



GPS Data/AI Strategy FY23

Delivered by CSA Team



Franck Gaillard
Cloud Solution Architect
Data AI
frgail@microsoft.com



Narjes Majdoub
Cloud Solution Architect
Data AI
nmajdoub@microsoft.com



Ali Bouhaddou
Cloud Solution Architect
Data Analytics
albouhad@microsoft.com



Frederic Gisbert
Cloud Solution Architect
Data Analytics
frgisber@microsoft.com

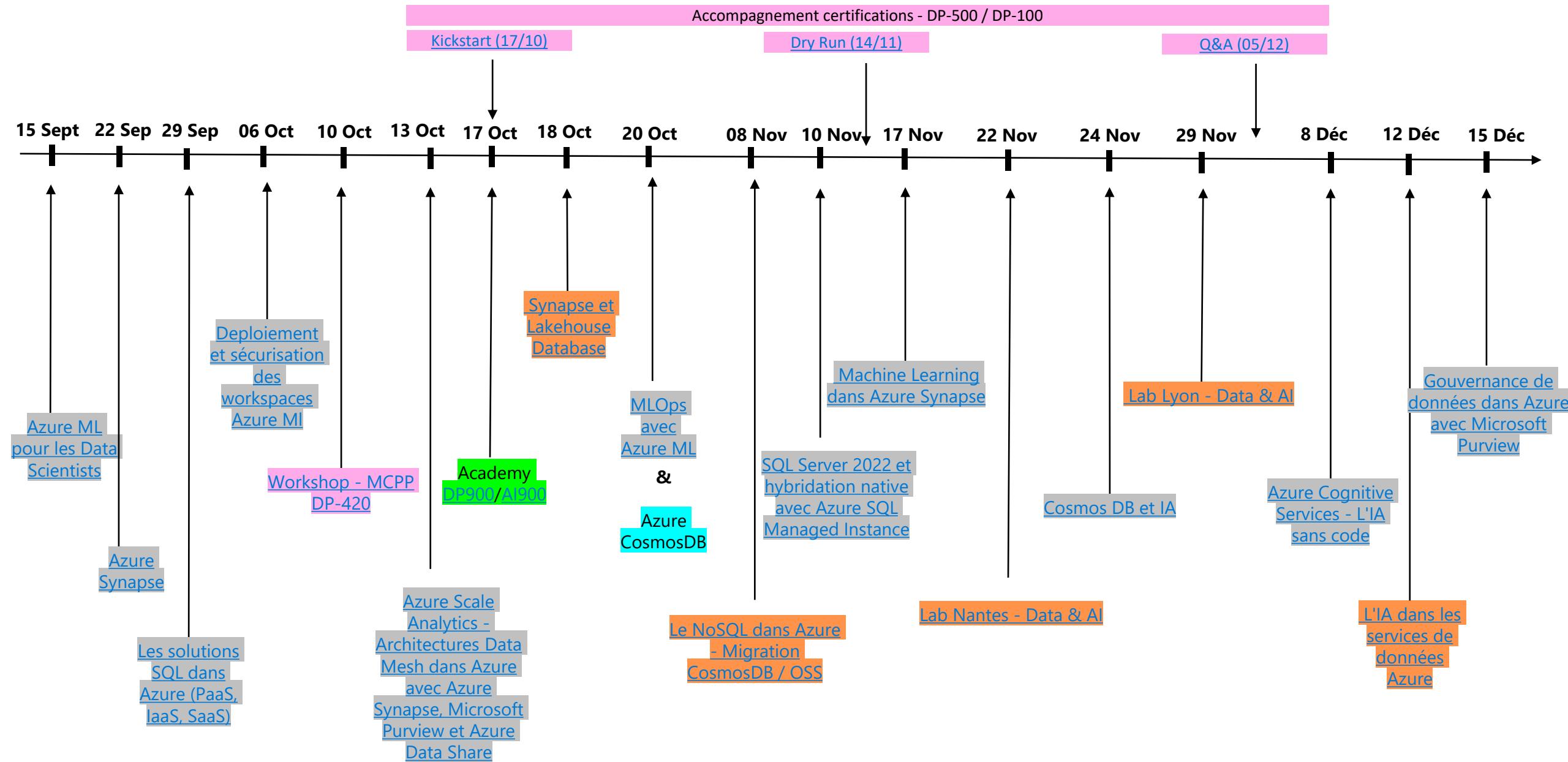


Azure Data & AI technical intensity plan

- From June 2022 to June 2023
- Focus on "Azure Data & AI" tech intensity
- Many content, from L100 Beginner to L400 Expert level:
 - Academy L100
 - Webinar L200/L300
 - Workshop L300/L400
 - Certification kickstart L300/L400
 - Openhack / Microhack L400

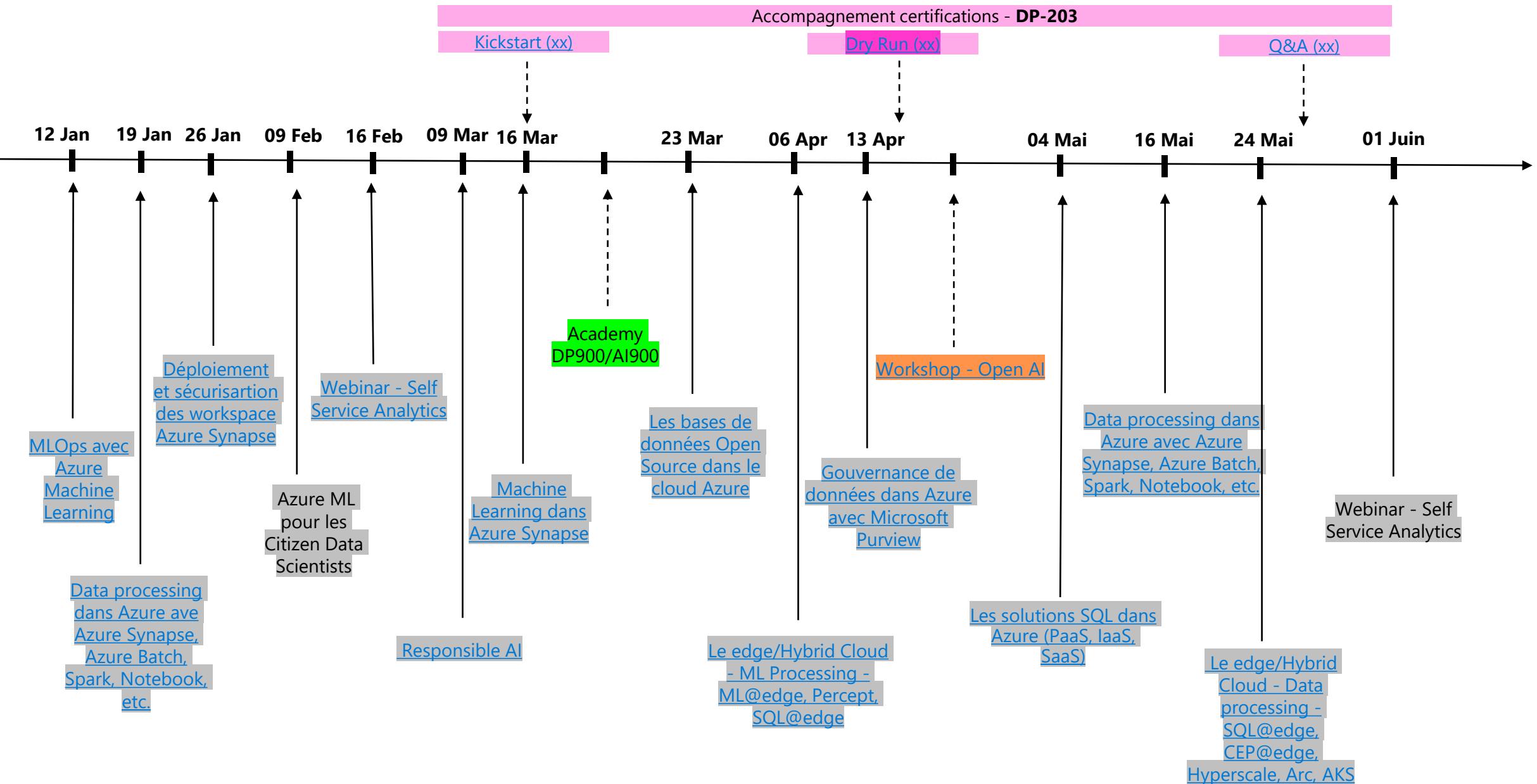
Data & AI events timeline – H1

Webinar/Academy - L 200/300
Workshop/ Openhack/ Certifications - L 300/400



Data & AI events timeline – H2

Webinar/Academy - L 200/300
Workshop/ Openhack/ Certifications - L 300/400



Liste des évènements de type Webinar 2H

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Machine Learning pour les Data Scientists	15/09/2022	120	https://msevents.microsoft.com/event?id=2454281594
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Déploiement et sécurisation des workspace Azure Synapse	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Azure Machine Learning pour les Citizen Data Scientists	09/02/2023	120	En cours
PowerBI - Self Service Analytics	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Hybridation des services de données Azure	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours

Liste des évènements de type Workshop/Prepa Cert/Academy

Event Workshop L300/400	Date	Duration (min)	Link
Synapse et Lakehouse Database	18/10/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Le NoSQL dans Azure - Migration CosmosDB / OSS	08/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Lab Lyon - Data & AI	22/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMIZZOURET0RSWjcyTERYRkJGTIFFUJaUi4u
Lab Nantes - Data & AI	29/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMIZZOURET0RSWjcyTERYRkJGTIFFUJaUi4u
L'IA dans les services de données Azure	12/12/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Open AI	H2	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u

Event Academy, kickstart certifications, workshop certifications	Date	Duration (min)	Link
MCPP - DP-420	10/10/2022	420	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMkJSIRKSU1RRFA0OVgzSFdTSTY0RE9WQy4u
Micro Hack CosmosDB	20/10/2022	420	H1 - Inscriptions PTA
Academy DP900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=3250818161
Academy AI900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=2717528090
Kickstart DP-500	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEDNNTQ2Uk85V0cxQzM3TE9ZRS4u
Dry Run DP-500	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEDNNTQ2Uk85V0cxQzM3TE9ZRS4u
Q&A DP-500	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEDNNTQ2Uk85V0cxQzM3TE9ZRS4u
Kickstart DP-100	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYxSkRIMi4u
Dry Run DP-100	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYxSkRIMi4u
Q&A DP-100	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYxSkRIMi4u
Kickstart DP-203	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Dry Run DP-203	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Q&A DP-203	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u



