



GLOBAL AI FEST:
VIRTUAL ARTIFICIAL
INTELLIGENCE AND
MACHINE LEARNING
CONFERENCE
31ST JULY 2020

WWW.GLOBALAZUREEVENTS.COM

Data Scientist versus Automated ML, who's best ?

Automated ML GUI versus SDK, who's best ?



In less than one hour and for less than one dollar of Azure consumption, I obtained an accuracy of 0.98529 with an algorithm I didn't even know the name of, while the salary of a Data Scientist is \$90k () a year minimum!*

Run 56 ✓ Completed

⟳ Refresh ⚡ Deploy ⬇ Download 🌐 Explain model ✖ Cancel

Details Model Explanations (preview) Metrics Outputs + logs Images Child runs Snapshot

Model summary

Algorithm name
MaxAbsScaler, ExtremeRandomTrees

Accuracy
0.98529 ☰ View all other metrics

Sampling
100% ⓘ

Registered models
No registration yet

Deploy status
No deployment yet

(*) No idea if this value is true...

Agenda

The common tasks of Data Scientists

Definition of Automated Machine Learning

1st round : interface under Azure ML studio (new portal)

2nd round: SDK azureml.core

Result of the game

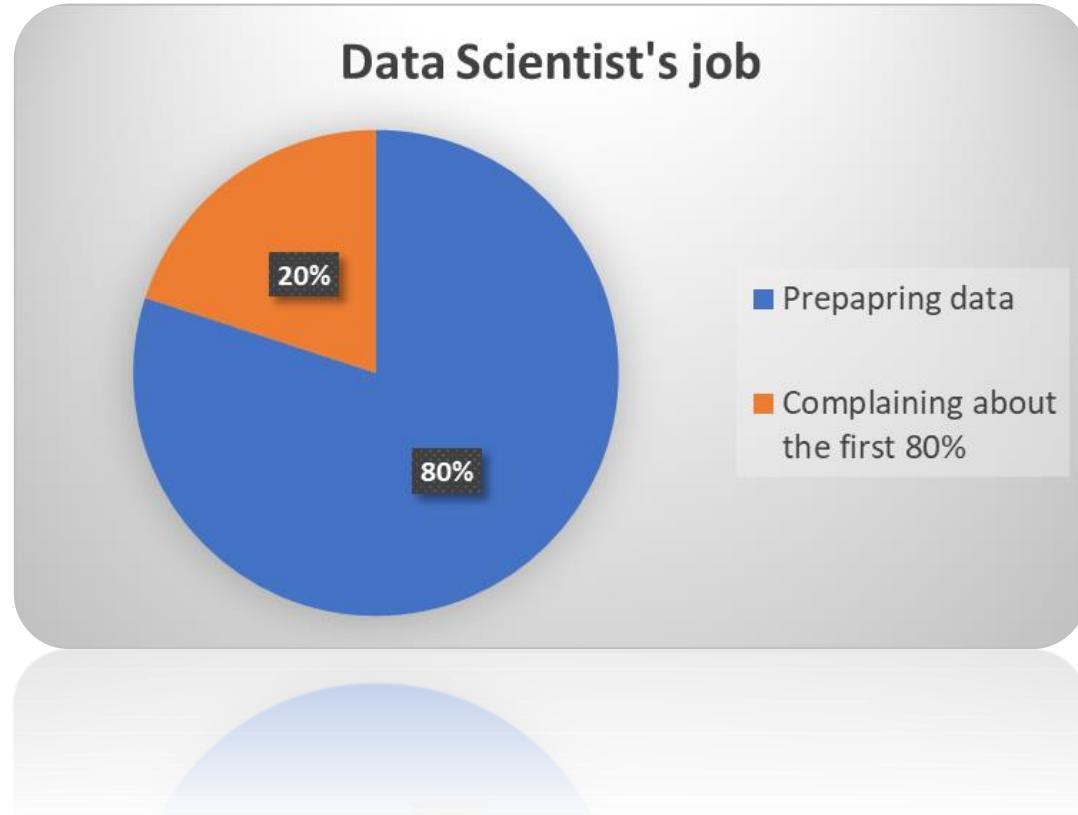
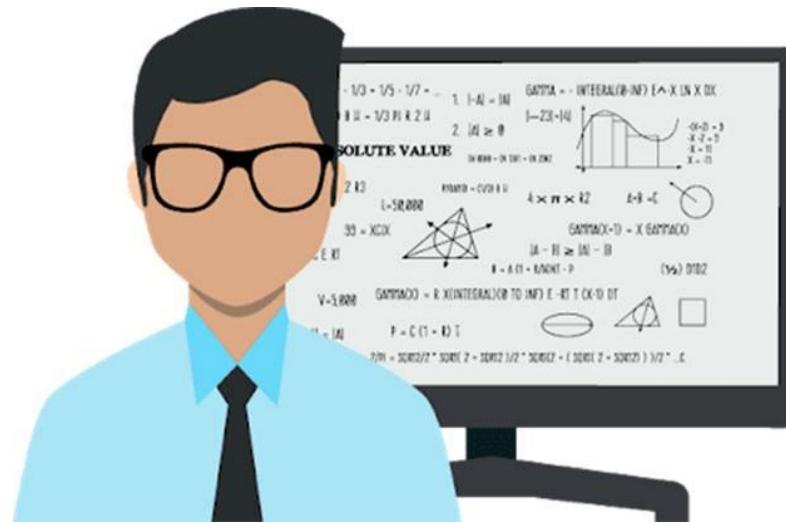


Datasets & notebooks are available on :
<https://github.com/methodidacte/azureml>



Paul PETON – Tech Lead Data & AI

What is the Pareto's law for Data Scientists ?



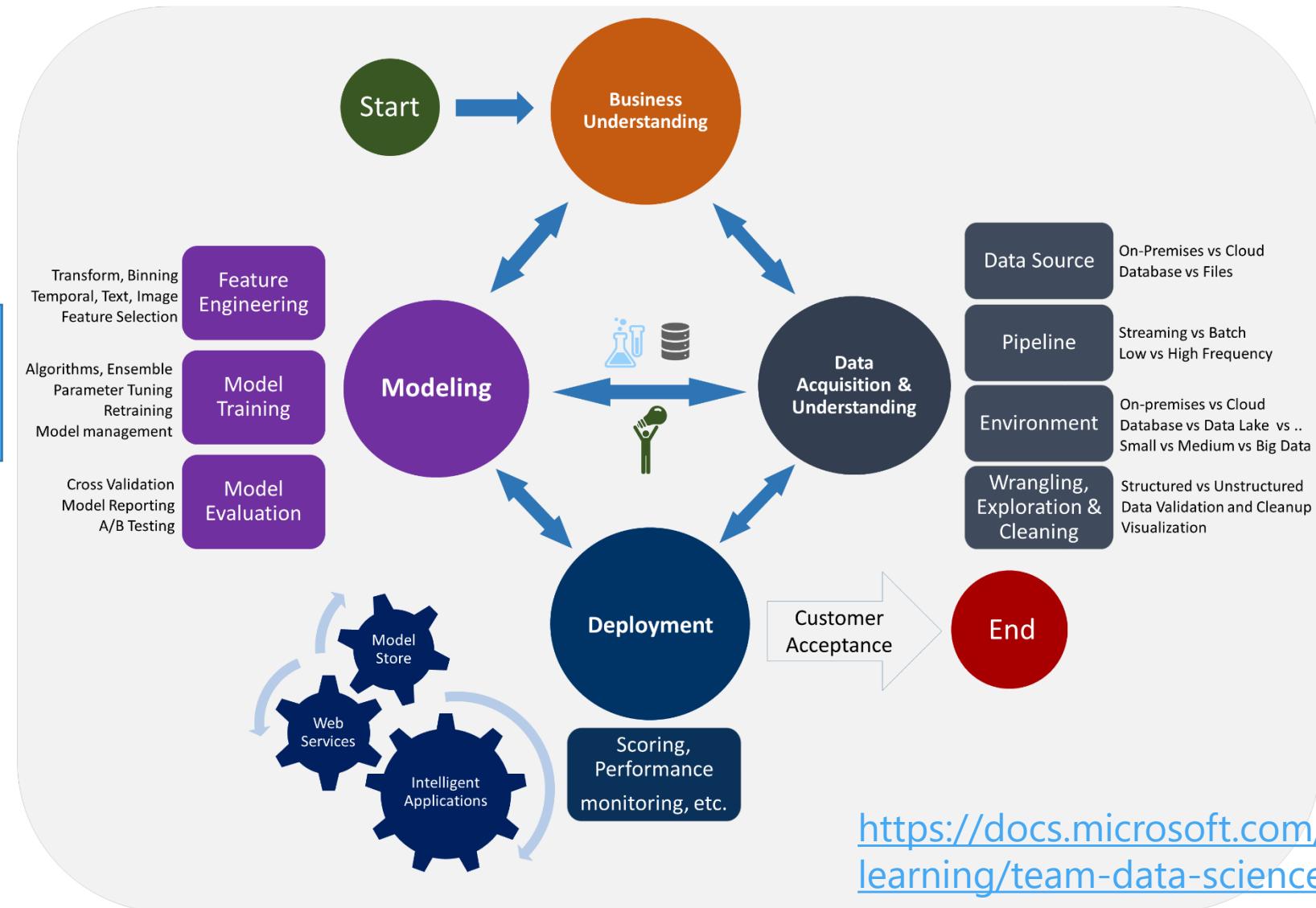
Advisory : no data scientist will be harmed during this talk.

