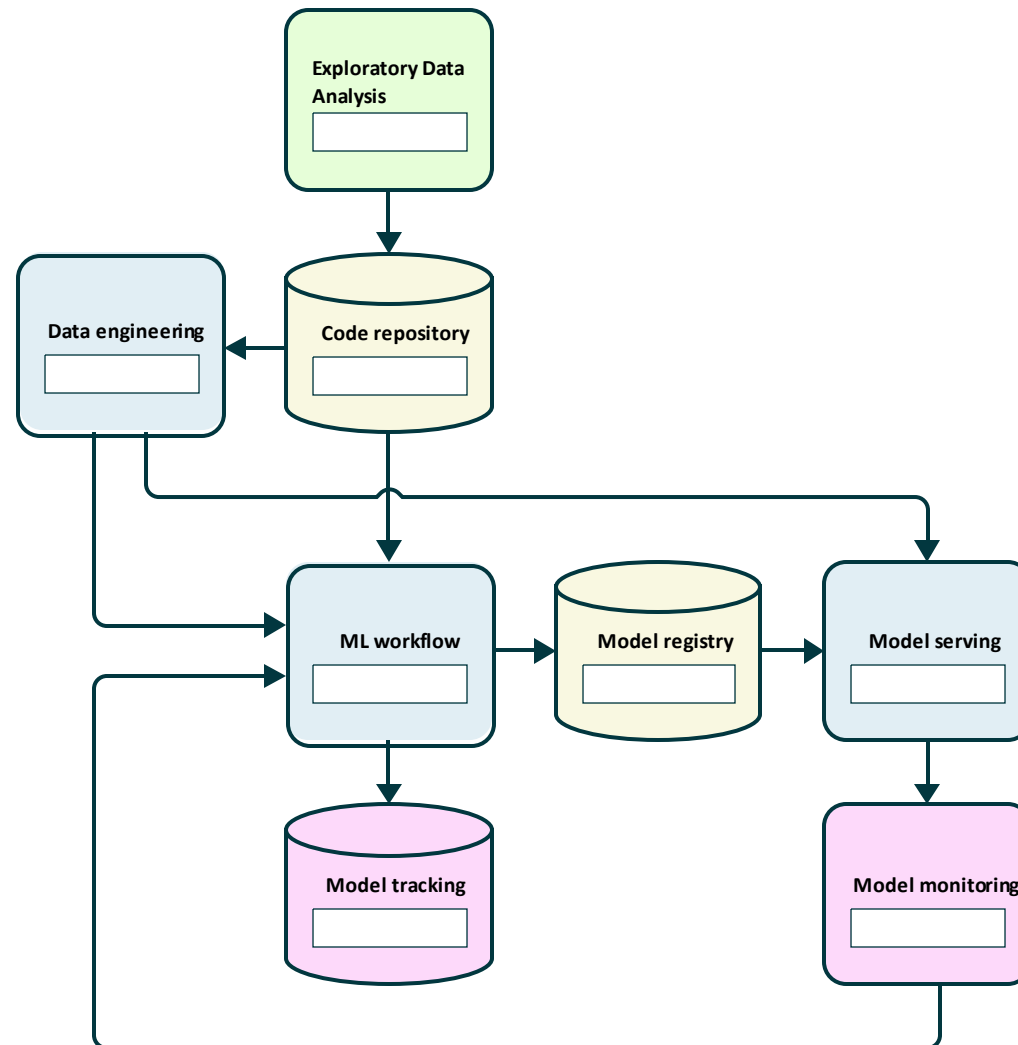
Définir les process, se répartir les rôles

Focus sur les trois pipelines principaux





MLOps end to end process



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



Expérimenter

Utiliser des notebooks ?