Azure ML for Citizen Data Scientists

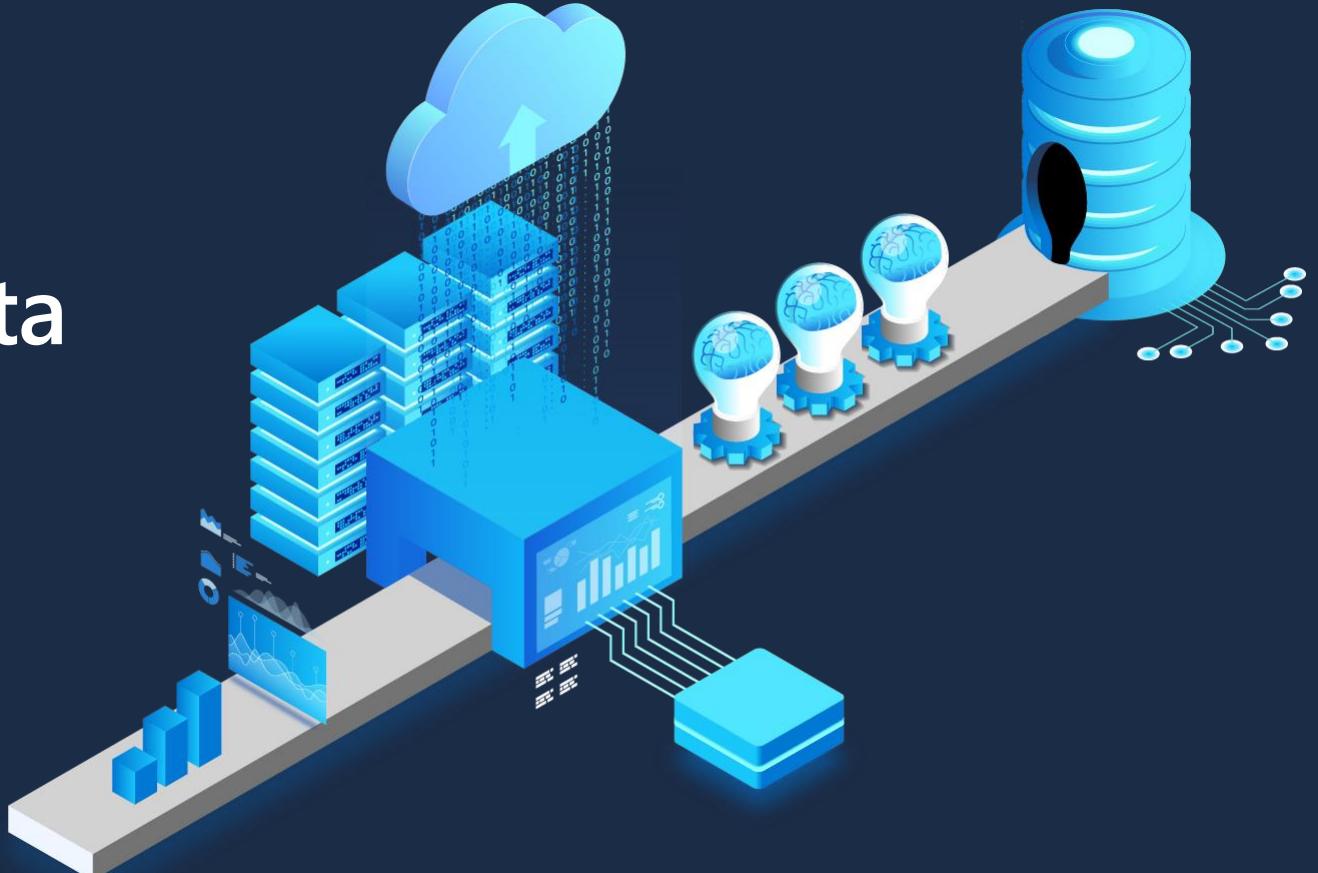
09/02/2023



Franck Gaillard
Cloud Solution Architect
Data AI
frgail@microsoft.com



Narjes Majdoub
Cloud Solution Architect
Data AI
nmajdoub@microsoft.com



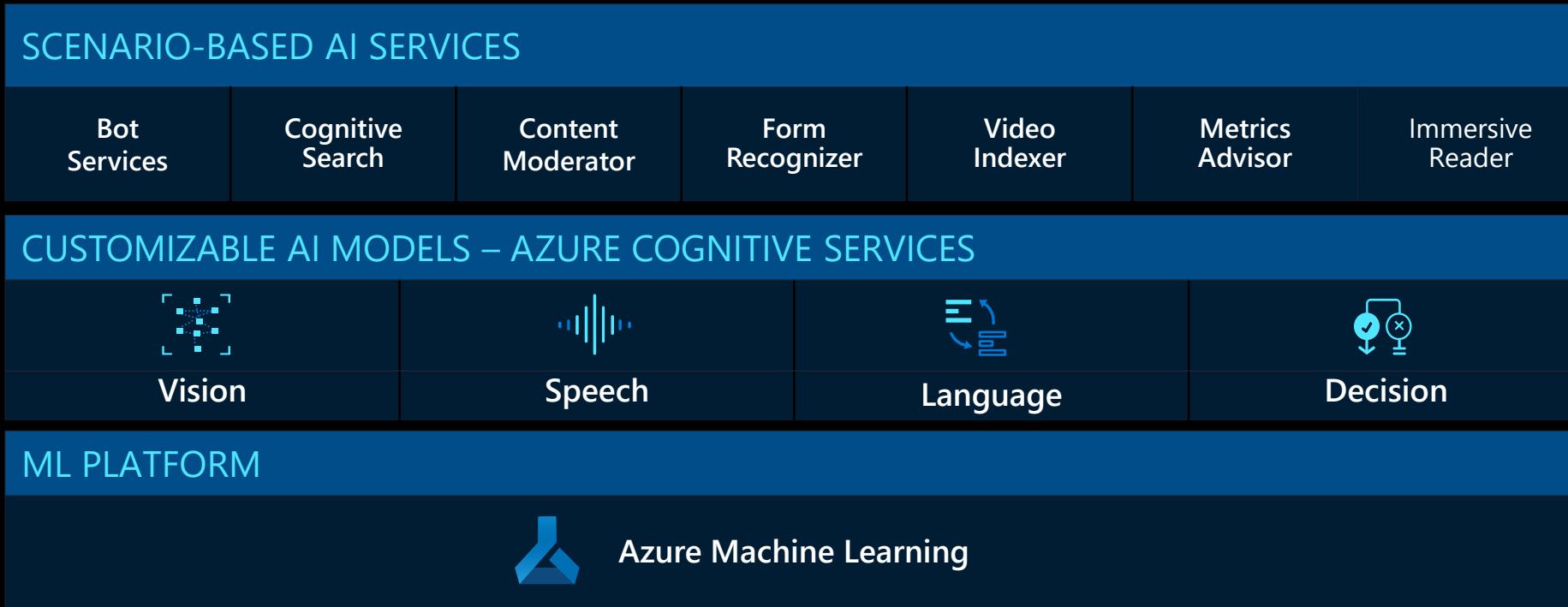
Agenda

-
1. Introduction to Azure ML
 2. AML Designer
 3. Data labeling with AML
 4. Automated ML

Introduction to Azure ML

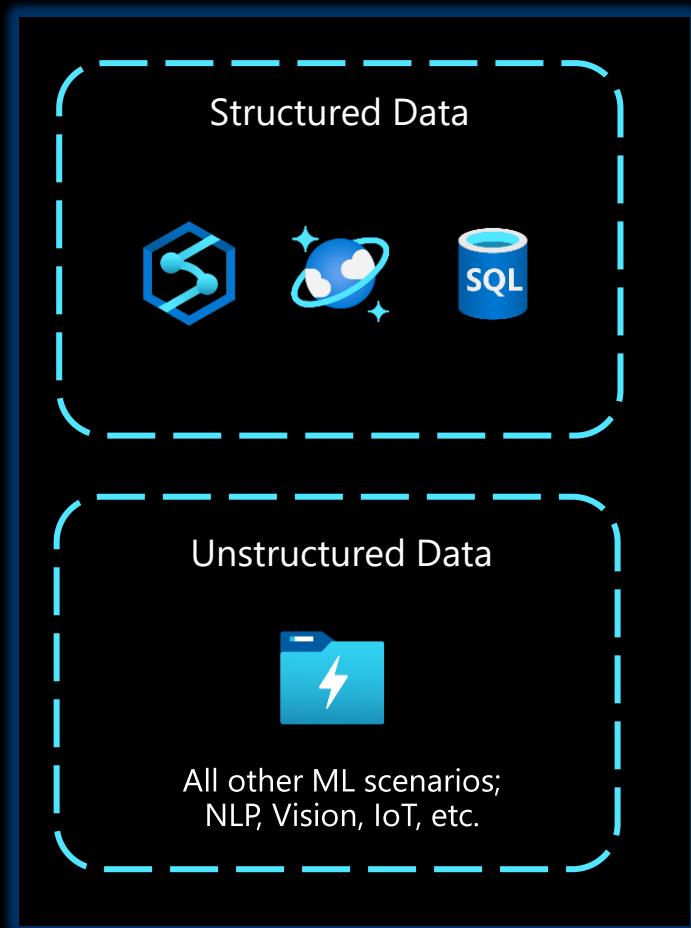


Azure AI

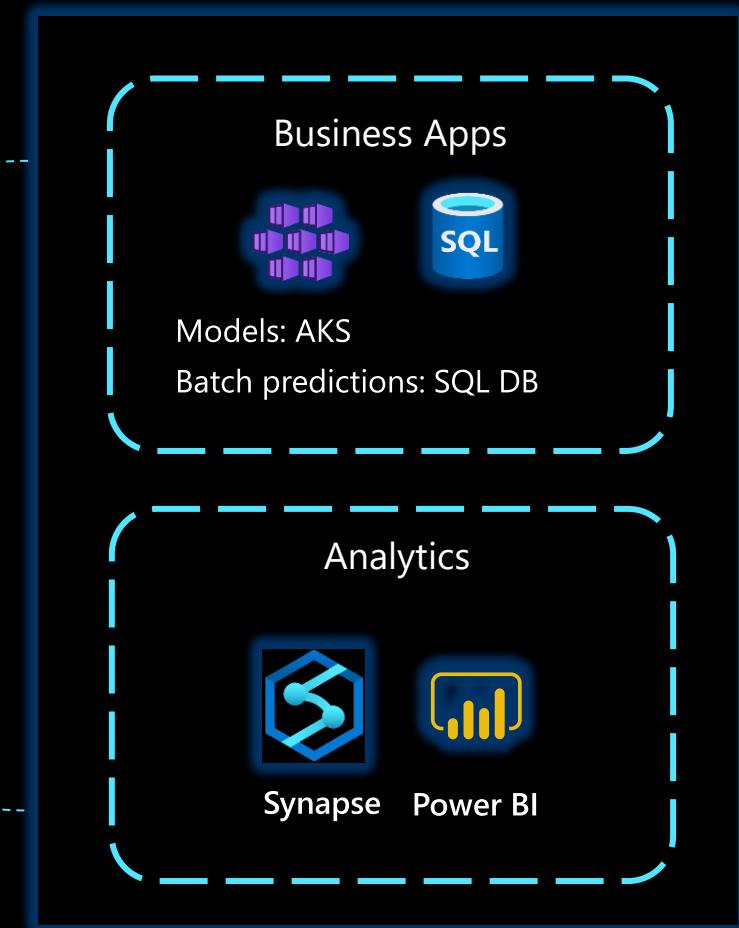


Machine Learning Platform Architecture

Ingest Data ►



Visualize ►



Prepare Data

Build & Train

Deploy

Manage & Monitor

Machine Learning Platform Architecture



Azure ML

Structured Data



Unstructured Data



All other ML scenarios;
NLP, Vision, IoT, etc.