The team data science process

Data Science Lifecycle

Choose the right model : the most uncertain task

Dataprep : the longest step



<https://docs.microsoft.com/fr-fr/azure/machine-learning/team-data-science-process/overview>

Data Scientist whole job

Understand the business need and reformulate it into Data Science process

Carry out data acquisition from one or more sources

Explore the data

Prepare the data :

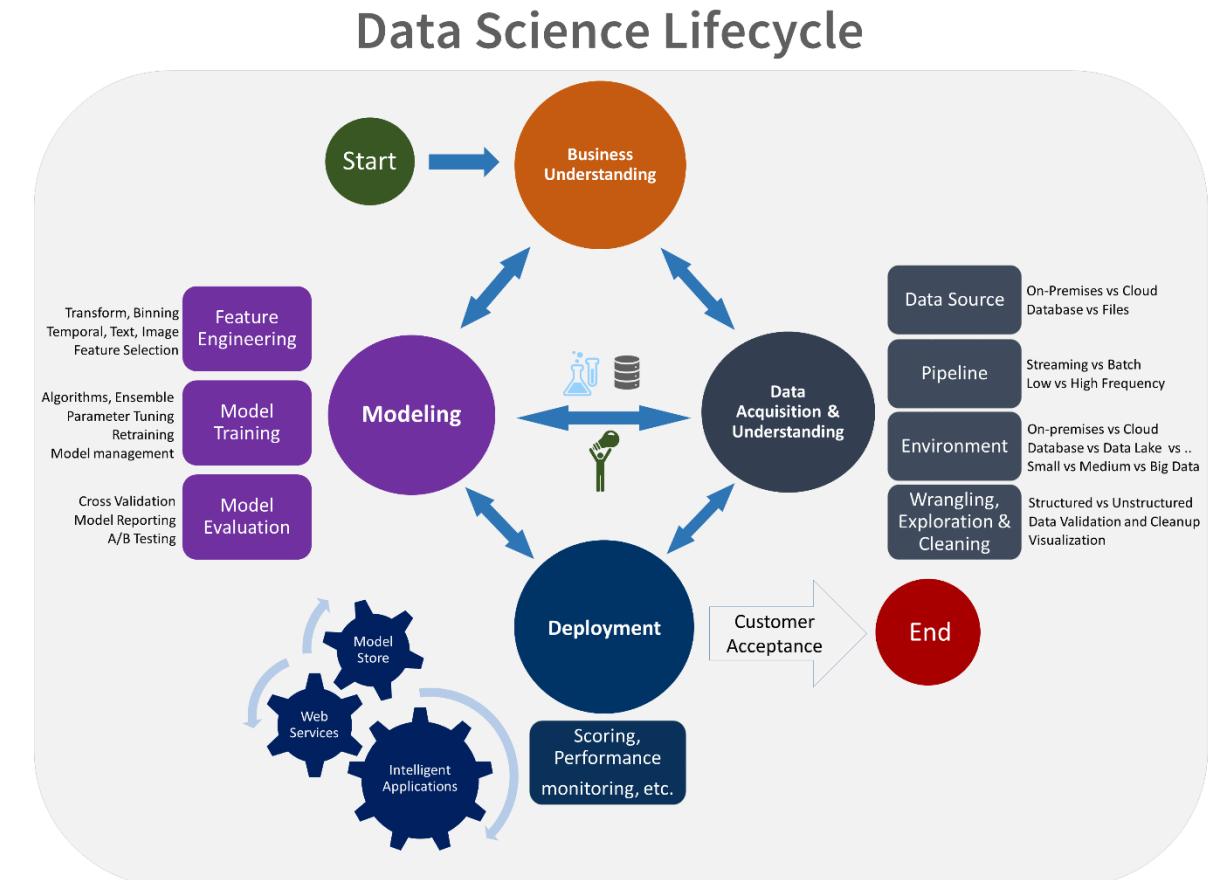
- Clean up missing values or outliers
- Reduce the dimension
- Select the best features

Training several models:

- Trying different approaches (linear or non-linear)
- Select and calculate evaluation metrics
- Identify the best hyperparameters ("gridsearch")

Interpret the results obtained

Deploy the models (or help the Data Engineers to do so !)



Automated Machine Learning : definition

Automated ML is a "brute force" approach to testing many algorithms and keeping the best according to the chosen evaluation metric.

"Theoretical" advantages :

Implementing Machine Learning solutions without having an in-depth knowledge of programming.

Save time and resources.

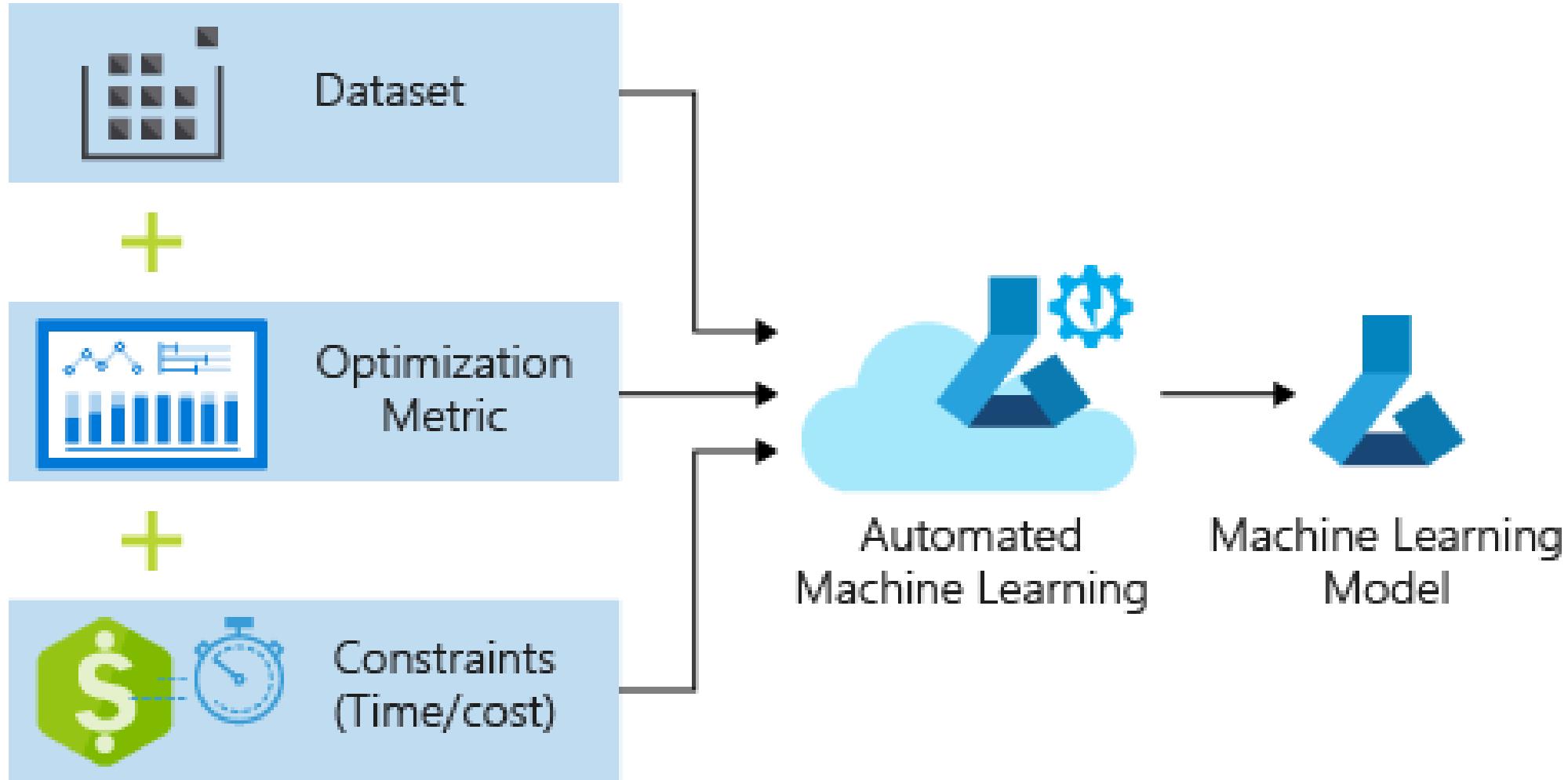
Leverage best practices in data science.

Provide agile problem solving.