- Flux non linéaire
- Résultats ne sont pas itérables

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'
```



Alibek
Data Scientist

Diabetes Ridge Regression Parameter Experimentation.ipynb

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

Open in Notebook Editor

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

Améliorer le code

Reproductibilité

Contrôle de la
version

Déboguer

Test et
Réutilisation

CI/CD



Alibek
Data Scientist

Passer d'un notebook à un script .py

- Les composants d'un flux de travail structuré pour les projets de ML



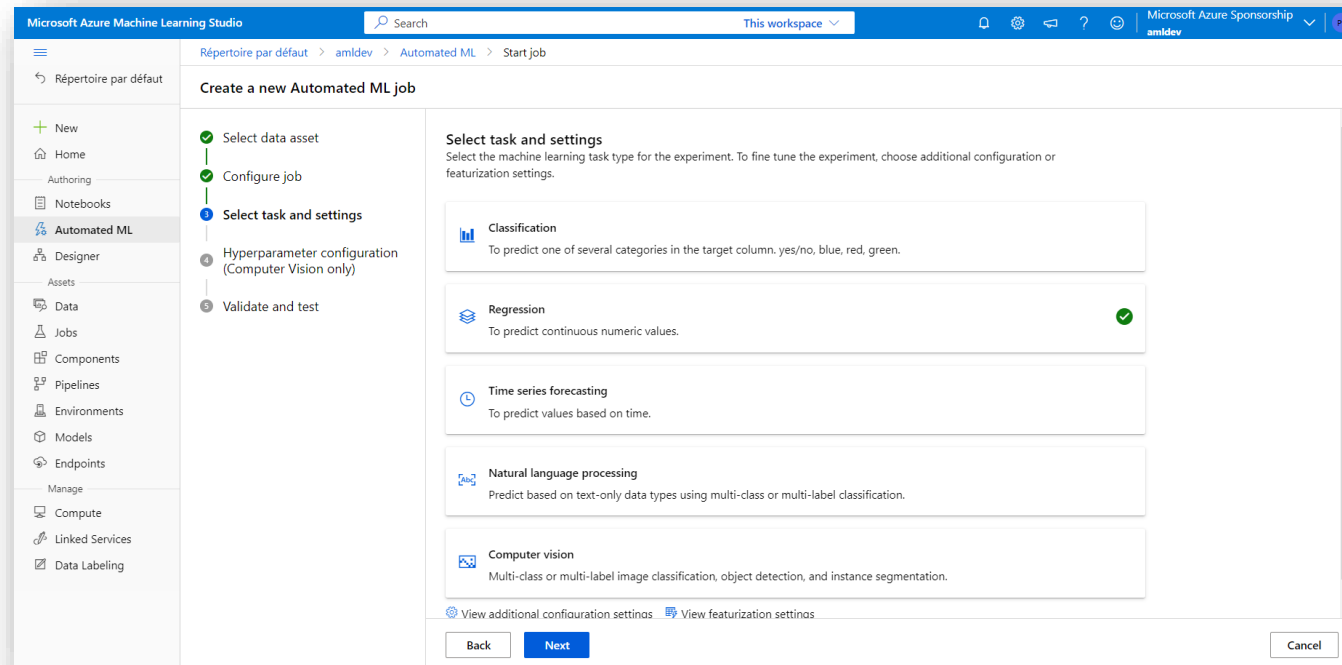
Alibek
Data Scientist

The screenshot displays the Azure DevOps web interface for a project named 'MyMLOpsPython'. The left sidebar shows the navigation menu with 'Files' selected. The main pane shows the file explorer for the 'diabetes_regression' directory, with 'train.py' highlighted. The right pane shows the code editor for 'train.py', which contains Python code for training a Ridge model and evaluating its metrics.