Seamless studio experience

Notebooks

Automated ML

Designer

Responsible ML tools

Comprehensive MLOps

Reproducibility

Automation

Deployment

Re-training

Governance

Unified management **across clouds and on-premises**



Azure

Serverless
Compute

Managed
Kubernetes



Edge & Hybrid

Azure Arc-enabled
Kubernetes

Edge/IoT Devices

Prepare Data

Build & Train

Deploy

Manage & Monitor

Business Apps



Models: AKS

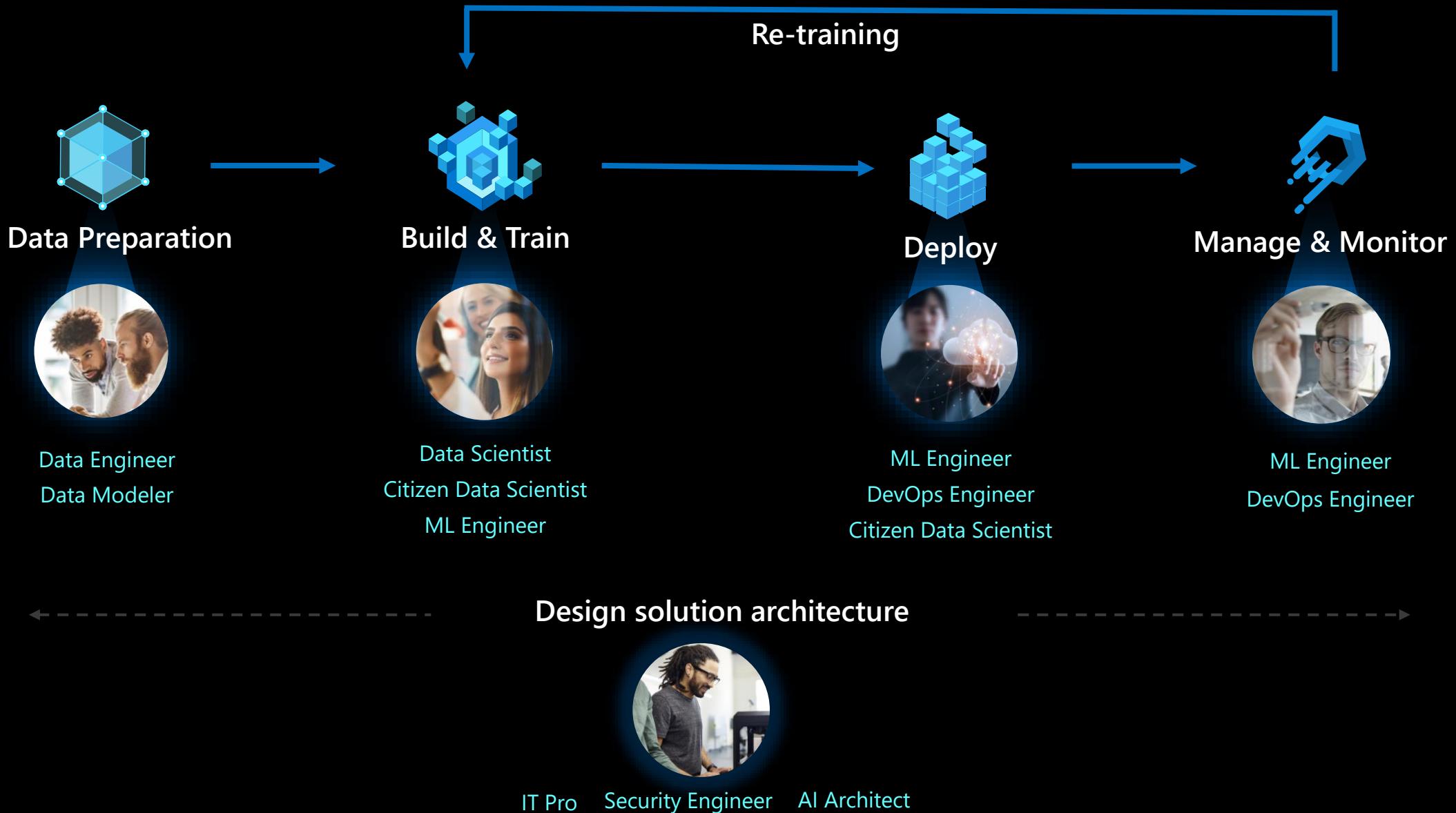
Batch predictions: SQL DB

Analytics



Synapse Power BI

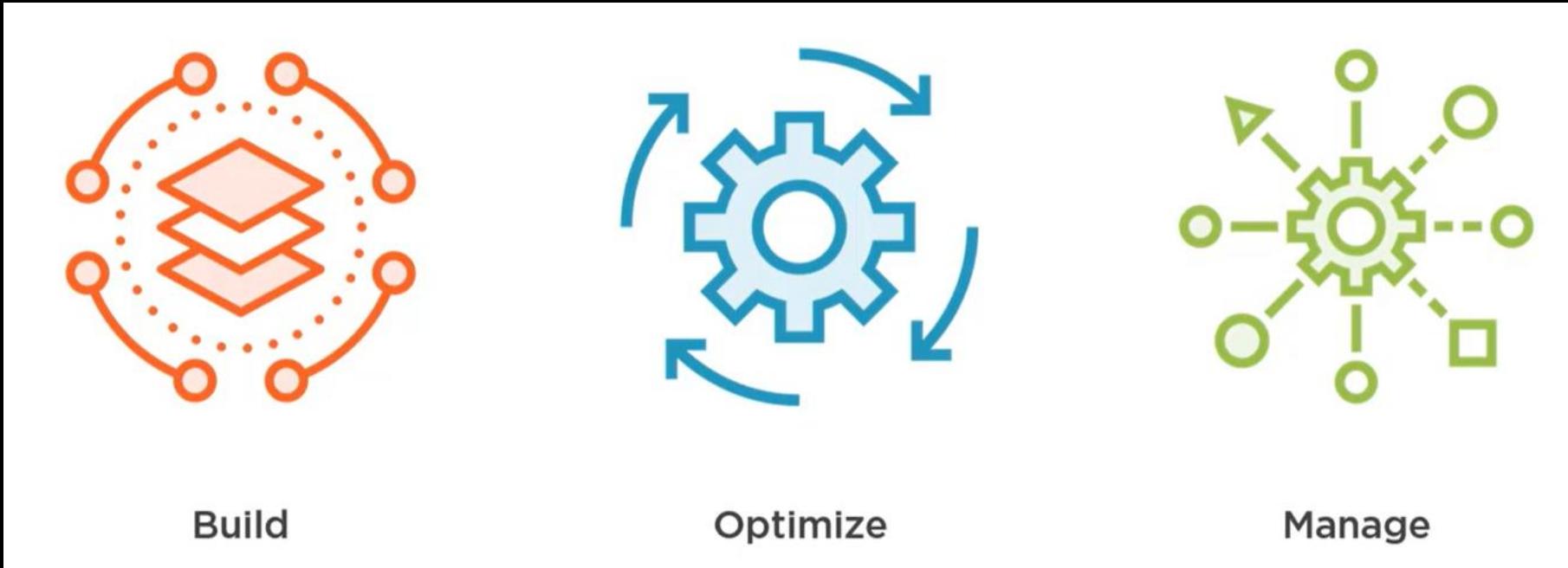
ML Lifecycle workflows



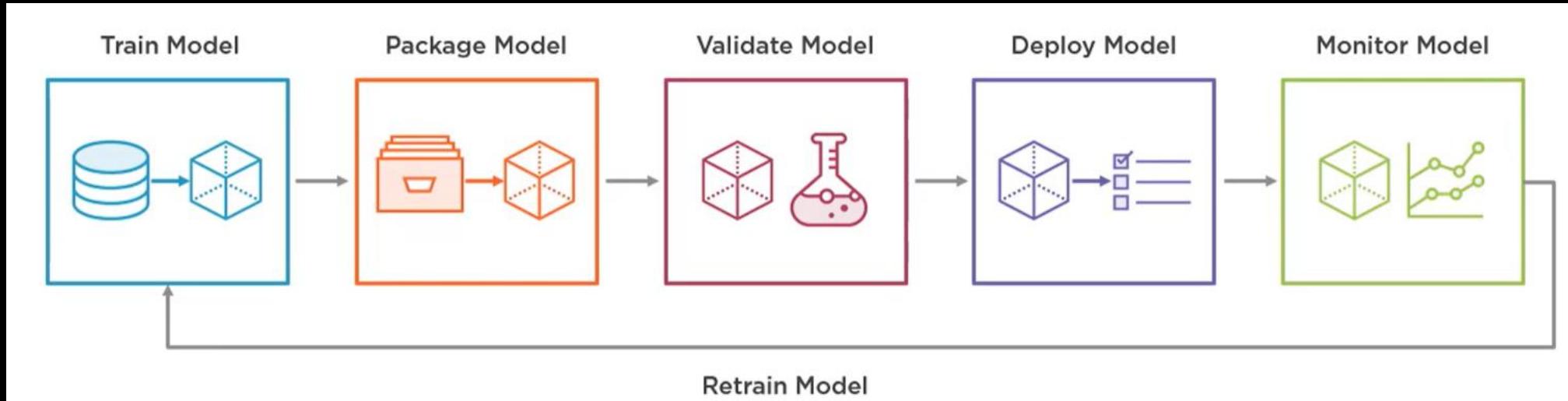
AML Designer



Azure Machine Learning Pipeline



Pipeline Expectation



Components of Azure ML Designer

Pipeline

Dataset

Module

Compute

Deploy

Publish

Azure Machine Learning studio

A new immersive experience for managing the end-to-end machine learning lifecycle

The screenshot shows the Microsoft Azure Machine Learning studio interface. On the left, there's a navigation sidebar with options like New, Home, Author, Notebooks, Automated ML, Designer (which is selected and highlighted with a blue box), Assets, Datasets, Experiments, Pipelines, Models, Endpoints, Manage, Compute, Datastores, and Data Labeling. The main area is titled "Azure Machine Learning studio" and contains four cards: "Create new" (with a plus icon), "Notebooks" (with a document icon), "Automated ML" (with a lightning bolt icon), and "Designer" (with a cluster icon). The "Designer" card is also highlighted with a blue box. Below these cards is a section titled "My recent resources" divided into "Runs" and "Compute". The "Runs" section lists several completed runs, each with a Run ID, Experiment name, Status, Submitted time, Submitted by, and Run type. The "Compute" section lists various compute resources with their names, types, provisioning states, and creation dates.

Run	Run ID	Experiment	Status	Submitted time	Submitted by	Run type
Run 9	daf34cdc-424a-4ed5-871...	sample1...	Completed	Nov 2, 2020 11:09 AM	Lu Zhang (AI)	Pipeline
Run 1	e4378dca-efe4-4300-a1b...	sample1...	Completed	Nov 2, 2020 10:49 AM	Lu Zhang (AI)	Pipeline
Run 34	ddf39624-e181-422e-979...	bookstor...	Completed	Nov 2, 2020 10:38 AM	Lu Zhang (AI)	Pipeline
Run 30	db48334c-c3aa-4ea0-b02...	bookstor...	Failed	Nov 2, 2020 10:26 AM	Lu Zhang (AI)	Pipeline
Run 18	eeaa13be-b8d5-45d8-996...	bookstor...	Completed	Nov 2, 2020 9:09 AM	Lu Zhang (AI)	Pipeline
Run 13	4dbcfecb-d40a-4435-81e...	bookstor...	Completed	Nov 2, 2020 8:55 AM	Lu Zhang (AI)	Pipeline
Run 1	7456337d-037d-42fb-99e...	bookstor...	Completed	Nov 2, 2020 8:33 AM	Lu Zhang (AI)	Pipeline
Run 3	labeling_Inference_64f41a...	labeling_I...	Completed	Oct 29, 2020 12:35 AM	Service Princi...	Script
Run 2	labeling_Validation_64f41...	labeling_I...	Completed	Oct 26, 2020 7:53 AM	Service Princi...	Script
Run 2	labeling_Inference_64f41a...	labeling_I...	Completed	Oct 26, 2020 7:35 AM	Service Princi...	Script

[View all experiments →](#)

Name	Type	Provisioning state	Created on
my-aks-9	Kubernetes service	Succeeded	Aug 14, 2020 8:32 PM
akscpucompute2	Kubernetes service	Succeeded	Aug 14, 2020 2:57 PM
akscpucompute1	Kubernetes service	Succeeded	Aug 14, 2020 1:47 PM
mercycompute	Compute instance	Succeeded	Aug 14, 2020 12:22 AM
cpucluster	Machine Learning com...	Succeeded (0 nodes)	Aug 13, 2020 10:44 PM
akscpucompute	Kubernetes service	Succeeded	Aug 13, 2020 9:50 PM
gpucluster	Machine Learning com...	Succeeded (0 nodes)	Aug 11, 2020 8:31 AM
vkannCompute1	Compute instance	Succeeded	Aug 8, 2020 6:45 AM
gpu-cluster2	Machine Learning com...	Succeeded (1 node)	Jun 5, 2020 3:10 AM
gpu-cluster3	Machine Learning com...	Succeeded (1 node)	Jun 5, 2020 1:47 AM

[View all compute →](#)

<https://ml.azure.com>

Who to use



Citizen Data Scientists

Build model and prefer low-code/no-code experience



Professional Data Scientists

Coding first but need a visual way to create & manage their ML pipelines



ML Engineers

Focus on operationalize a trained model as a training pipeline and deploy it for inferencing



Azure Machine Learning designer

Drag-and-drop workflow to build, test and deploy your ML models more easily and efficiently

Microsoft Azure Machine Learning

AML-Data-Labeling-WS > Designer

Designer

New pipeline

Show less samples ^

Easy-to-use prebuilt modules

Image Classification using DenseNet

Binary Classification using Vowpal Wabbit Model - Adu...