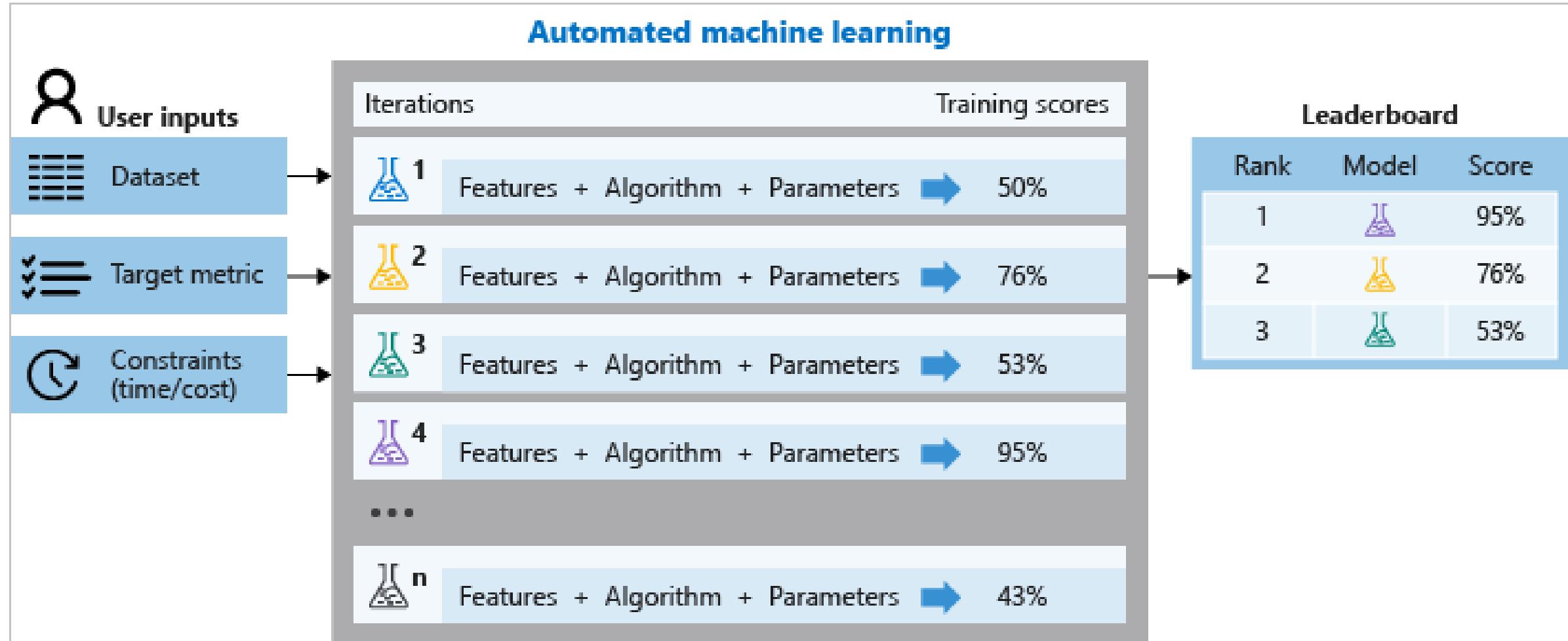
Machine Learning benefits from distributed parallel architectures.

Different algorithms can then be executed in parallel, taking advantage of the different nodes in a cluster.

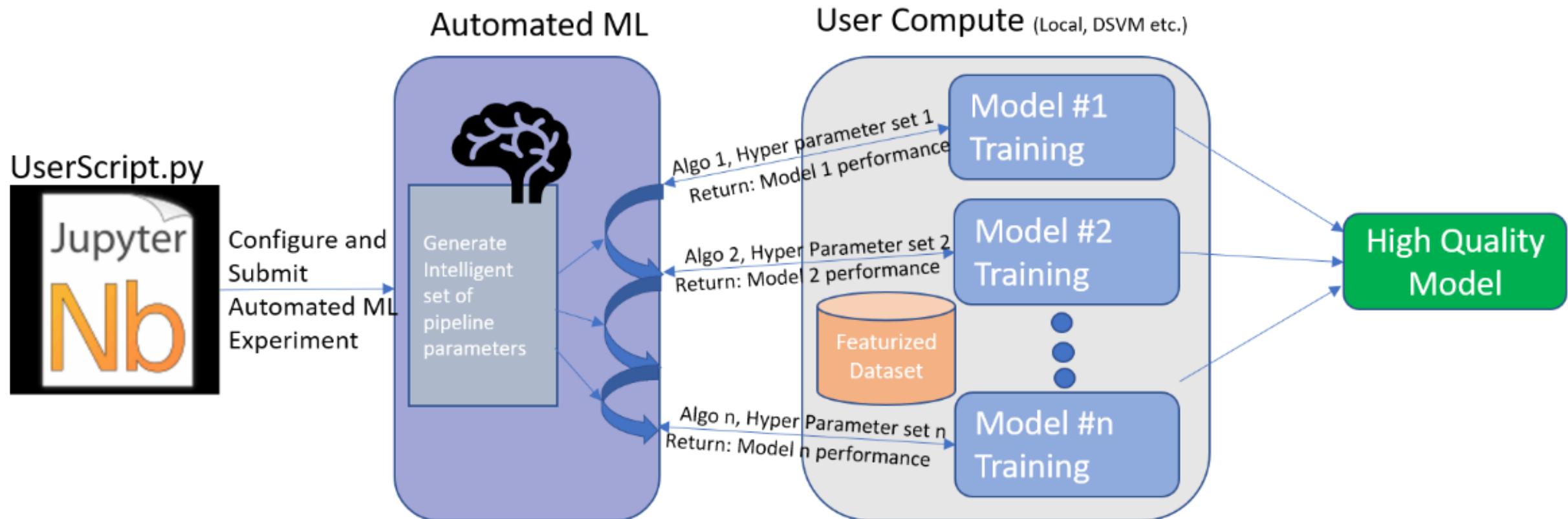
Automated ML : how it works ? (overview)



Run many different algorithms in parallel and rank on metrics



Automated ML : how it works ? (detailed view)





PREMIUM licence



Azure ML SDK Version: 1.10.0

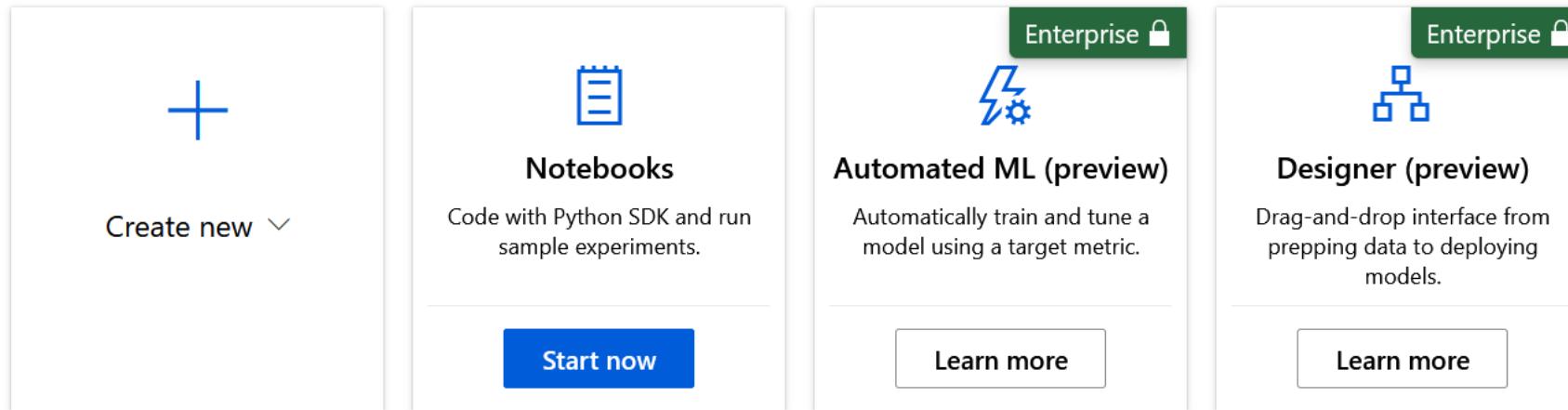
Automated Machine Learning on Microsoft

Two approaches in Azure Machine Learning



« No code / low code » : GUI but need Enterprise licence

Welcome to the studio!



The screenshot shows the 'Welcome to the studio!' screen. It includes a 'Create new' button, a 'Notebooks' section with a 'Start now' button, an 'Automated ML (preview)' section with a 'Learn more' button, and a 'Designer (preview)' section with a 'Learn more' button. Each of the preview sections has an 'Enterprise' lock icon in the top right corner.

- Create new**
- Notebooks**
Code with Python SDK and run sample experiments.
Start now
- Automated ML (preview)**
Enterprise
Automatically train and tune a model using a target metric.
Learn more
- Designer (preview)**
Enterprise
Drag-and-drop interface from prepping data to deploying models.
Learn more

« Code first » : package azureml-sdk (visible code but no contribution possible) and Standard license for storage, monitoring and deployment features.

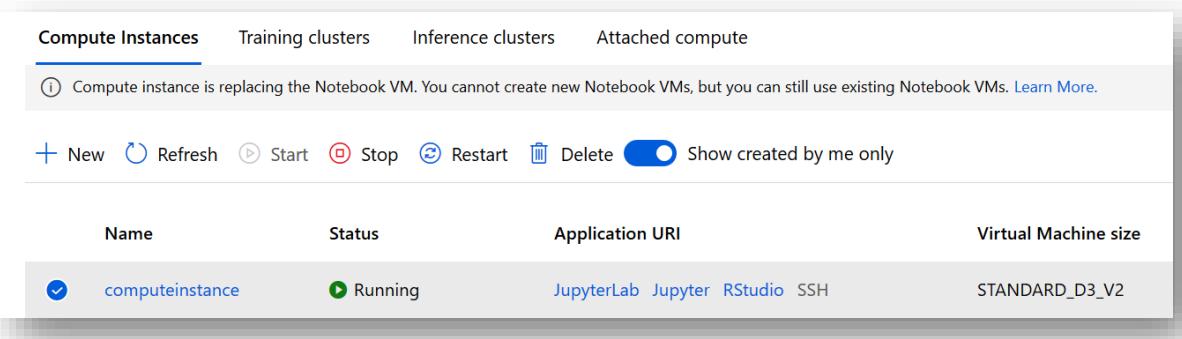
<https://pypi.org/project/azureml-sdk/>

Python Kernel and azureml.core library

From a compute instance :

Launch JupyterLab

Use the preconfigured kernel with
azureml.core



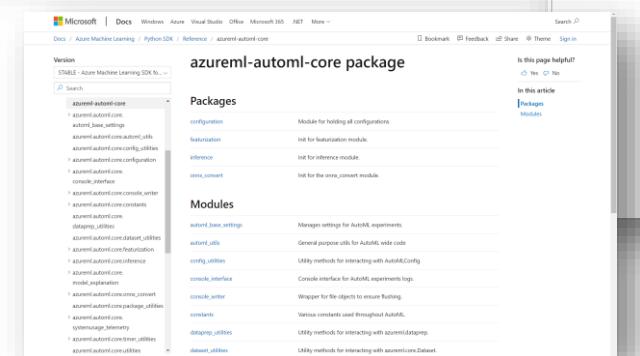
Compute Instances Training clusters Inference clusters Attached compute