```
24 POSSIBILITY OF SUCH DAMAGE.  
25 """  
26  
27 import os  
28 import pandas as pd  
29 from sklearn.linear_model import Ridge  
30 from sklearn.metrics import mean_squared_error  
31 from sklearn.model_selection import train_test_split  
32  
33  
34 # Split the dataframe into test and train data  
35 def split_data(df):  
36     X = df.drop("y", axis=1).values  
37     y = df["y"].values  
38  
39     X_train, X_test, y_train, y_test = train_test_split(  
40         X, y, test_size=0.2, random_state=0  
41     )  
42     data = {"train": (X_train, y_train),  
43           "test": (X_test, y_test)}  
44     return data  
45  
46 # Train the model, return the model  
47 def train_model(data, ridge_args):  
48     reg_model = Ridge(**ridge_args)  
49     reg_model.fit(data["train"][0], data["train"][1])  
50     return reg_model  
51  
52  
53 # Evaluate the metrics for the model  
54 def get_model_metrics(model, data):  
55     preds = model.predict(data["test"][0])  
56     mse = mean_squared_error(preds, data["test"][1])  
57     metrics = {"mse": mse}  
58     return metrics  
59  
60  
61 def main():  
62     print("Running train.py")
```

Faire de l'experiment tracking

- Utiliser l'**automated ML** pour accélérer les recherches



- Suivre les performances des différents entraînements



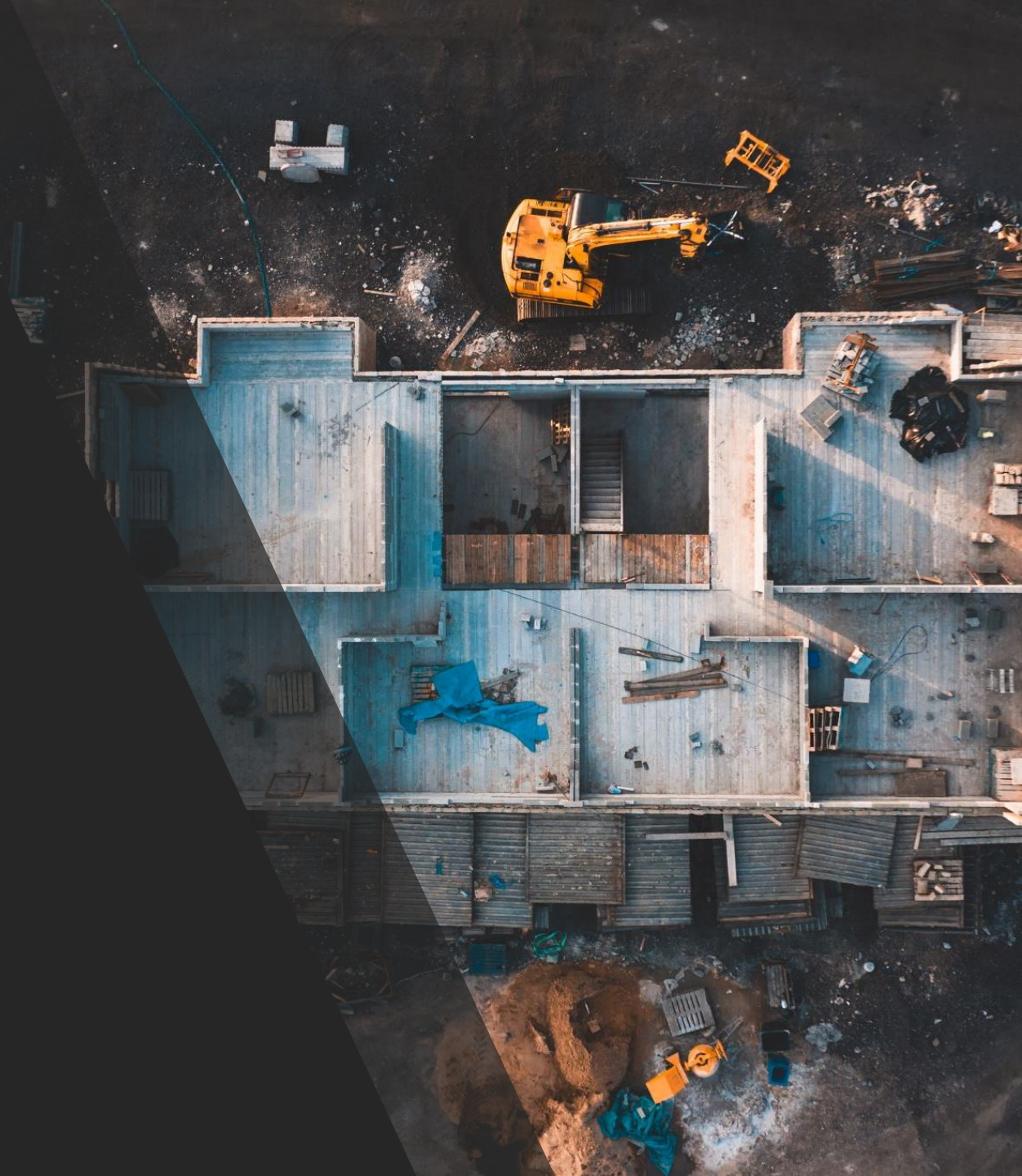
Alibek
Data Scientist

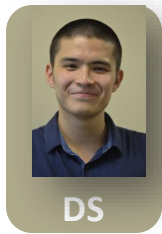


Démonstration

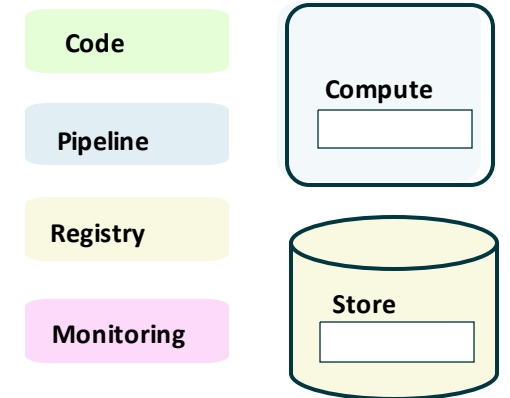
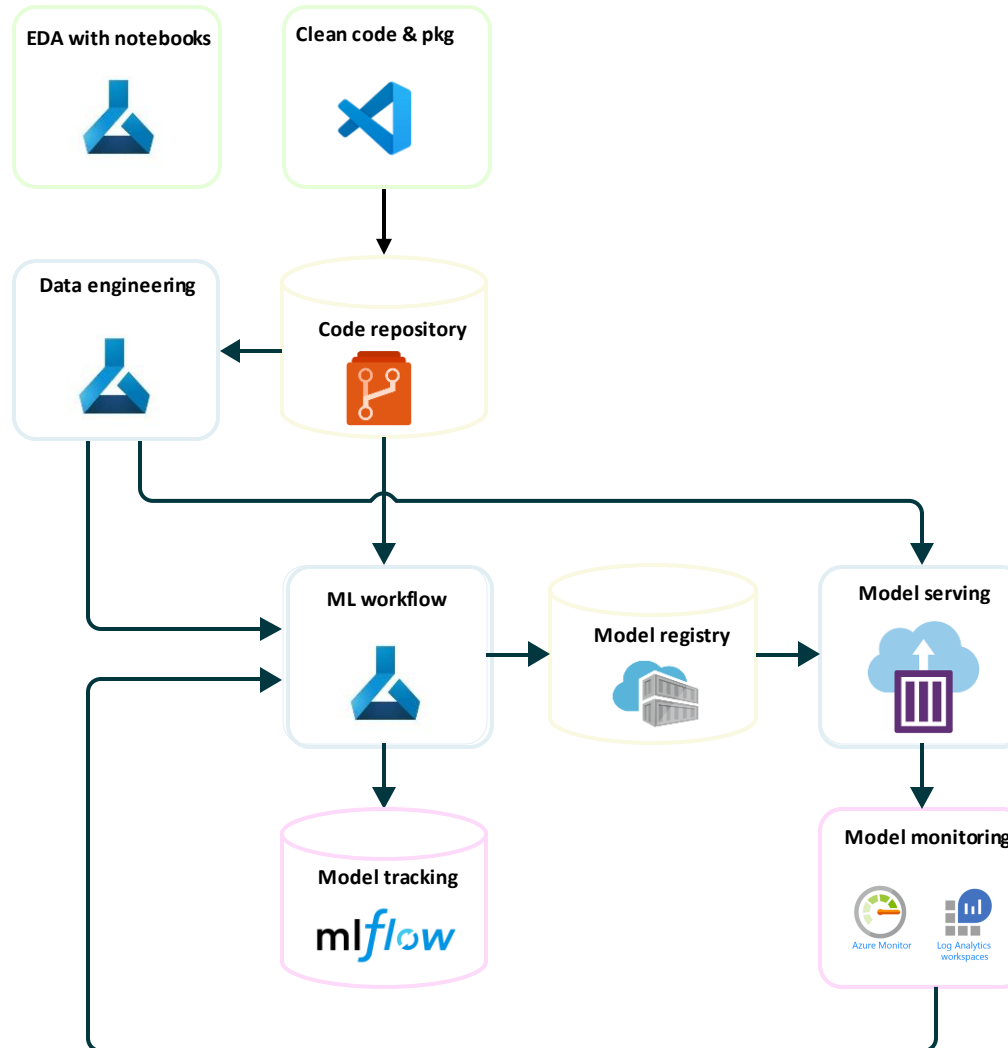
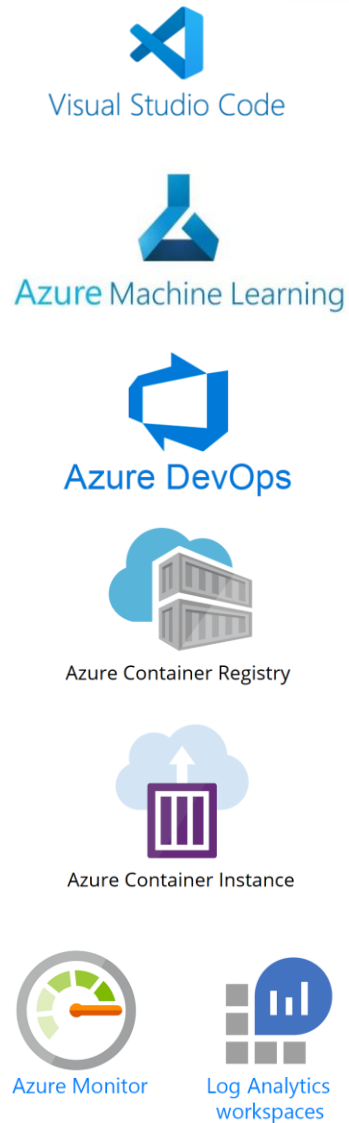
Améliorer l'utilisation des notebooks

Construire la plateforme





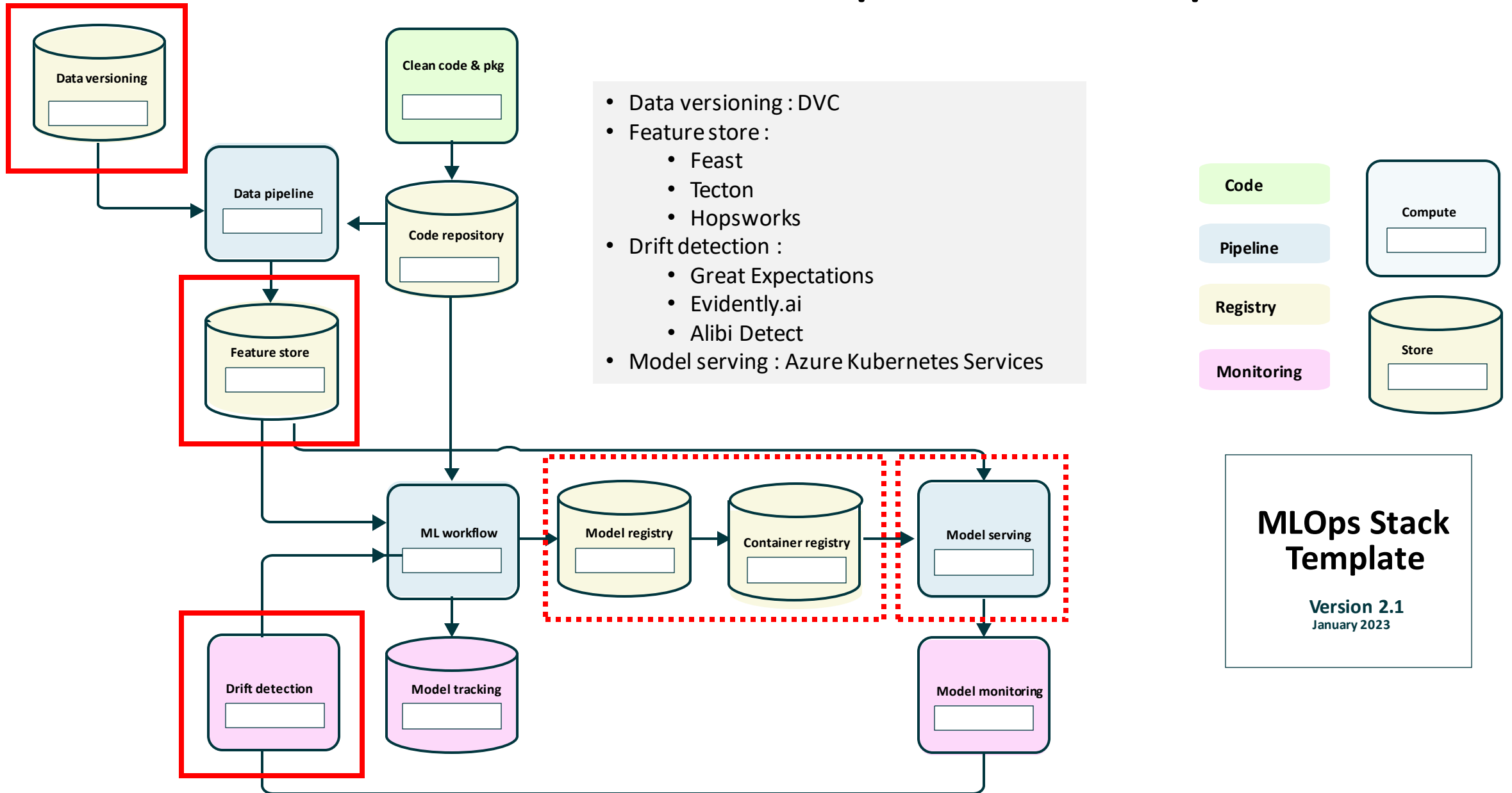
Composants de la stack technique



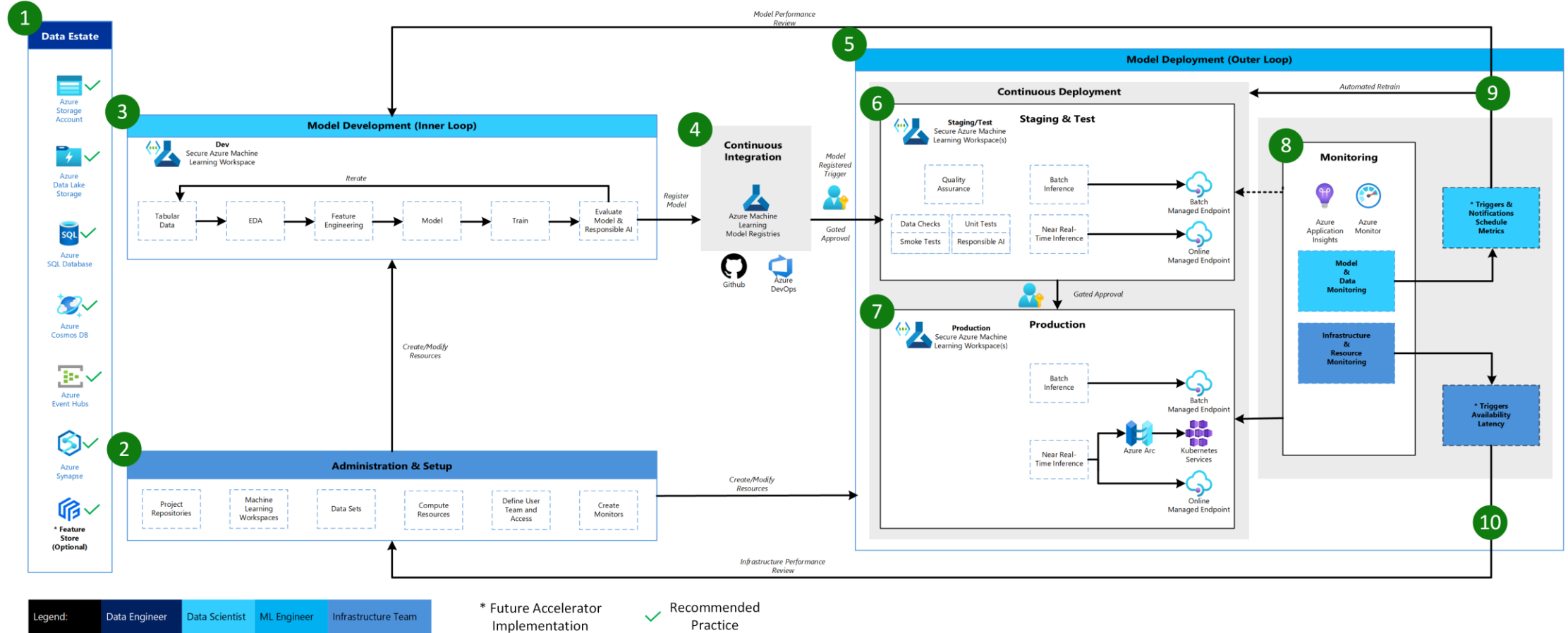
MLOps Stack Template
Version 2.0
January 2023

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

Version avancée du MLOps stack template