Wide & Deep based Recommendation - Restaur...

Regression - Automobile Price Prediction (Basic)

Regression - Automobile Price Prediction (Compare algorit...

Binary Classification with Feature Selection - Income ...

Binary Classification with custom Python script - Cred...

Binary Classification - Customer Relationship Prediction

Use custom R script - Flight Delay Prediction

Text Classification - Wikipedia SP 500 Dataset

Cross Validation for Binary Classification - Adult Incom...

Permutation Feature Importance

Recommendation - Movie Rating Tweets

Tune Parameters for Binary Classification - Adult Incom...

Multiclass Classification - Letter Recognition

Pipelines

Pipeline drafts Pipeline runs

Refresh Delete

Search to filter items...

Name	Pipeline type	Updated on ↓	Created by
Regression - Automobile Price Predictio...	Real-time inference	Nov 2, 2020 11:10 AM	Lu Zhang (AI)
Regression - Automobile Price Predictio...	Training	Nov 2, 2020 11:05 AM	Lu Zhang (AI)
Copy of Bookstore-Warehouse-Subway ...	Training	Nov 2, 2020 10:38 AM	Lu Zhang (AI)
Bookstore-Warehouse-Subway Image Cl...	Real-time inference	Nov 2, 2020 9:21 AM	Lu Zhang (AI)
Bookstore-Warehouse-Subway Image Cl...	Training	Nov 2, 2020 9:09 AM	Lu Zhang (AI)

Automobile price data (Raw)

Select Columns in Dataset

Clean Missing Data

Linear Regression

Split Data

Search by name, tag

Datasets

Modular pipeline

Modules

▶ Data Input and Output (3)

▶ Data Transformation (19)

▶ Feature Selection (2)

▶ Statistical Forecasting (3)

▶ Machine Learning Algorithms (10)

▶ Model Training (4)

▶ Model Scoring & Evaluation (2)

▶ Python Language (2)

▶ R Language (1)

▶ Text Analytics (7)

▶ Computer Vision (6)

▶ Recommendation (5)

▶ Anomaly Detection (2)

▶ Web Service (2)

Automobile price data (Raw) result visualization

X

Submit

Publish

Rows ⓘ Columns ⓘ

205 26

mpg	horsepower	peak-rpm	city-mpg	highway-mpg	price
18	82	438	15	22	13495
18	110	500	19	27	16500
18	154	500	19	26	16500
18	102	5500	24	20	12050

price

Statistics

Mean	13207.1294
Median	10295
Min	5118
Max	45400
Standard deviation	7947.0663
Unique values	186
Missing values	4
Feature type	Numeric Feature

Visualizations

Close

Train Model

Score Model

Navigator

Datasets

Modules

Models



Autosave on



70%

Run finished [View run](#)

Animal Images Dataset

Convert to Image Directory

Convert image dataset to "Image Directory"

Completed

Split Image Directory

Split the input dataset into training/validation/test datasets

Completed

④ Init Image Transformation

Initialize transformation which will be applied

Completed

④ Split Image Directory

Split the input dataset into training dataset

Completed

④ Apply Image Transformation

Apply transformation defined by Init Image

Completed

④ Apply Image Transformation

Apply transformation defined by Init Image

Completed

④ Train PyTorch Model

Train DenseNet model, recommended to run

Completed

④ Score Image Model

Predict image class of test dataset using

Completed

Build and train models without writing code

Navigator

Search by name, tag

Modules

▶ Data Input and Output (3)

▶ Data Transformation (19)

▶ Feature Selection (2)

▶ Statistics (1)

▶ Machine Learning Algo

▶ Model Training (4)

▶ Model Scoring & Evaluati

▶ Python Language (2)

▶ R Language (1)

▶ Text Analytics (7)

▶ Computer Vision (6)

▶ Recommendation (5)

▶ Anomaly Detection (2)

▶ Web Service (2)



Evaluate Model result visualization



Pipeline ▾

Publish

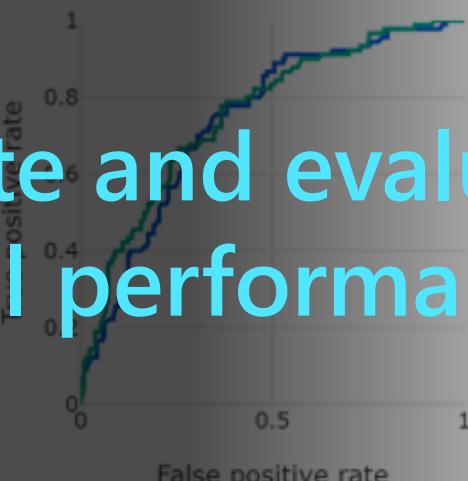
Left port

Right port

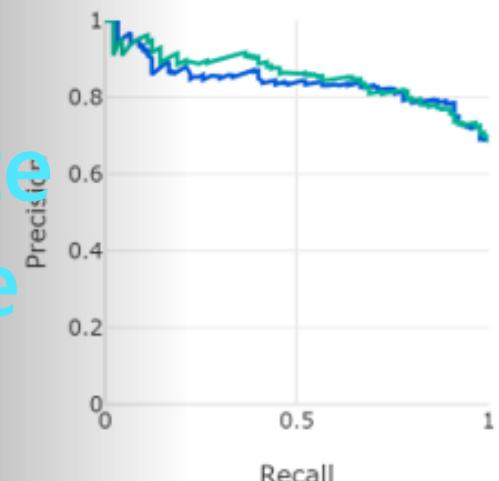
● Scored dataset (left port)

● Scored dataset to compare (right port)

ROC curve



Precision-recall curve



Lift curve

Compare the performance of two-class SVM
Completed

Close

 Navigator

Validate and evaluate model performance

Training pipeline

Real-time inference p

Set up real-time endpoint

X

Deploy new real-time endpoint Replace an existing real-time endpoint

Real-time endpoint name *

price-prediction-endpoint

Endpoint description (optional)



Modules

▶ Data Input and Output (3)

▶ Data Transformation (19)

▶ Feature Selection (1)

▶ Statistical Functions (1)

▶ Machine Learning Algorithms (4)

▶ Model Training (4)

▶ Model Scoring & Evaluation (6)

▶ Python Language (2)

▶ R Language (1)

▶ Text Analytics (7)

▶ Computer Vision (6)

▶ Recommendation (5)

▶ Anomaly Detection (2)

▶ Web Service (2)

Navigator

Advanced topics

- Bring in your own code
- UI <-> code interoperability
- Automate your machine learning pipeline

Real-time inference pipeline

Search by name, tags and description

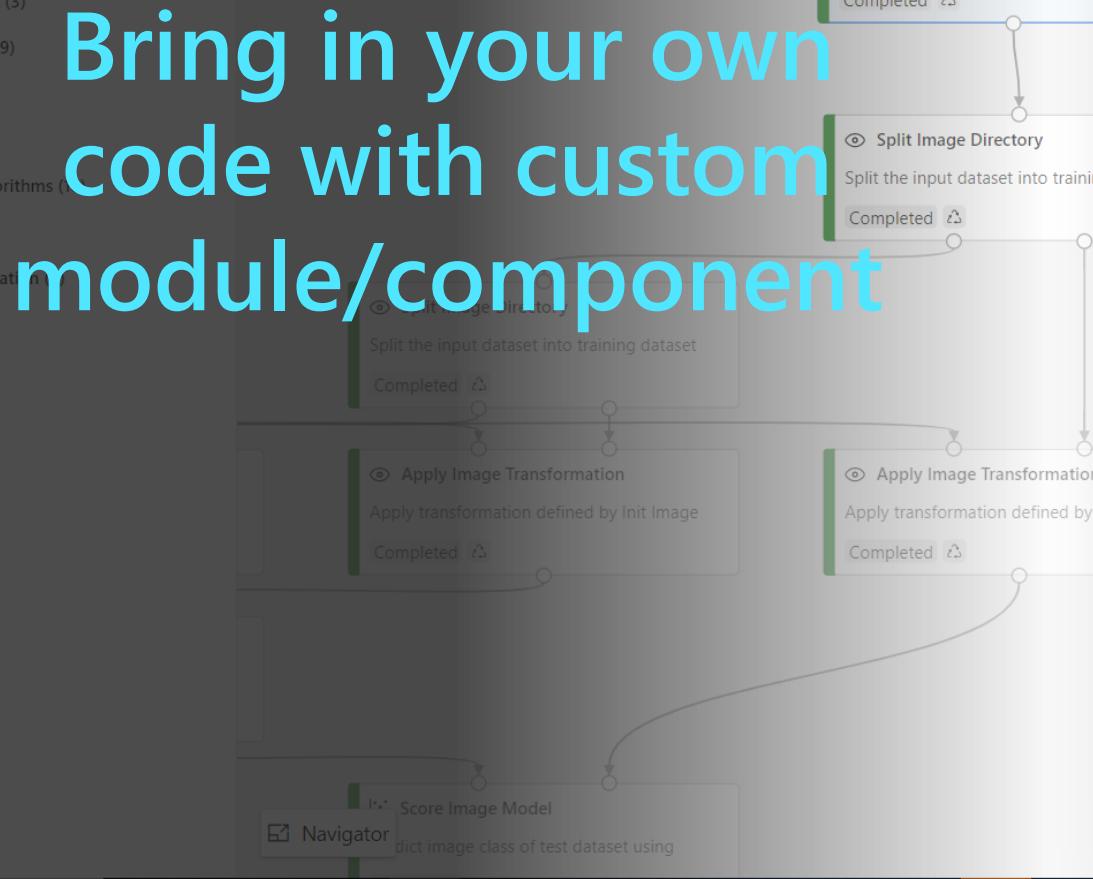
Assets in total



Bookstore-Warehouse-Subway Image Classification using DenseNet



100%



Module definition

```

1 amlModuleIdentifier:
2   namespace: AML-Data-Labeling-WS
3   moduleName: Convert Labeling Data to Image Directory
4   moduleVersion: 0.0.2
5   isDeterministic: True
6   jobType: Basic
7   inputs:
8     - name: Labeling data
9       type: AzureMLdataset
10      optional: False
11      argumentName: labeling_data
12   outputs:
13     - name: Output image directory
14       type: ImageDirectory
15       argumentName: output_image_dir
16   implementation:
17     container:
18       amlEnvironment:
19         python:
20           condaDependencies:
21             channels:
22               - defaults
23             dependencies:
24               - python=3.6.5
25               - pip:
26                 - azureml-contrib-dataset
27                 - azureml-dataset-runtime==1.15.0
28                 - azureml-designer-core[image]==0.0.41
29                 - fire
30             name: project_environment
31     docker:
32       baseImage: mcr.microsoft.com/azureml/base:intelmpi2018.3-ubuntu16.04
33       command: [python, convert_labeling_data_to_image_directory.py]
34       args: [-labeling_data, {inputPath: Labeling data}, --output_image_dir, {outputPath: Output image directory}]
35       os: Linux
36 ...
37
  
```

Copy**Cancel**

Inference pipeline

Task description

Bookstore-Warehouse-Subway Image Classification using DenseNet



Submit

Create inference pipeline

Update inference pipeline



Autosave on



100%



Convert Labeling Data to Image Di...