Compute instance is replacing the Notebook VM. You cannot create new Notebook VMs, but you can still use existing Notebook VMs. [Learn More.](#)

+ New Refresh Start Stop Restart Delete Show created by me only

Name	Status	Application URI	Virtual Machine size
computeinstance	Running	JupyterLab Jupyter RStudio SSH	STANDARD_D3_V2

Available documentation



Version: Standard - Azure Machine Learning SDK for Python

Search

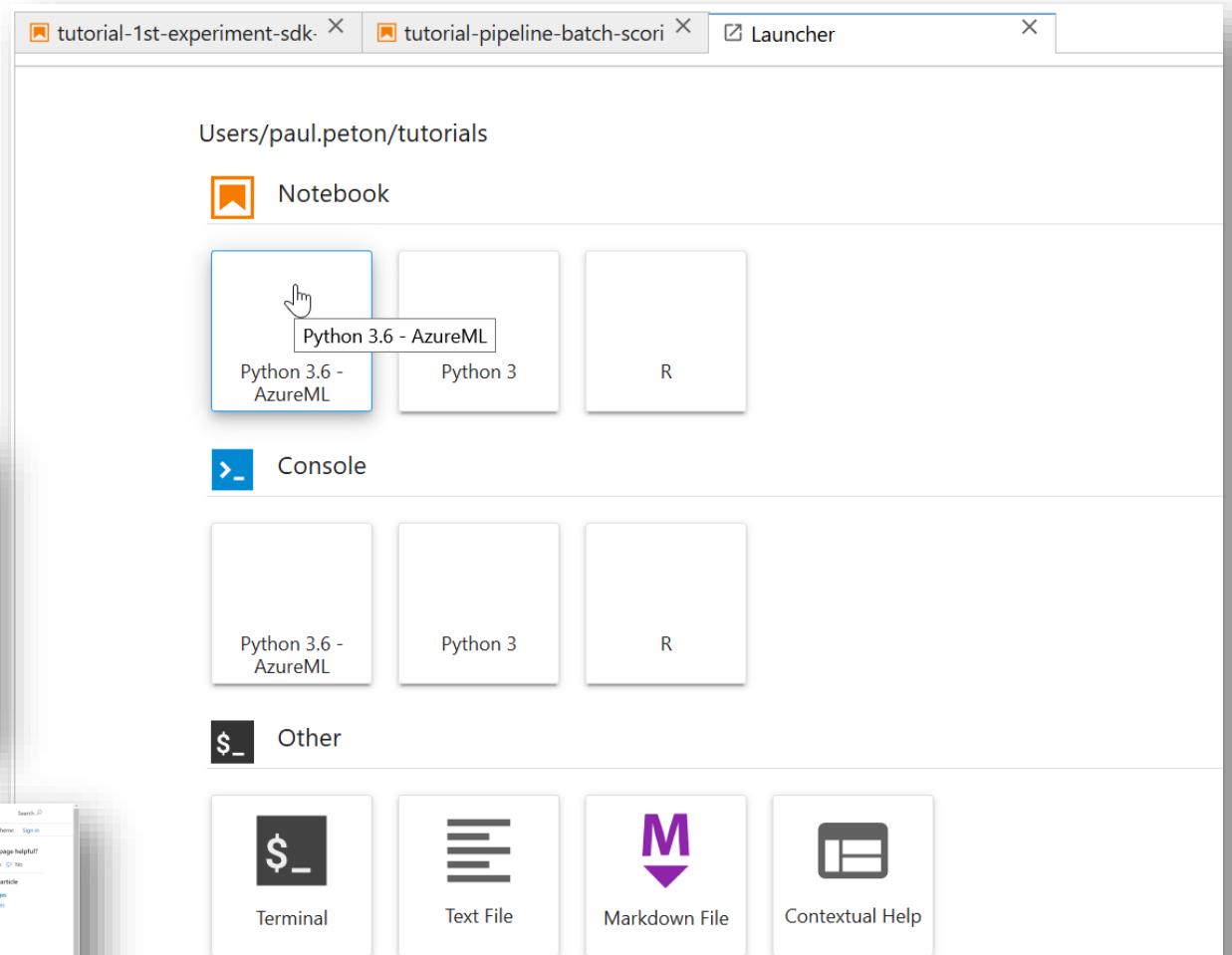
azureml-autml-core package

Packages

- configuration
- featurestore
- inference
- utils

Modules

- autml_base_settings
- autml_utils
- config_utilities
- inference_interface
- console_writer
- constants
- deployment_utilities
- dataset_utilities
- model_deployment
- model_explainability
- model_explainability_package_utilities
- autml_core
- sysmanagement_utilities
- autml_core_trainer_utilities
- autml_core_utilities



Users/paul.peton/tutorials

Notebook

Python 3.6 - AzureML Python 3 R

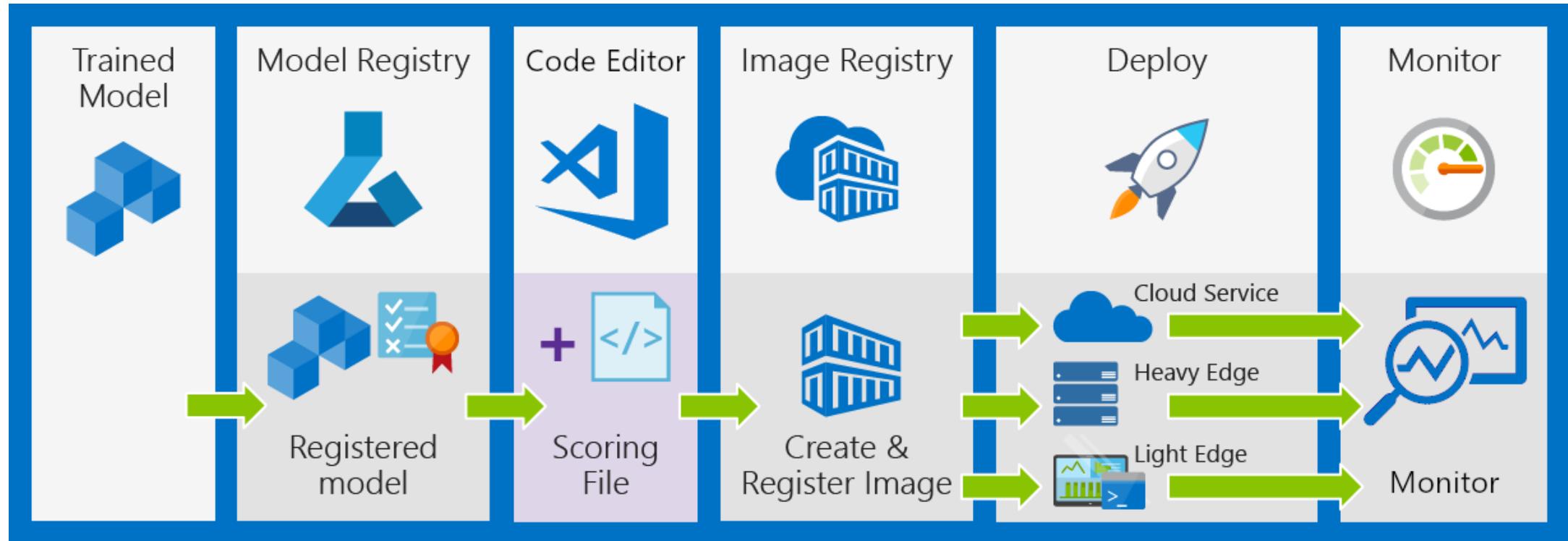
Console

Python 3.6 - AzureML Python 3 R

Other

Terminal Text File Markdown File Contextual Help

Azure Machine Learning Service features

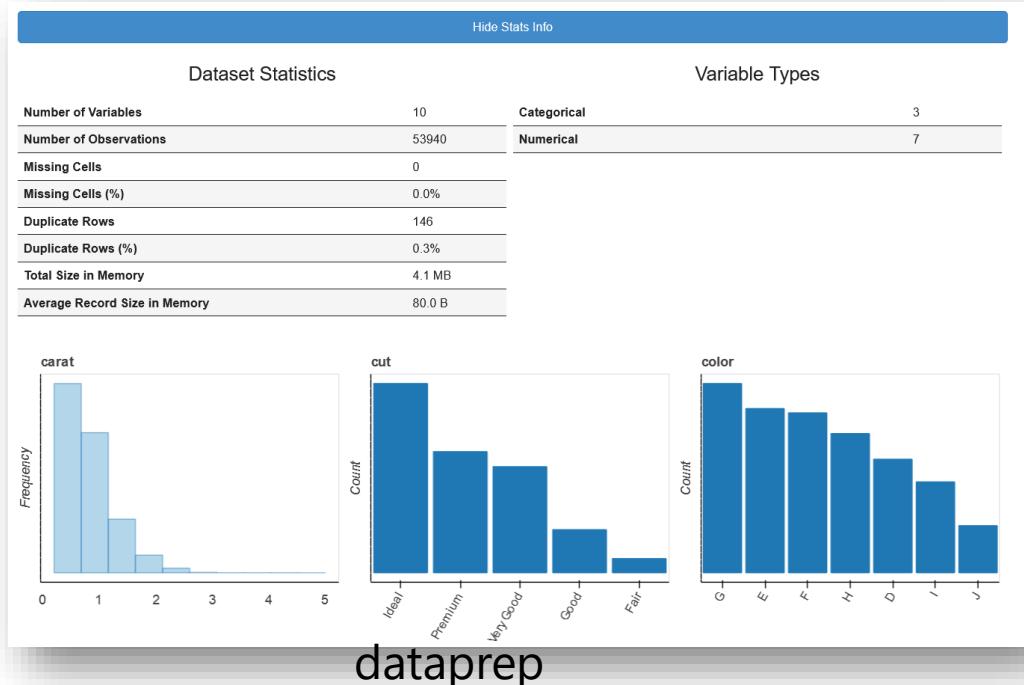


Don't forget Exploratory Data Analysis (EDA)