MLOps architecture for classical ML on Azure



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



Démonstration

Industrialisation des déploiements



Alibek
Data Scientist

Conclusion : *better together !*



Paul
ML Engineer

Partager une culture du **Software Engineering** (clean code, tests) et du DevOps (automatisation)

Avancer par **palier de maturité** (registres, automatisation, monitoring)

La plateforme sera **exponentiellement efficace** au fur et à mesure que de nouveaux cas d'usage y seront implémentés.

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 * i

Y (Decimal)

Select compute type

Compute cluster

Select Azure ML compute cluster *

aml-cluster

+ New

Refresh computes

Back

Next

Cancel

Microsoft Azure Machine Learning Studio

Search

This workspace

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 > Automated ML > Start job

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)

Column name	Feature type	Impute with	Data example
AGE	Auto	Auto	0.0380759064334241, -0.0
SEX	Auto	Auto	0.0506801187398187, -0.0
BMI	Numeric	Auto	0.0616962065186885, -0.0
BP	DateTime	Auto	0.0218723549949558, -0.0
S1	Categorical	Auto	0.0218723549949558, -0.0
	Categorical hash	Auto	0.0218723549949558, -0.0
S2	Text	Auto	-0.0442234984244464, -0.0
S3	Auto	Auto	-0.0348207628376986, -0.0
S4	Auto	Auto	-0.0434008456520269, 0.0
S5	Auto	Auto	-0.00259226199818282, -0.0
S6	Auto	Auto	0.0199084208763183, -0.0






Save

Cancel



Microsoft Azure Machine Learning Studio

Search

This workspace



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 > 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 > Automated ML > diabetes_autoML > careful_deer_lyrygkdd

careful_deer_lyrygkdd

Running

Overview

Data guardrails

Models

Outputs + logs

Child jobs

Refresh

Edit and submit (preview)

Register model

Cancel

Delete

Compare (preview)