Completed

Split Image Directory

Split the input dataset into training/validation

Completed

UI <-> code interoperability

Split Image Directory

Split the input dataset into training dataset

Completed

Apply Image Transformation

Apply transformation defined by Init Image

Completed

Apply Image Transformation

Apply transformation defined by Init Image

Completed

```
# get dataset
from azureml.core import Dataset
bookstore_subway_warehouse_20201102_002426 = Dataset.get_by_name(ws, name='bookstore_subway_warehouse_20201102_002426', version=1)

# define pipeline
@ds.pipeline(name='Bookstore-Warehouse-Subway Image Classification using DenseNet', description='this sample shows how to use computer vision modules to build image classification model. The algorithm is DenseNet based on PyTorch.', default_compute_target='gpu-cluster1', default_datastore='workspaceblobstore')
def generated_pipeline():
    init_image_transformation_0 = init_image_transformation_func(
        resize=True,
        size=256,
        center_crop=True,
        crop_size=224,
        pad=False,
        color_jitter=False,
        grayscale=False,
        random_resized_crop=False,
        random_crop=False,
        random_horizontal_flip=True,
        random_vertical_flip=False,
        random_rotations=False,
        random_affine=False,
        random_grayscale=False,
        random_perspective=False
    )

    convert_labeling_data_to_image_directory_0 = convert_labeling_data_to_image_directory_func(
        labeling_data=bookstore_subway_warehouse_20201102_002426
    )

    split_image_directory_0 = split_image_directory_func(
        input_image_directory=convert_labeling_data_to_image_directory_0.outputs.output_image_dir,
        fraction_of_images_in_the_first_output=0.9
    )

    split_image_directory_1 = split_image_directory_func(
        input_image_directory=split_image_directory_0.outputs.output_image_directory,
        fraction_of_images_in_the_first_output=0.9
    )

    apply_image_transformation_0 = apply_image_transformation_func(
        input_image_directory=split_image_directory_1.outputs.output_image_directory,
        input_image_transformation=init_image_transformation_0.outputs.output_image_transformation,
        mode='for training'
    )

    apply_image_transformation_1 = apply_image_transformation_func(
        input_image_transformation=init_image_transformation_0.outputs.output_image_transformation,
        input_image_directory=split_image_directory_1.outputs.output_image_directory,
        mode='for inference'
    )

    apply_image_transformation_2 = apply_image_transformation_func(
        input_image_transformation=init_image_transformation_0.outputs.output_image_transformation,
        input_image_directory=split_image_directory_0.outputs.output_image_directory,
        mode='for inference'
    )
```

Notebook.ipynb

Python.py



Search by name, tags and description

New module default versions are available. Update to reflect the latest changes.

Check details Keep as-is for all

in total (43) datasets (16) Module (1)

Labeling Data to I... Version 0.0.2 11/2/2020

ut and Output (3) nsformation (19) Selection (2) Functions (1) Learning Algorithms (18) aining (4) Scoring & Evaluation (6) language (2) age (1) ytics (7) er Vision (6) endation (5) Detection (2) vice (2)

Automate your ML pipeline

Copy of Bookstore-Warehouse-Subway Image Classification using DenseNet on 11-02-2020

Autosave on 100% 1:1

Draft autosaved on 11/2/2020, 1:36:07 PM View ru

Submit Publish

Set up published pipeline

PipelineEndpoint

Select existing Create new

Existing PipelineEndpoint *

Published pipeline

Copy of Bookstore-Warehouse-Subway Image Classification ...

Set as default pipeline for this endpoint.

Continue on failure step

Pipeline parameters and default values

DataSet1 8f4e0d6a-9a81-45ee-bba9-4...

Convert Labeling Data to Image Di...

Split Image Directory

the input dataset into training/validation set (0.9) and test dataset (0.1).

Apply Image Transformation

Apply transformation defined by Init Image Transformation module on the training

Train PyTorch Model

Train DenseNet model, recommended to run on GPU compute for better performance.

Score Image Model

Predict image class of test dataset using trained DenseNet model.

Navigator

```
graph TD; A[Convert Labeling Data to Image Di...]; A --> B[Split Image Directory]; B --> C1[Apply Image Transformation]; C1 --> D[Train PyTorch Model]; D --> E[Score Image Model];
```

Resources

- Sign up in Azure: <https://azure.microsoft.com>
- Azure Machine Learning designer:
<https://azure.microsoft.com/en-us/services/machine-learning/designer/>
- Tutorial: <https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-train-score>
- Samples: <https://docs.microsoft.com/en-us/azure/machine-learning/samples-designer>
- Machine Learning cheat sheet: <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet>

Data labeling with AML



Usage Scenario

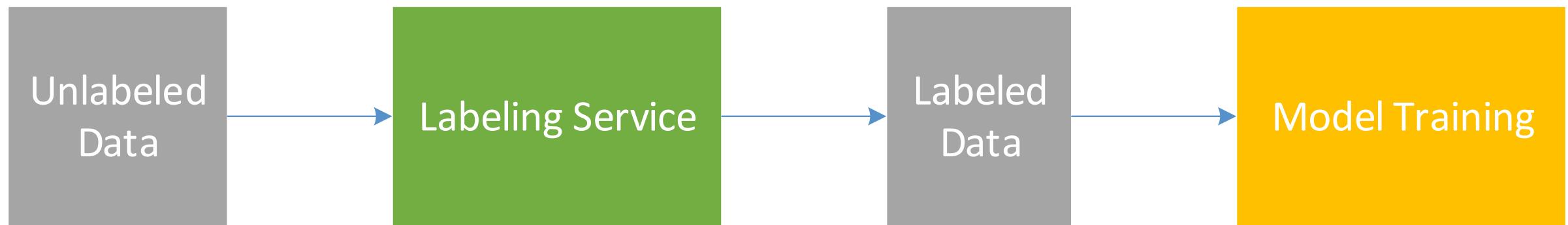


Image labelling

The screenshot shows the Microsoft Azure Machine Learning Data Labeling interface. The title bar reads "Preview Microsoft Azure | Machine Learning Data Labeling". Below it, the navigation bar shows "All Projects > Animal Classes (Multiclass)". The main area is titled "Animal Classes (Multiclass)". It has two tabs: "Instructions" (selected) and "Tasks". A button "Select all (2 selected)" is visible. There are four image thumbnails: a lizard, a bear, a whale, and a shark. The shark thumbnail has a green label "Cartilaginous fish" with a close button "X". On the right, there is a "Tags" sidebar with checkboxes for Mammal, Bird, Bony fish, Cartilaginous fish, Reptile, Anthozoan, and Other/Unknown. The "Mammal" checkbox is checked. At the bottom is a "Submit" button.

Multi-class classification

The screenshot shows the Microsoft Azure Machine Learning Data Labeling interface for medical image labeling. The title bar reads "Home > Data Labeling > Chest X-rays thorax diseases. > Label". Below it, the navigation bar shows "Chest X-rays thorax diseases.". A disclaimer states: "Disclaimer: The capability to label DICOM or similar image types is not intended or made available for use as a medical device, clinical support, diagnostic tool, or ot...". The main area has "Instructions" and "Tasks" tabs. A search bar and filter icon are at the top. A chest X-ray image is displayed with "AP" and "PORTABLE" labels. On the right, a "Tags" sidebar lists eight categories: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, and Pneumothorax. The "Atelectasis" category is selected, indicated by a blue circle. A "Search tags" input field and "Shortcut Keys" link are also present.

Support medical images

[Set up image labeling project - Azure Machine Learning | Microsoft Learn](#)

Image labelling

All Projects > Photo labels (Multilabel)

Photo labels (Multilabel)

Instructions Tasks

Select all (2 selected)

Closeup X Ocean X

Wideangle X Land X

Submit

Tags

- Land
- Ocean
- Closeup
- Wideangle

Multi-label classification

Photo Subject ID (Object Identification)

Instructions Tasks

1

Search tags

2

3

Submit

Tags

- (#1) Shark
- (#2) Dog
- (#3) Cat
- (#4) Bird
- (#5) Other

Object identification

Image labelling

Instance segmentation

Text labeling

Multi-label classification

Multi-class classification

	Product	Consumer_complaint_narrative	category_id
1	Credit reporting	I have outdated information on my credit repor...	0
2	Consumer Loan	I purchased a new car on XXXX XXXX. The car de...	1
7	Credit reporting	An account on my credit report has a mistaken ...	0
12	Debt collection	This company refuses to provide me verificatio...	2
16	Debt collection	This complaint is in regards to Square Two Fin...	2

	ID	TITLE	ABSTRACT	Computer Science	Physics	Mathematics	Statistics	Quantitative Biology	Quantitative Finance
0	1	Reconstructing Subject-Specific Effect Maps	Predictive models allow subject-specific inf...	1	0	0	0	0	0
1	2	Rotation Invariance Neural Network	Rotation invariance and translation invarian...	1	0	0	0	0	0
2	3	Spherical polyharmonics and Poisson kernels fo...	We introduce and develop the notion of spher...	0	0	1	0	0	0
3	4	A finite element approximation for the stochas...	The stochastic Landau--Lifshitz--Gilbert (LL...	0	0	1	0	0	0
4	5	Comparative study of Discrete Wavelet Transfor...	Fourier-transform infra-red (FTIR) spectra o...	1	0	0	1	0	0

Ousted WeWork founder Adam Neumann lists his Manhattan penthouse for \$37.5 million

[organization]

[person]

[location]

[monetary value]

Named Entity Recognition

Challenges & Solutions

Challenges

Data

Volume of Data

Data movement

Tooling

Requires labeling tools

Workflow

Need to iterate

Distribution of labeling tasks

Project Management

Quality Control
(human bias)

Rate of labeling

Cost concerns

Fundamentals

Security

Privacy

Builds on Azure ML datasets, integrated with Azure ML

Prebuilt tools
- browser based
- Multi-user

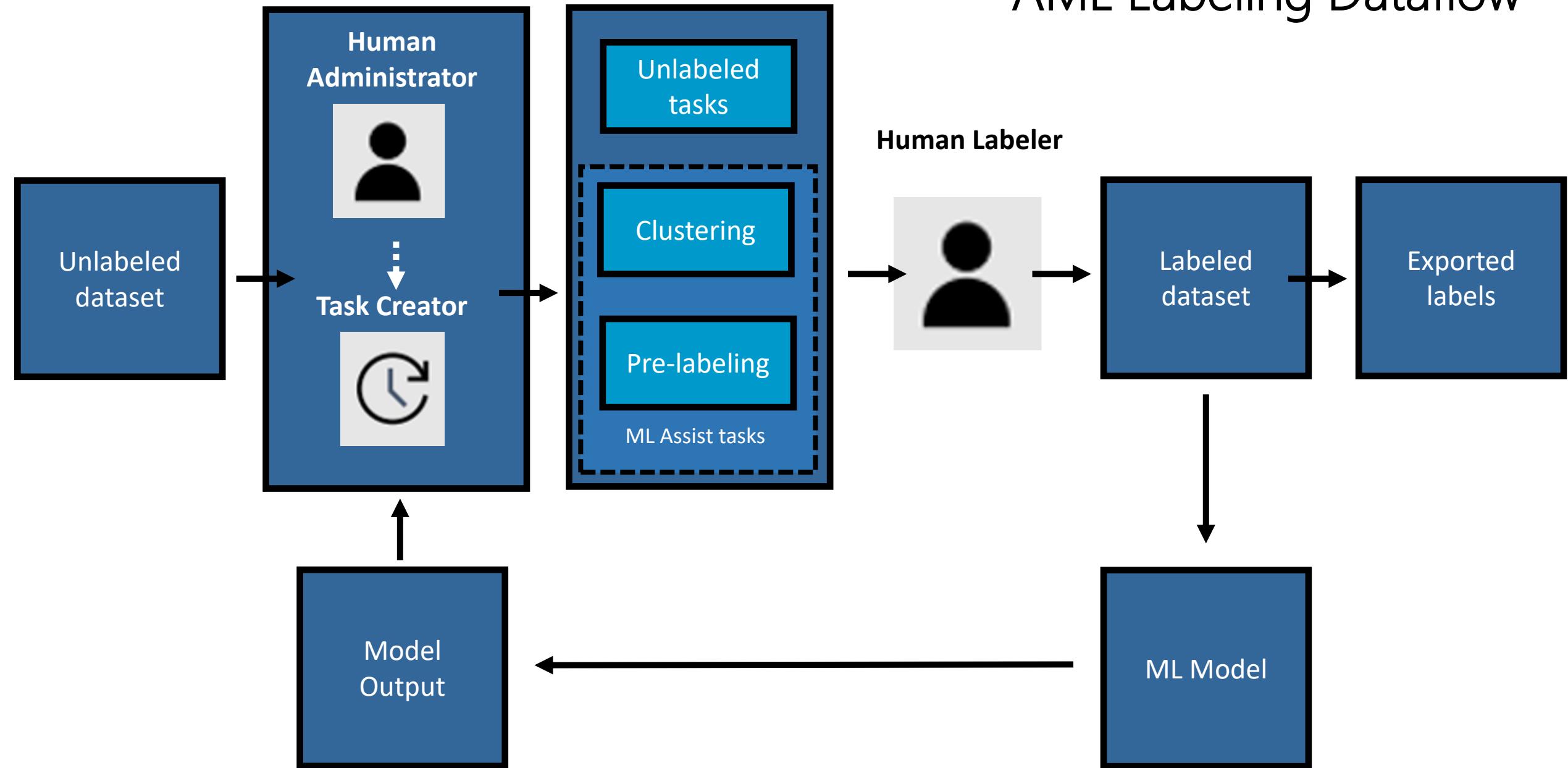
Supports iteration, automatic assignment of tasks, and task review