Use packages like :

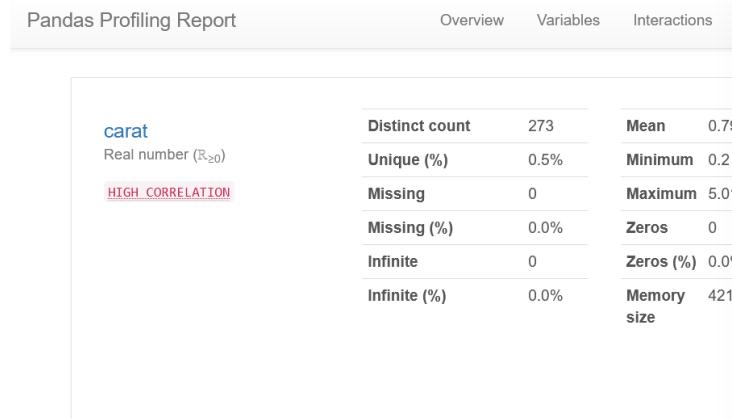
- azureml-dataprep ([doc](#) & [example](#))
- pandas_profiling
- dataprep

```
from dataprep.eda import plot, plot_correlation, plot_missing
plot(df)
```



dataprep

```
import pandas_profiling
import pandas.util.testing as tm
pandas_profiling.ProfileReport(df)
```



pandas_profiling

```
from azureml.core import Dataset
dataset = Dataset.get_by_name(ws, name='Penguins')
df = dataset.to_pandas_dataframe()
```

```
import azureml.dataprep as dprep
dataset.get_profile()
```

ColumnProfile:
 column_name: Culmen Depth (mm)
 type: FieldType.DECIMAL

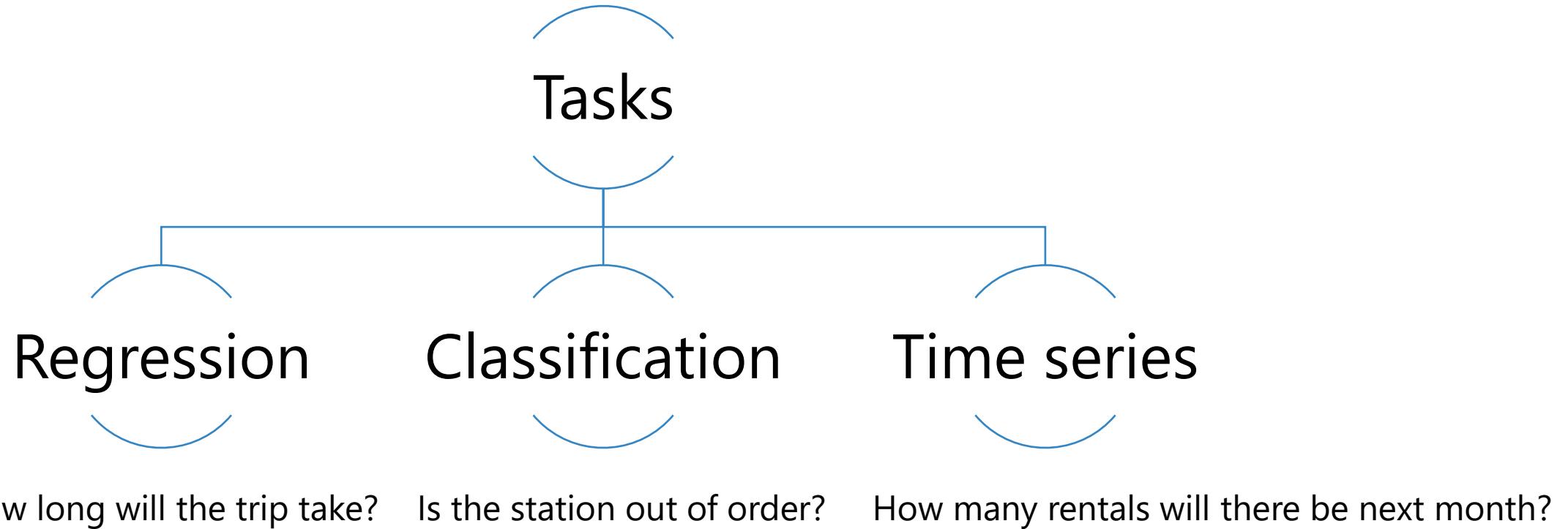
min: 32.1
 max: 59.6
 count: 344.0
 missing_count: 2.0
 not_missing_count: 342.0
 percent_missing: 0.005813953488372093
 error_count: 0.0
 empty_count: 0.0
 unique_values: 165

Quantiles (est.):
 0.1%: 32.1
 1%: 36.57
 5%: 36.5
 25%: 39.28
 50%: 44.38
 75%: 48.5
 95%: 52.0
 99%: 55.808
 99.9%: 59.6

mean: 43.92192982456136
 std: 5.459583713926547
 variance: 29.807054329371987
 skewness: 0.052653027315875206
 kurtosis: -0.8931396557023659
 whisker_top: 59.6
 whisker_bottom: 32.1

azureml-dataprep

Supervised machine learning tasks



Available algorithms (scikit-learn library)



classification ;	régression ;	Prévision de séries chronologiques
Logistic Regression	Elastic Net	Elastic Net
Light GBM	Light GBM	Light GBM
Gradient Boosting	Gradient Boosting	Gradient Boosting
Decision Tree	Decision Tree	Decision Tree
K Nearest Neighbors	K Nearest Neighbors	K Nearest Neighbors
Linear SVC	LARS Lasso	LARS Lasso
Support Vector Classification (SVC)	Stochastic Gradient Descent (SGD)	Stochastic Gradient Descent (SGD)
Random Forest	Random Forest	Random Forest
Extremely Randomized Trees	Extremely Randomized Trees	Extremely Randomized Trees
Xgboost	Xgboost	Xgboost
Classifieur DNN	Régresseur DNN	Régresseur DNN
Classifieur linéaire DNN	Régression linéaire	Régression linéaire
Naive Bayes	FastLinearRegressor	Auto-ARIMA
Stochastic Gradient Descent (SGD)	OnlineGradientDescentRegressor	Prophet
AveragedPerceptronClassifier		ForecastTCN
LinearSVMClassifier		

Data sources and datasets

- Azure file systems
 - Azure Blob Container
 - Azure File Share
 - Azure Data Lake
 - Azure Data Lake Gen2
- Azure managed databases
 - Azure SQL Database
 - Azure Database for PostgreSQL
 - Azure Database for MySQL
- Databricks File System



DEMO 1 – Automated ML GUI

Which dataset will we use ?



Iris Versicolor



Iris Setosa



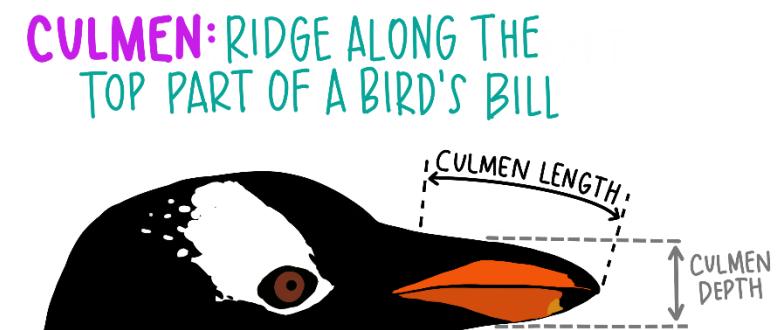
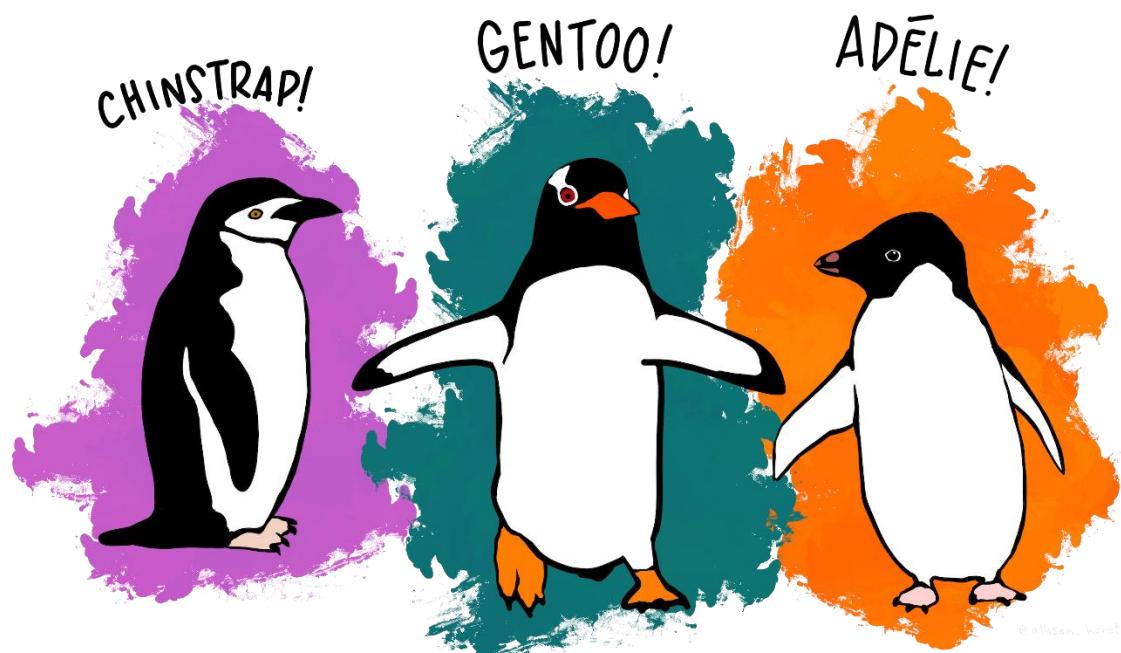
Iris Virginica

Datasets - classification

Palmer penguins

Data were collected and made available by [Dr. Kristen Gorman](#) and the [Palmer Station, Antarctica LTER](#), a member of the [Long Term Ecological Research Network](#).