Properties

Status

Running

Running featurization

Created on

Feb 26, 2023 9:12 PM

Start time

Feb 26, 2023 9:13 PM

Compute target

aml-cluster

Name

AutoML_fd15433d-3ef3-4bc5-aa01-c677a9bf8662

Script name

--

Created by

Paul PETON

Job type

Automated ML

Experiment

diabetes_autoML

Arguments

None

See all properties

Raw JSON

See YAML job definition

Job YAML

Tags

No tags

Description

Click edit icon to add a description

Inputs

Input name: training_data
Dataset: diabetes:1

Best model summary

No data

Run summary

Task type

Regression

Featurization

Enabled

Primary metric

Normalized root mean squared error

Experiment name

diabetes_autoML

View configuration settings

View featurization settings

Microsoft Azure Machine Learning Studio

Search

This workspace

1

Microsoft Azure Sponsorship
amldev

pp

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

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

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

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.

data_source - 10 col

10 col

Numeric

MeanImputer

1 col

MaxAbsScaler

ElasticNet

New

Home

Authoring

Notebooks

Automated ML

Designer

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Linked Services

Data Labeling

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

🔄 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

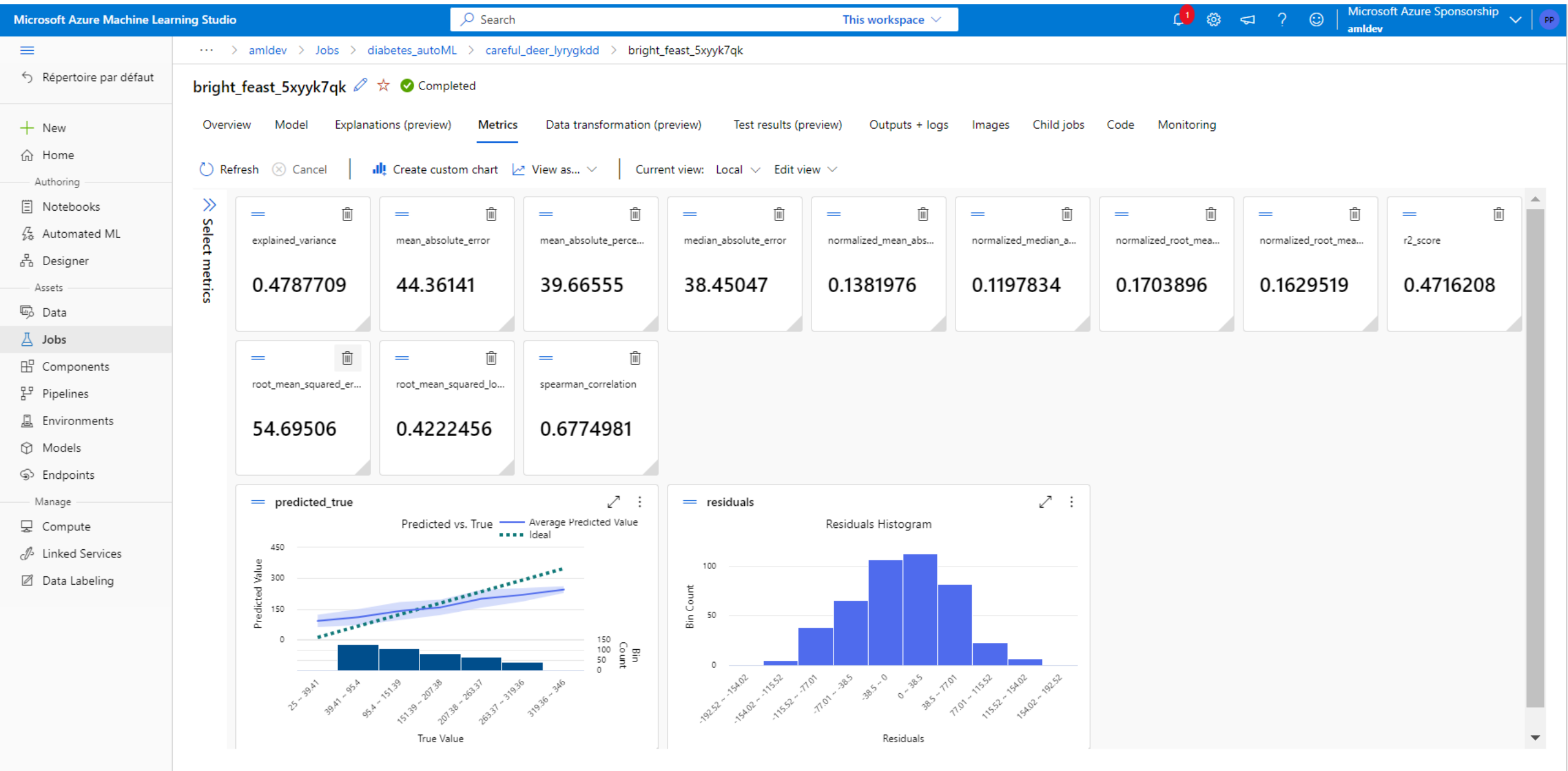
Data transformation:

```
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_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: mimimic.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

Microsoft Azure Machine Learning Studio

Search

This workspace

1

Microsoft Azure Sponsorship
amldev

PP

← 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

2 Select output

3 Model settings

4 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