ML-assisted labeling

Various capabilities

Controlled access, authenticated using AAD or MSA

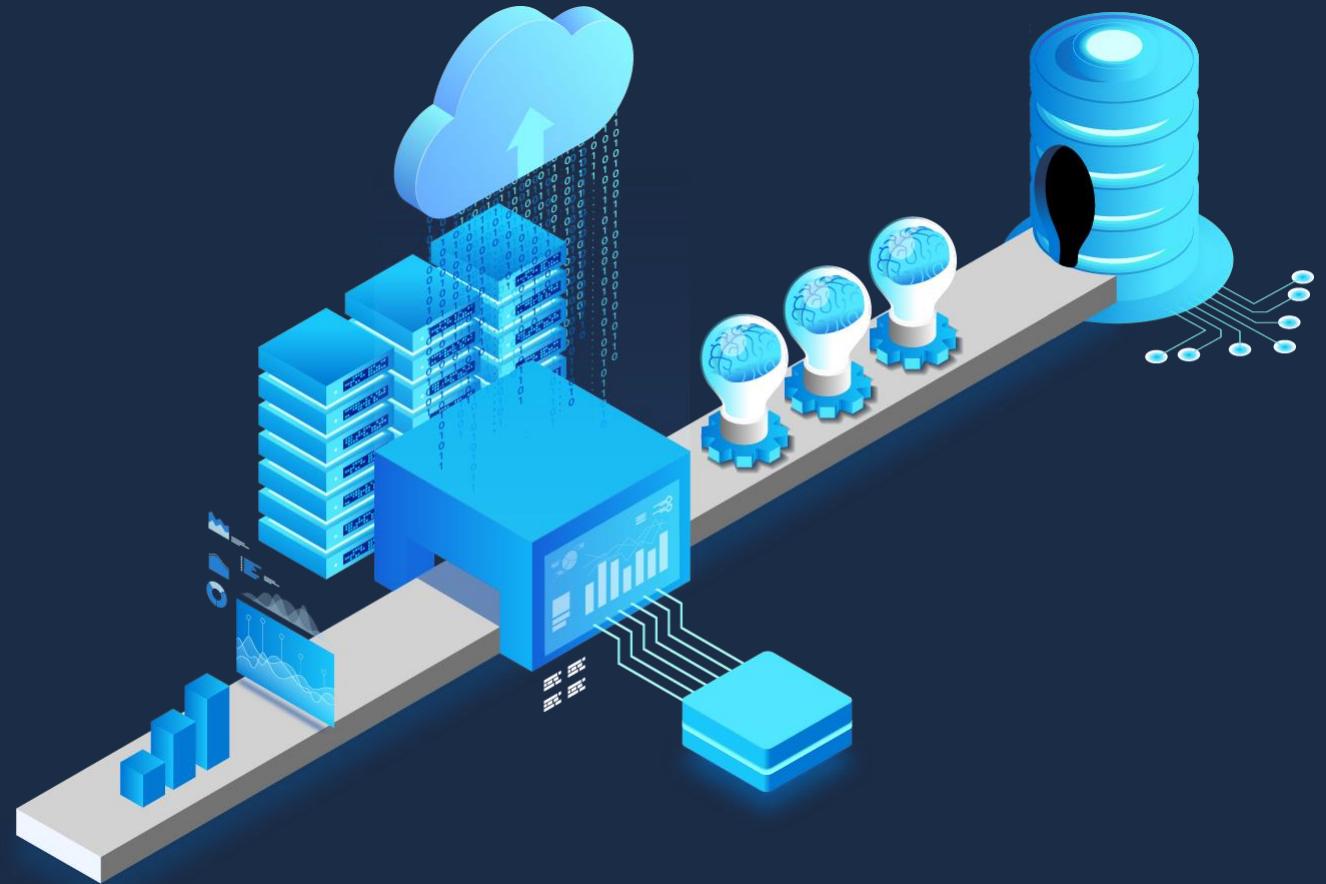
AML Labeling Dataflow

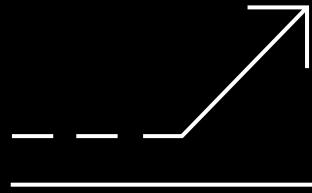


Data Labeling Demo



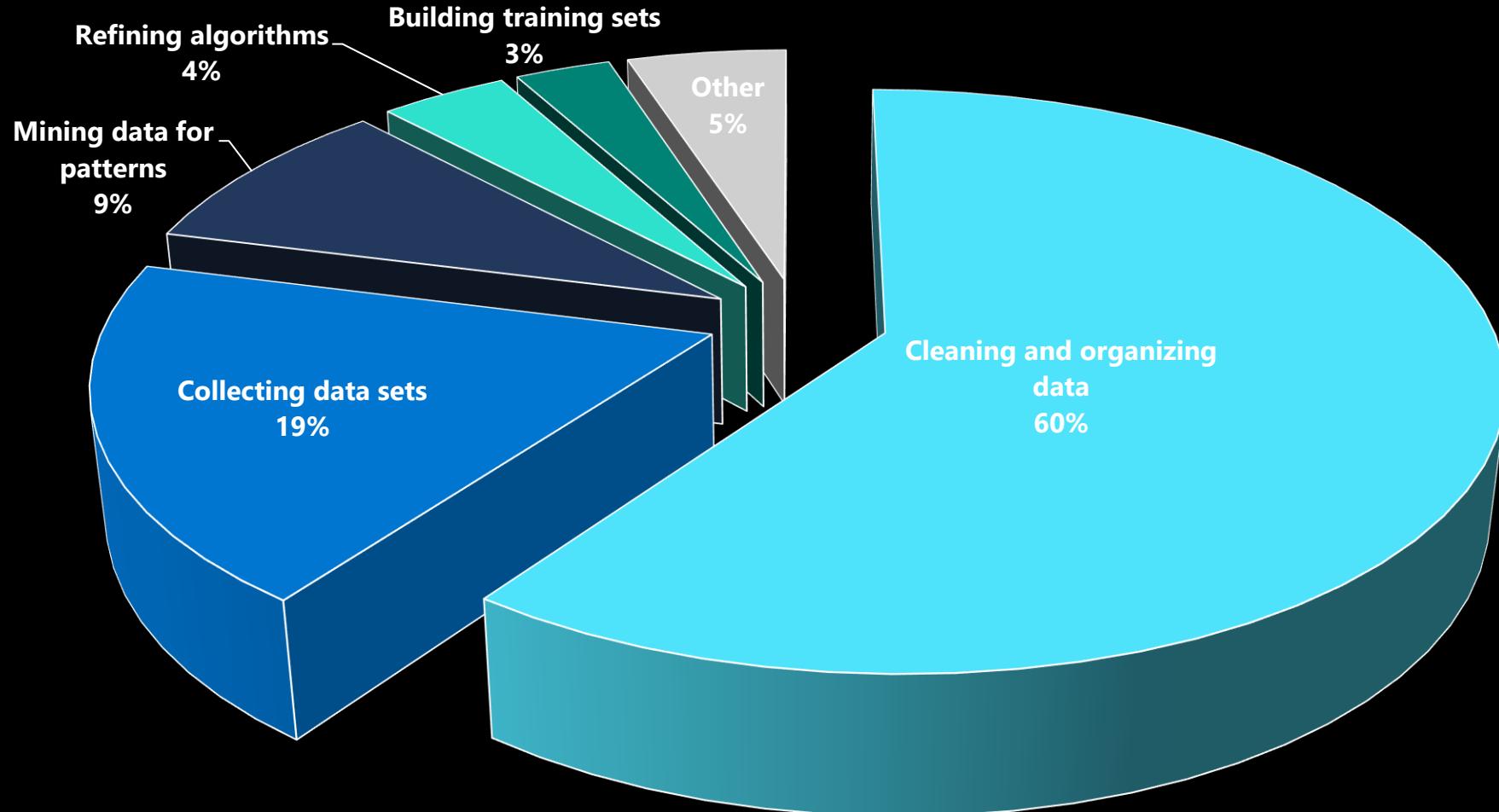
Automated ML





Automated Machine Learning
'simplifies' the creation and selection
of the optimal model

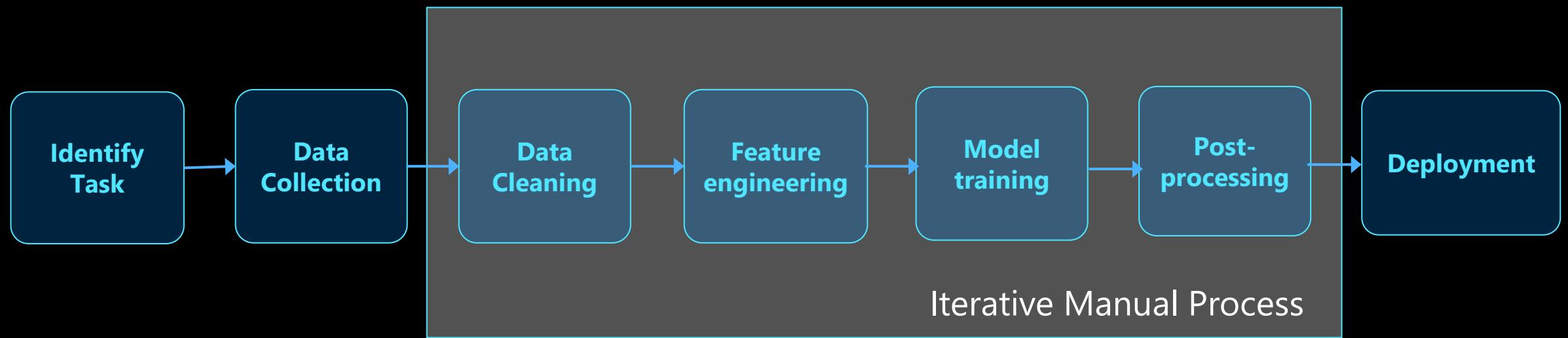
What data scientist spend the most time doing?