<https://allisonhorst.github.io/palmerpenguins/>



<https://data-enhanced.com/2020/06/10/two-cheers-for-penguins-data/>

Now, we are playing the game of...

The Seven Common Mistakes with an Automated ML GUI

Launch a new Automated ML run

The screenshot shows the Microsoft Azure Machine Learning studio interface. The top navigation bar includes 'Preview' (orange), 'Microsoft Azure Machine Learning', and various icons for settings, help, and user profile.

The left sidebar has a tree view with 'eacbmlservicews' selected, and 'Automated ML' is highlighted with a red oval. Other items in the tree include 'New', 'Home', 'Author', 'Notebooks', 'Designer', 'Datasets', 'Experiments', 'Pipelines', 'Models', 'Endpoints', 'Compute', 'Datastores', and 'Data Labeling'. A secondary red oval highlights the '+ New Automated ML run' button.

The main content area is titled 'Automated ML' and contains the following text: 'Let Automated ML train and find the best model based on your data without writing a single line of code. [Learn more about Automated ML](#)'. Below this is a table titled 'Recent Automated ML runs' with a 'View all experiments' link. The table lists six runs:

Run	Run ID	Experiment	Status	Created time	Duration	Submitted by	Compute target	Tags
Run 1	AutoML_72237c43-9380-4beb-a...	machineDataA...	Completed	Feb 13, 2020 3:55 PM	1h 19m 1s	Paul PETON	trainingcluster	
Run 1	AutoML_32d65004-de1f-4406-9...	prevision_trott...	Completed	Sep 21, 2019 2:08 PM	1h 1m 59s		myfirstcompute	
Run 1	AutoML_034c1e0b-480e-41e1-a...	prevision_trott...	Completed	Sep 21, 2019 1:39 PM	1h 2m 47s		myfirstcompute	
Run 12	AutoML_707590aa-0240-4c61-8...	trottinettes-for...	Completed	Sep 18, 2019 2:33 PM	2m 55s		local	
Run 1	AutoML_5ff4b5e6-fc94-4b46-9a...	trottinettes-for...	Completed	Sep 18, 2019 2:28 PM	2m 58s		local	

Below this is a 'Documentation' section with three items: 'Concept: What is Automated ML?', 'Tutorial: Create your first classification model with Automated ML', and 'Blog: Build more accurate forecasts with new capabilities in Automated ML'. Each item has a blue icon and a link to its respective documentation page.

Choose training dataset

Preview Microsoft Azure Machine Learning

eacbm1servicews > Automated ML > Start run

Create a new Automated ML run

Select dataset

Select a dataset from the list below, or create a new dataset. Automated ML currently only supports tabular data for authoring runs.

+ Create dataset | Show supported datasets only | Search to filter items...

Dataset type	Created on	Modified
Tabular	Feb 13, 2020 3:40 PM	Feb 13, 2020 3:40 PM
Tabular	Feb 6, 2020 6:47 PM	Feb 6, 2020 6:47 PM

① Customers should not include personal data or other sensitive information in fields marked with because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting. [Learn more](#)

Select Open Dataset

Dataset details

Azure Open Datasets offers ML ready data from the open domain. [Registering](#) open datasets in the workspace lets you easily access open data in your experiments from a common storage location without creating a copy of the data in your storage account.

Register US Consumer Price Index

Name * Dataset version

Back Next Cancel

The screenshot shows the 'Create a new Automated ML run' wizard in the Microsoft Azure Machine Learning studio. The 'Select dataset' step is active, indicated by a red oval around the 'Select dataset' link in the left sidebar. Another red oval highlights the 'Select Open Dataset' link in the 'Dataset details' section of the main content area. The sidebar also features links for 'Configure run' and 'Task type and settings'. The top navigation bar includes 'Preview', 'Microsoft Azure Machine Learning', and various user interface icons.

Create a new dataset

First mistake : use a raw dataset. It's better to use a prepared one.

Create dataset from datastore

X

- Basic info
- Datastore selection
- Settings and preview
- Schema
- Confirm details

Settings and preview

These settings were automatically detected. Please verify that the selections were made correctly or update

File format

Delimited

Delimiter

Example

Comma

Field1,Field2,Field3

Encoding

UTF-8

Column headers

Use headers from the first file

Skip rows

None

studyName	Sample Number	Species	Region
PAL0708	1	Adelie Penguin (Pyg... Anvers	
PAL0708	2	Adelie Penguin (Pyg... Anvers	
PAL0708	3	Adelie Penguin (Pyg... Anvers	
PAL0708	4	Adelie Penguin (Pyg... Anvers	
PAL0708	5	Adelie Penguin (Pyg... Anvers	
PAL0708	6	Adelie Penguin (Pyg... Anvers	
PAL0708	7	Adelie Penguin (Pyg... Anvers	
PAL0708	8	Adelie Penguin (Pyg... Anvers	

Back

Next

Cancel

Include columns and format settings

Second mistake : don't verify each feature type. Be careful with missing values.

Create dataset from datastore

X

Schema

Include	Column name	Properties	Type	Format settings and example
<input checked="" type="checkbox"/>	Path	Not applicable to selected type	String	
<input checked="" type="checkbox"/>	Column1	Not applicable to selected type	Integer	1, 2, 3
<input checked="" type="checkbox"/>	studyName	Not applicable to selected type	String	PAL0708, PAL0708, PAL0708
<input checked="" type="checkbox"/>	Sample Number	Not applicable to selected type	Integer	1, 2, 3
<input checked="" type="checkbox"/>	Species	Not applicable to selected type	String	Adelie Penguin (Pygoscelis adel...
<input checked="" type="checkbox"/>	Region	Not applicable to selected type	String	Anvers, Anvers, Anvers
<input checked="" type="checkbox"/>	Island	Not applicable to selected type	String	Torgersen, Torgersen, Torgersen
<input checked="" type="checkbox"/>	Stage	Not applicable to selected type	String	Adult, 1 Egg Stage, Adult, 1 Egg...
<input checked="" type="checkbox"/>	Individual ID	Not applicable to selected type	String	N1A1, N1A2, N2A1
<input checked="" type="checkbox"/>	Clutch Completion	Not applicable to selected type	Boolean	true, true, true
<input checked="" type="checkbox"/>	Date Egg	Timestamp	Date	2007-11-11 00:00:00, 2007-11...
<input checked="" type="checkbox"/>	Culmen Length (mm)	Not applicable to selected type	Decimal	39.1, 39.5, 40.3
<input checked="" type="checkbox"/>	Culmen Depth (mm)	Not applicable to selected type	Decimal	18.7, 17.4, 18
<input checked="" type="checkbox"/>	Flipper Length (mm)	Not applicable to selected type	Decimal	181, 186, 195
<input checked="" type="checkbox"/>	Body Mass (g)	Not applicable to selected type	Decimal	3750, 3800, 3250

Back Next Cancel

Confirm details and profile this dataset

Create dataset from datastore

X

Confirm details

Basic info

Name	Penguins
Dataset version	1
Dataset type	Tabular
Description	Raw version

Datastore selection

Datastore	sandbox_penguins
Path	penguins_raw.csv

File settings

File format	Delimited
Delimiter	Comma
Encoding	UTF-8
Column headers	Use headers from the first file
Skip rows	None

Profile this dataset after creation

Select compute for profiling

myComputeCluster

Refresh

Back Create Cancel

Third mistake : forgot to explore the dataset

Profile : some help for Exploratory Data Analysis

Microsoft Azure Machine Learning

eacbmilservicews > Datasets > Penguins

Penguins Version 2 (latest)

Details Consume Explore Models

Refresh Generate profile + Create dataset monitor ⚡ Unregister New version

Preview Profile

Number of columns: 18 Number of rows: 344

Column	Profile	Type	Min	Max	Count	Missing count	Empty count	Ent
Culmen Depth (mm)		Decimal	32.10	59.60	344	2	0	0
Flipper Length (mm)		Decimal	13.10	21.50	344	2	0	0
Body Mass (g)		Integer	172	231	344	2	0	0
Sex		Integer	2700	6300	344	2	0	0

Select dataset

Microsoft Azure Machine Learning

eacbmlservicews > Automated ML (preview) > Start run

Success: Penguins dataset created successfully

Create a new Automated ML run

Select dataset

Select a dataset from the list below, or create a new dataset. Automated ML currently only supports tabular data for authoring runs.

+ Create dataset | Show supported datasets only

Search to filter items...

Dataset name	Dataset type	Created on	Modified
Penguins	Tabular	Jul 24, 2020 10:46 AM	Jul 24, 2020 10:46 AM

Automated ML (preview)

New

Home

Author

Notebooks

Automated ML (preview)

Designer (preview)

Assets

Datasets

Experiments

Pipelines

Choose the target

Preview Microsoft Azure Machine Learning

eacbmlservicews > Automated ML > Start run

Success: USConsumerPriceIndex dataset created successfully

Create a new Automated ML run

① Customers should not include personal data or other sensitive information in fields marked with ⓘ because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting. [Learn more](#)

Configure run

Configure the experiment. Select from existing experiments or define a new name, select the target column and the training compute to use. [Learn more on how to configure the experiment](#)

Dataset
USConsumerPriceIndex ([View dataset](#))

Experiment name *
USCPI-regression

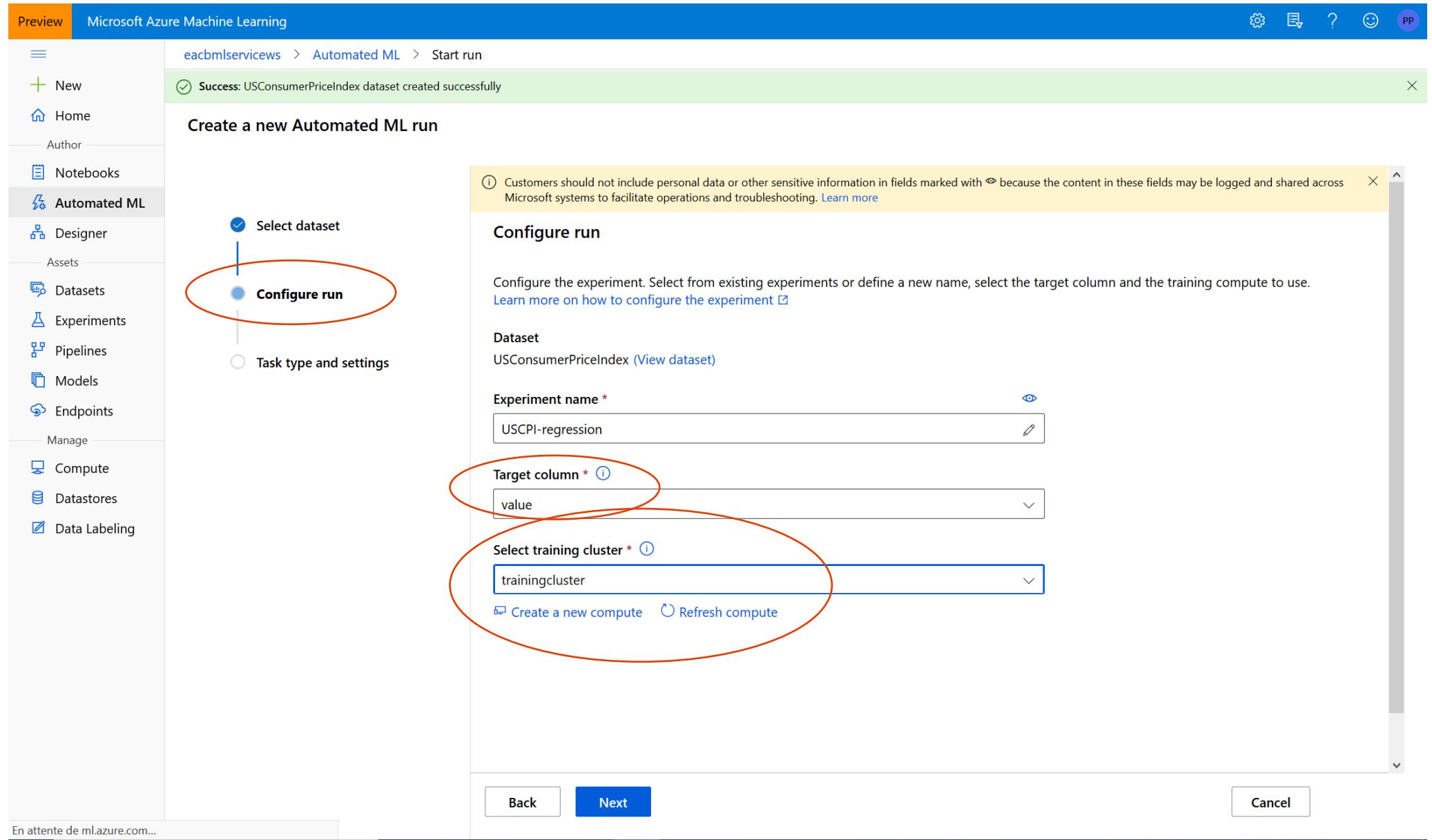
Target column * ⓘ
value

Select training cluster * ⓘ
trainingcluster

[Create a new compute](#) [Refresh compute](#)

Back Next Cancel

En attente de ml.azure.com...



Select task type (classification, regression, time series forecast)

Microsoft Azure Machine Learning

eacbmlservicews > Automated ML (preview) > Start run

Create a new Automated ML run

Select dataset

Configure run

Task type and settings

Select task type

Select the machine learning task type for the experiment. Additional settings are available to fine tune the experiment if needed.

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

Enable deep learning ⓘ

Regression
To predict continuous numeric values

Time series forecasting
To predict values based on time

[View additional configuration settings](#) [View featurization settings](#)

Back Finish Cancel

Fourth mistake : use deep learning for tabular dataset (keep it for unstructured datas)

The screenshot shows the Microsoft Azure Machine Learning interface for creating a new Automated ML run. The left sidebar shows navigation options like 'New', 'Home', 'Author', 'Notebooks', 'Automated ML (preview)', 'Designer (preview)', 'Assets', 'Datasets', 'Experiments', 'Pipelines', 'Models', 'Endpoints', 'Manage', 'Compute', 'Datastores', and 'Data Labeling'. The 'Automated ML (preview)' option is currently selected. The main area is titled 'Create a new Automated ML run' and shows a flow: 'Select dataset' (completed), 'Configure run' (completed), and 'Task type and settings' (in progress). The 'Task type and settings' step is titled 'Select task type' with the instruction: 'Select the machine learning task type for the experiment. Additional settings are available to fine tune the experiment if needed.' It lists three options: 'Classification' (selected, indicated by a green checkmark), 'Regression' (unchecked), and 'Time series forecasting' (unchecked). A red oval highlights the 'Enable deep learning' checkbox under 'Classification', which is described as 'To predict one of several categories in the target column. yes/no, blue, red, green.' To the right of this section, a callout box with a red background contains the text: 'Fourth mistake : use deep learning for tabular dataset (keep it for unstructured datas)'. At the bottom of the screen are 'Back', 'Finish', and 'Cancel' buttons.

Métrique d'évaluation (classification)

Fifth mistake (bis) : choose accuracy for imbalanced dataset

Additional configurations X

Primary metric ⓘ AUC weighted

Explain best model ⓘ

Blocked algorithms ⓘ
A list of algorithms that Automated ML will not use during training.

Exit criterion

Training job time (hours) ⓘ

Metric score threshold ⓘ

Validation

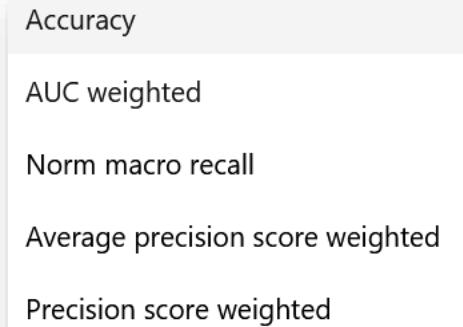
Validation type ⓘ Train-validation split

Percentage validation of data * ⓘ ▲
Automated ML recommends the percentage validation of data to be between 10 and 30

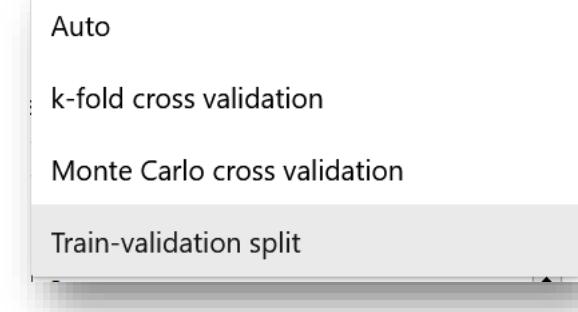
Concurrency

Max concurrent iterations ⓘ ▼

Save



Sixth mistake : choose train-validation on too small dataset , prefer k-fold CV



Featurization

Select features

Select feature type

Select imputation method

Featurization

Feature selection identifies the actions performed on the dataset 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

Column name	Included	Feature type	Impute with	Data example
studyName	<input checked="" type="checkbox"/>	Auto	Auto	PAL0708, PAL0708, PAL0708
Sample Number	<input checked="" type="checkbox"/>	Auto	Auto	1, 2, 3
Species (Target column)	<input checked="" type="checkbox"/>	Auto	Auto	Adelie Penguin (Pygoscelis adeliae... Gentoo Penguin (Pygoscelis papua) Chinstrap Penguin (Pygoscelis antarcticus)
Region	<input checked="" type="checkbox"/>	Auto	Auto	Torgersen
Island	<input checked="" type="checkbox"/>	Auto	Auto	Torgersen
Stage	<input checked="" type="checkbox"/>	Auto	Auto	Adult, 1 Egg Stage, Adult, 1 Egg St...
Individual ID	<input checked="" type="checkbox"/>	Auto	Auto	N1A1, N1A2, N2A1
Clutch Completion	<input checked="" type="checkbox"/>	Categorical	Auto	true, true, true
Date Egg	<input checked="" type="checkbox"/>	DateTime	Auto	2007-11-11 00:00:00, 2007-11-11 ...
Culmen Length (mm)	<input checked="" type="checkbox"/>	Numeric	Auto	39.1, 39.5, 40.3
Culmen Depth (mm)	<input checked="" type="checkbox"/>	Numeric	Auto	18.7, 17.4, 18
Flipper Length (mm)	<input checked="" type="checkbox"/>	Numeric	Auto	181, 186, 195
Body Mass (g)	<input checked="" type="checkbox"/>	Numeric	Auto	3750, 3800, 3250
Sex	<input checked="" type="checkbox"/>	Categorical	Auto	MALE, FEMALE, FEMALE
Delta 15 N (o/oo)	<input checked="" type="checkbox"/>	Numeric	Median	Null, 8.94956, 8.36821
Delta 13 C (o/oo)	<input checked="" type="checkbox"/>	Numeric	Median	Null, -24.69454, -25.33302
Comments	<input checked="" type="checkbox"/>	Auto	Auto	Not enough blood for isotopes, n...