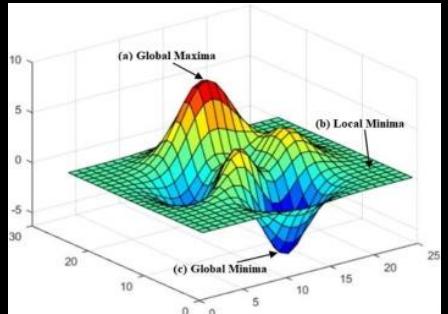
Source : Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says (forbes.com)



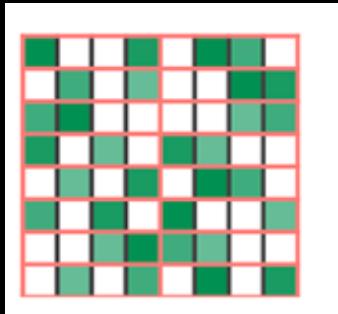
Challenges in Designing ML pipelines



Complex search space



Sparsity of good configurations



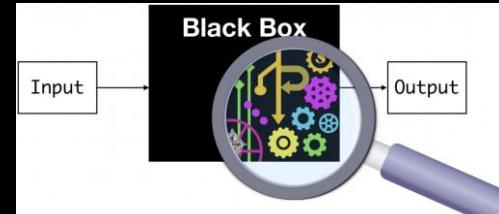
Expensive evaluations



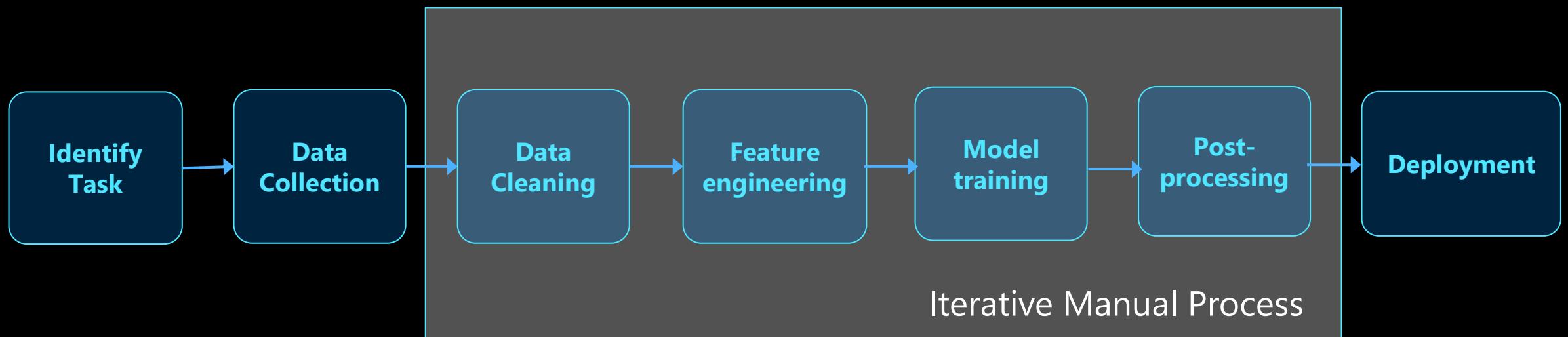
Noise on observations



Black-Box Problem



From Manual to Automated ML



Azure Machine Learning accelerates model development with automated machine learning

Input

Intelligently test multiple models in parallel

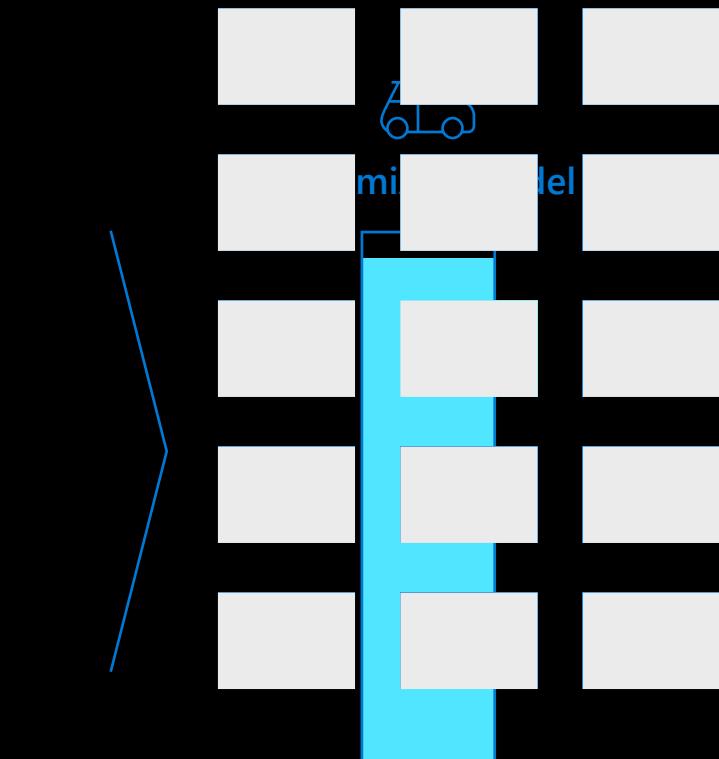
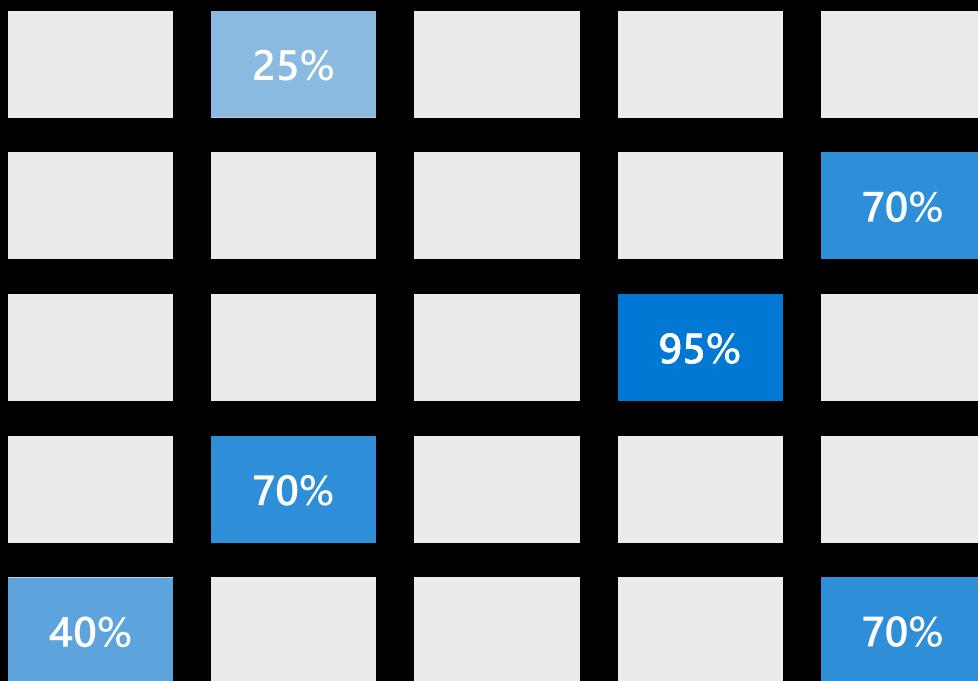
101010
010101
101010



Define goals



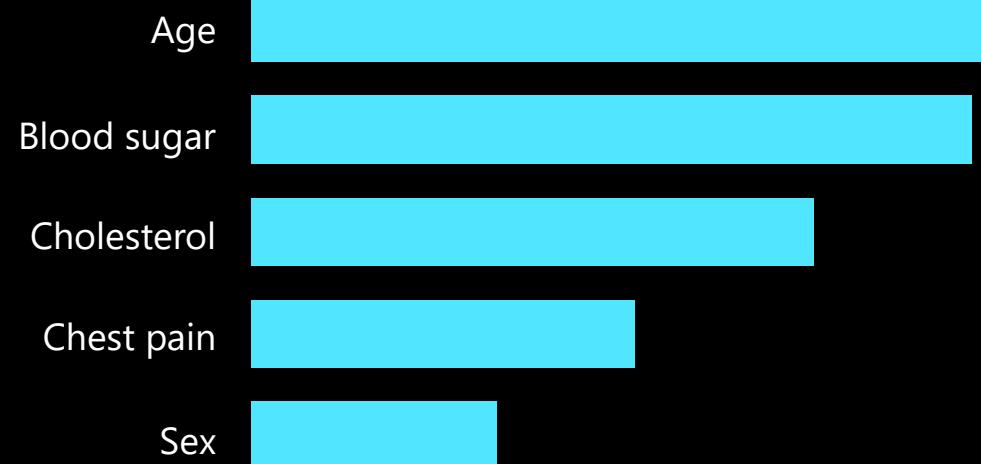
Apply constraints



Azure Machine Learning accelerates model selection with model explainability

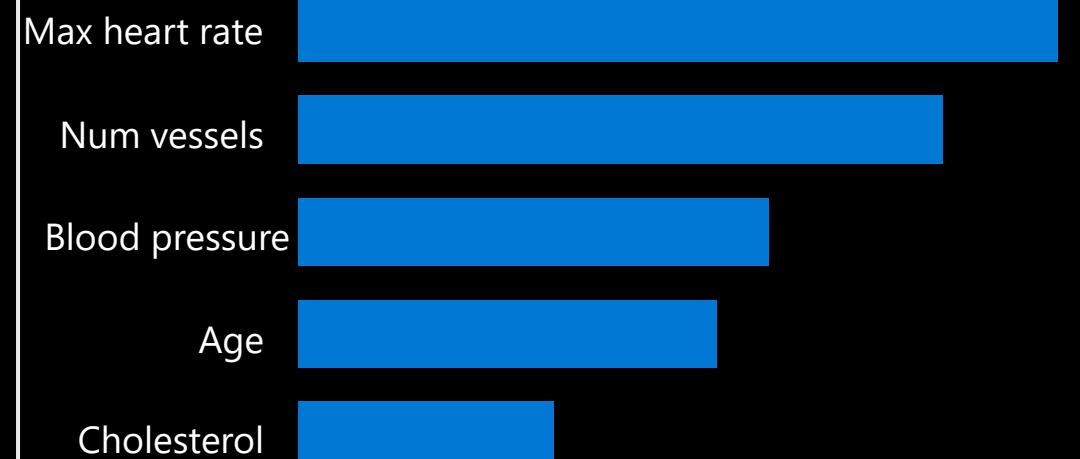
Feature importance

Model A (95%)



Feature importance

Model B (70%)



Azure Machine Learning accelerates model selection with code generation [Public Preview]

AutoML Model's training code generation (i.e. Scikit-Learn code, Transformers and Algorithms)

AutoML UI
(Experiment's model leader board)

Microsoft > cesardi-automl-centraluseuap-ws > Experiments > SDK_Codegen_remote_porto_seguro

CodeGen_Internal_SDK_remote_porto_seguro_1 🖊 ⭐

Refresh Edit and submit Cancel Delete

Details Data guardrails Models Outputs + logs Child runs Snapshot

Refresh Deploy Download Explain model # View generated code (preview) 📄

Search

Showing 1-10 of 10 models

Algorithm name	Explained	AUC weighted ↓
VotingEnsemble	View explanation	0.63816
MaxAbsScaler, XGBoostClassifier		0.63739
MaxAbsScaler, LightGBM		0.63594
TruncatedSVDWrapper, XGBoostClassifier		0.62862
MaxAbsScaler, LightGBM		0.62337
StandardScalerWrapper, XGBoostClassifier		0.62047

script.py with model's training generated code (Scikit-Learn, etc.)

```
def generate_algorithm_config():
    from xgboost.sklearn import XGBClassifier

    algorithm = XGBClassifier(
        base_score=0.5,
        booster='gbtree',
        colsample_bylevel=1,
        colsample_bynode=1,
        colsample_bytree=1,
        gamma=0,
        learning_rate=0.1,
        max_delta_step=0,
        max_depth=3,
        min_child_weight=1,
        missing=None,
        n_estimators=100)
    return algorithm

#...
def build_model_pipeline():
    from sklearn.pipeline import Pipeline
    pipeline = Pipeline(
        steps=[
            ('featurization', generate_data_transformation_config()),
            ('preproc', generate_preprocessor_config()),
            ('model', generate_algorithm_config()),])
    return pipeline
#...

model_pipeline = build_model_pipeline()

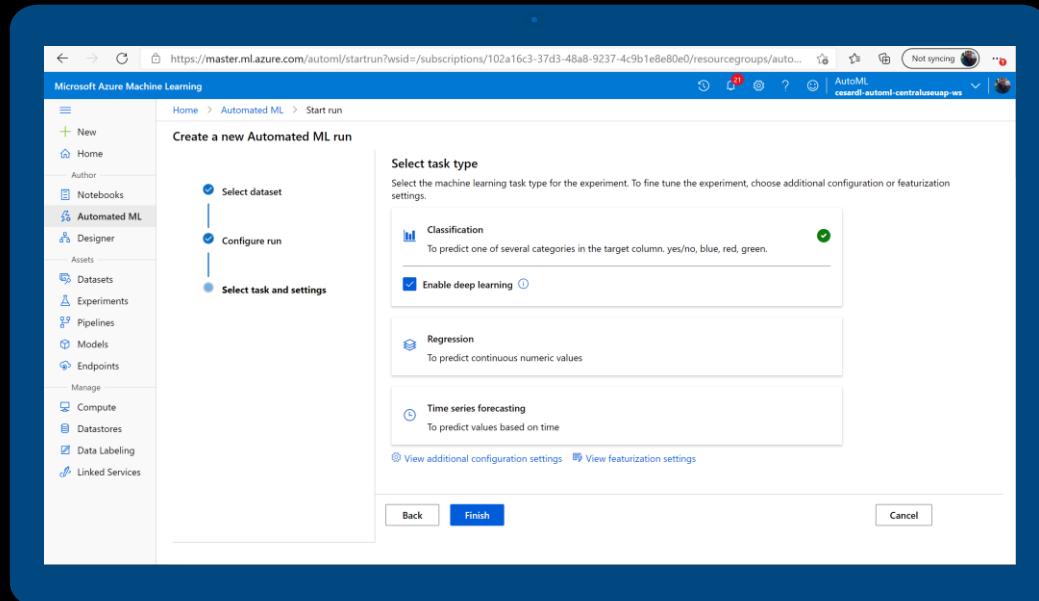
model = model_pipeline.fit(X, y)
```

- Data preprocessing
- Algorithm selection
- Featurization
- Hyperparameters

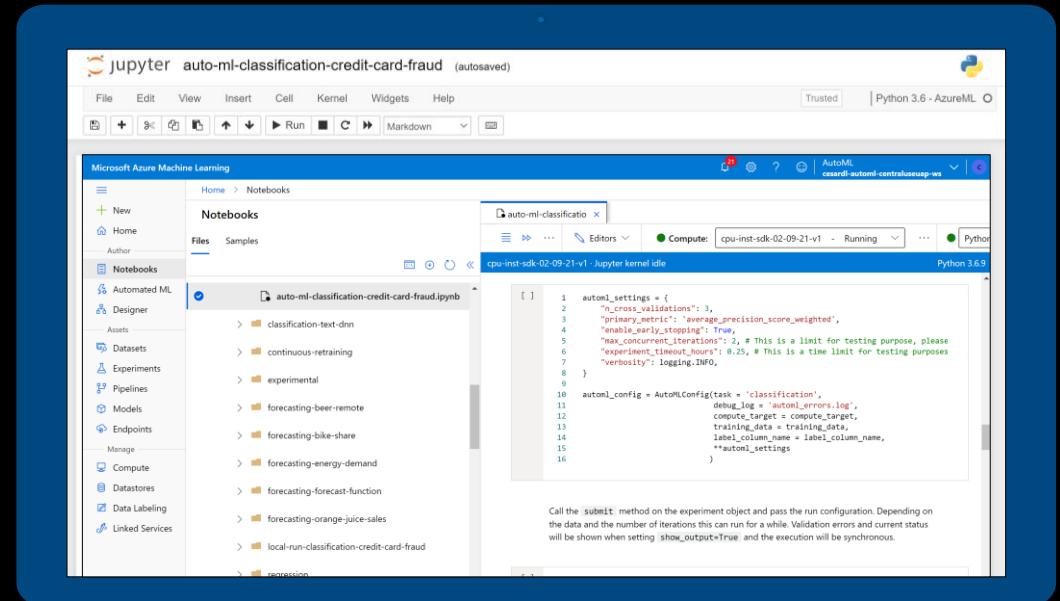
AutoML

UI and Python SDK

Automated ML UI in Azure portal



Python SDK (Notebooks)



End-to-end no-code experience for non-data scientists to train ML models

Easy to get started

Deploy models easily and quickly

Most common scenarios supported

UI also covers results from SDK runs:

- Model metrics
- Model explainability

Full control

Maximum flexibility based on Python
Supports all scenarios and parameters
Sample notebooks for SDK 1.x:

<https://github.com/Azure/MachineLearningNotebooks/tree/master/how-to-use-azureml/automated-machine-learning>

Automated ML

Current Capabilities

Category	Value
ML Problem Spaces	Classification Regression Forecasting Computer Vision (Preview) NLP (Preview)
Frameworks	Scikit Learn
Languages	Python
Data Type and Data Formats	Numerical Text Scikit-learn supported data formats (Numpy, Pandas)
Data sources	Local Files, Azure Blob Storage
Compute Target	Automated Hyperparameter Tuning Azure ML Compute (Batch AI), Azure Databricks Automated Model Selection Local Compute, Azure ML Compute (Batch AI), Azure Databricks



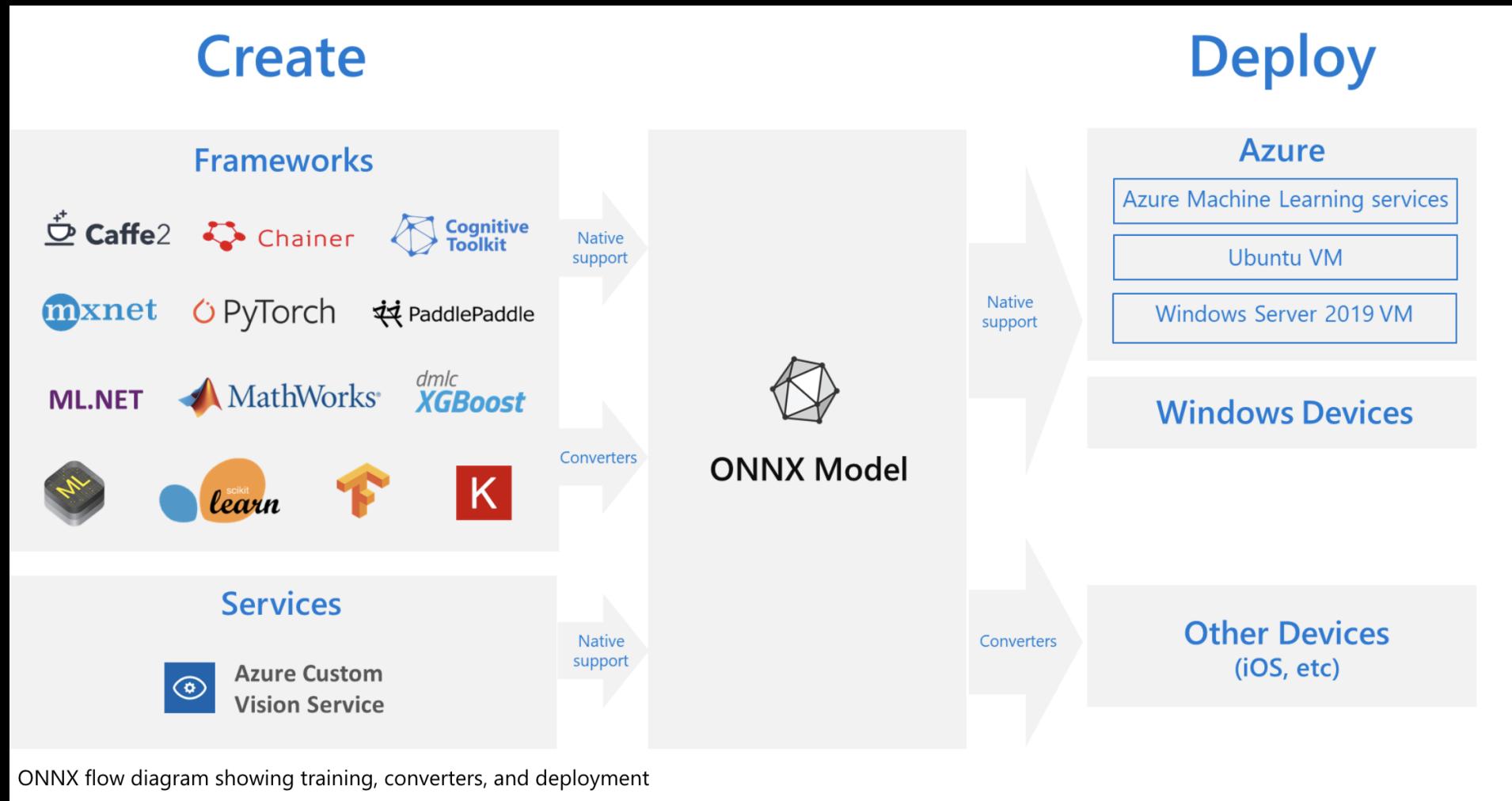
NLP & CV supported algorithms

All the pre-trained text DNN models currently available in AutoML NLP for fine-tuning are listed below:

- bert_base_cased
- bert_large_uncased
- bert_base_multilingual_cased
- bert_base_german_cased
- bert_large_cased
- distilbert_base_cased
- distilbert_base_uncased
- roberta_base
- roberta_large
- distilroberta_base
- xlm_roberta_base
- xlm_roberta_large
- xlnet_base_cased
- xlnet_large_cased

Task	model architectures	String literal syntax <code>default_model</code> * denoted with *
Image classification (multi-class and multi-label)	MobileNet: Light-weighted models for mobile applications ResNet: Residual networks ResNeSt: Split attention networks SE-ResNeXt50: Squeeze-and-Excitation networks ViT: Vision transformer networks	<code>mobilenetv2</code> <code>resnet18</code> <code>resnet34</code> <code>resnet50</code> <code>resnet101</code> <code>resnet152</code> <code>resnest50</code> <code>resnest101</code> <code>seresnext</code> <code>vits16r224</code> (small) <code>vitb16r224</code> *(base) <code>vitl16r224</code> (large)
Object detection	YOLOv5: One stage object detection model Faster RCNN ResNet FPN: Two stage object detection models RetinaNet ResNet FPN: address class imbalance with Focal Loss	<code>yolov5</code> * <code>fasterrcnn_resnet18_fpn</code> <code>fasterrcnn_resnet34_fpn</code> <code>fasterrcnn_resnet50_fpn</code> <code>fasterrcnn_resnet101_fpn</code> <code>fasterrcnn_resnet152_fpn</code> <code>retinanet_resnet50_fpn</code>
Instance segmentation	MaskRCNN ResNet FPN	<code>maskrcnn_resnet18_fpn</code> <code>maskrcnn_resnet34_fpn</code> <code>maskrcnn_resnet50_fpn</code> * <code>maskrcnn_resnet101_fpn</code> <code>maskrcnn_resnet152_fpn</code>

What is ONNX?



Q&A





Enregistrez vous dès maintenant au prochain Webinars Data AI

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Déploiement et sécurisation des workspace Azure Synapse	09/02/2023	120	En cours
Azure Machine Learning pour les Citizen Data Scientists	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Hybridation des services de données Azure	01/06/2023	120	En cours

Annexes