Seventh mistake : use all columns. Avoid Id, names or too much correlated features.

Save

Cancel

The « best » model according to the primary metric

Preview Microsoft Azure Machine Learning

eacbmllservices > Experiments > diamondsautoml > Run 1

Run 1 Completed

Refresh Cancel

Switch to old experience ?

Details Data guardrails Models Logs Outputs

Recommended model

Model name
MaxAbsScaler, GradientBoosting

Metric value
0.976001738274974

Started on
Feb 29, 2020 11:51 AM

Duration
4m 41s

Sdk version
1.1.1rc0

Deploy status
No deployment yet

Run summary

Task type
Regression ([View all run settings](#))

Primary metric
R2 score

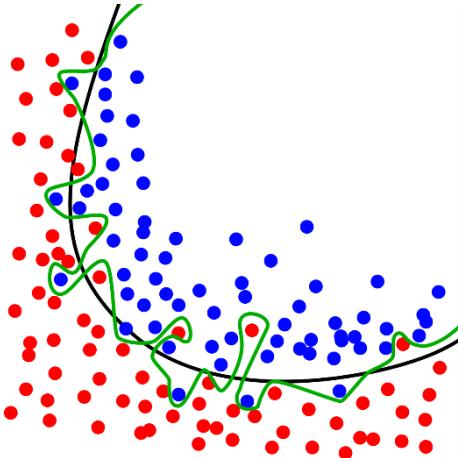
Run status
Completed

Experiment name
diamondsautoml

Run ID
AutoML_631682e5-f044-4a54-ad53-ea0a435526ff

Deploy best model View model details Download best model

Prevent over-fitting



Model	Train accuracy	Test accuracy
A	99.9%	95%
B	87%	87%
C	99.9%	45%

The best way to prevent over-fitting is to follow ML best-practices including:

- Using more training data
- Preventing target leakage
- Using fewer features
- Regularization and hyperparameter optimization
- Model complexity limitations
- Cross-validation

Training and cross validation

Training data size	Validation technique
Larger than 20,000 rows	Train/validation data split is applied. The default is to take 10% of the initial training data set as the validation set. In turn, that validation set is used for metrics calculation.
Smaller than 20,000 rows	Cross-validation approach is applied. The default number of folds depends on the number of rows. If the dataset is less than 1,000 rows , 10 folds are used. If the rows are between 1,000 and 20,000 , then three folds are used.

AutoMLConfig parameters :

- training_data
- validation_data
- validation_size
- n_cross_validations
- cv_split_column_names

Featurization (automatic or customize)

Feature engineering is the process of using domain knowledge of the data to create features that help machine learning algorithms to learn better.

In Azure Machine Learning

- data-scaling and
- normalization techniques

are applied to make feature engineering easier.

Collectively, these techniques and this feature engineering are called *featurization* in automated machine learning experiments.

Featurization steps	Description
Drop high cardinality or no variance features*	Drop these features from training and validation sets. Applies to features with all values missing, with the same value across all rows, or with high cardinality (for example, hashes, IDs, or GUIDs).
Impute missing values*	For numeric features, impute with the average of values in the column. For categorical features, impute with the most frequent value.
Generate additional features*	For DateTime features: Year, Month, Day, Day of week, Day of year, Quarter, Week of the year, Hour, Minute, Second. For Text features: Term frequency based on unigrams, bigrams, and trigrams. Learn more about how this is done with BERT .
Transform and encode*	Transform numeric features that have few unique values into categorical features. One-hot encoding is used for low-cardinality categorical features. One-hot-hash encoding is used for high-cardinality categorical features.
Word embeddings	A text featurizer converts vectors of text tokens into sentence vectors by using a pretrained model. Each word's embedding vector in a document is aggregated with the rest to produce a document feature vector.
Target encodings	For categorical features, this step maps each category with an averaged target value for regression problems, and to the class probability for each class for classification problems. Frequency-based weighting and k-fold cross-validation are applied to reduce overfitting of the mapping and noise caused by sparse data categories.
Text target encoding	For text input, a stacked linear model with bag-of-words is used to generate the probability of each class.
Weight of Evidence (WoE)	Calculates WoE as a measure of correlation of categorical columns to the target column. WoE is calculated as the log of the ratio of in-class vs. out-of-class probabilities. This step produces one numeric feature column per class and removes the need to explicitly impute missing values and outlier treatment.
Cluster Distance	Trains a k-means clustering model on all numeric columns. Produces k new features (one new numeric feature per cluster) that contain the distance of each sample to the centroid of each cluster.

AutoMLConfig parameters

Property	Description
<code>task</code>	forecasting
<code>primary_metric</code>	This is the metric that you want to optimize. Forecasting supports the following primary metrics <i>spearman_correlation</i> <i>normalized_root_mean_squared_error</i> <i>r2_score</i> <i>normalized_mean_absolute_error</i>
<code>experiment_timeout_hours</code>	Experimentation timeout in hours.
<code>enable_early_stopping</code>	If early stopping is on, training will stop when the primary metric is no longer improving.
<code>training_data</code>	Input dataset, containing both features and label column.
<code>label_column_name</code>	The name of the label column.
<code>compute_target</code>	The remote compute for training.
<code>n_cross_validations</code>	Number of cross-validation folds to use for model/pipeline selection
<code>enable_voting_ensemble</code>	Allow AutoML to create a Voting ensemble of the best performing models
<code>enable_stack_ensemble</code>	Allow AutoML to create a Stack ensemble of the best performing models
<code>debug_log</code>	Log file path for writing debugging information
<code>time_column_name</code>	Name of the datetime column in the input data
<code>grain_column_names</code>	Name(s) of the columns defining individual series in the input data
<code>max_horizon</code>	Maximum desired forecast horizon in units of time-series frequency
<code>featurization</code>	'auto' / 'off' / FeaturizationConfig Indicator for whether featurization step should be done automatically or not, or whether customized featurization should be used. Setting this enables AutoML to perform featurization on the input to handle <i>missing data</i> , and to perform some common <i>feature extraction</i> .
<code>max_cores_per_iteration</code>	Maximum number of cores to utilize per iteration. A value of -1 indicates all available cores should be used.

DEMO 2 – Python SDK

Automated ML with azureml SDK

```
[ ]: !pip install --upgrade azureml-sdk
[ ]: !pip install --upgrade azureml-widgets
[ ]: !pip install --upgrade azureml-train-automl-runtime
[1]: import pandas as pd
import numpy as np
import azureml.core
[2]: from azureml.core import Workspace
ws = Workspace.from_config()
[3]: from azureml.core import Experiment
experiment_name = 'penguins-automatedML'
experiment = Experiment(workspace=ws, name=experiment_name)
...
[14]: from azureml.core import Dataset
dataset = Dataset.get_by_name(ws, name='Penguins').drop_columns(['Individual ID', 'studyName', 'Sample Number', 'Stage', 'Region', 'Island', 'Comments'])
dataset.take(3).to_pandas_dataframe()
```

	Species	Clutch Completion	Date Egg	Culmen Length (mm)	Culmen Depth (mm)	Flipper Length (mm)	Body Mass (g)	Sex	Delta 15 N (o/oo)	Delta 13 C (o/oo)
0	Adelie Penguin (Pygoscelis adeliae)	True	2007-11-11	39.10	18.70	181	3750	MALE	nan	nan
1	Adelie Penguin (Pygoscelis adeliae)	True	2007-11-11	39.50	17.40	186	3800	FEMALE	8.95	-24.69
2	Adelie Penguin (Pygoscelis adeliae)	True	2007-11-16	40.30	18.00	195	3250	FEMALE	8.37	-25.33

Automated ML with azureml SDK

```
[21]: import logging

automl_settings = {
    "iteration_timeout_minutes": 1,
    "iterations": 40,
    "primary_metric": 'norm_macro_recall', #precision_score_weighted, average_precision_score_weighted
    "featurization": 'auto', #off, auto, FeaturizationConfig
    "verbosity": logging.INFO,
    "enable_local_managed": True # to submit a local conda or local docker run
}

[22]: #https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig.automlconfig?view=azure-ml-py
from azureml.train.automl import AutoMLConfig

automl_config = AutoMLConfig(task='classification',
                             debug_log='automated_ml_errors.log',
                             #compute_target = compute_target,
                             training_data=train_data,
                             validation_data=test_data,
                             label_column_name=label,
                             blocked_models=['ExtremeRandomTrees', 'SVM'], # https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train
                             **automl_settings)

featurization_config = FeaturizationConfig()
featurization_config.blocked_transformers = ['LabelEncoder']
featurization_config.drop_columns = ['aspiration', 'stroke']
featurization_config.add_column_purpose('engine-size', 'Numeric')
featurization_config.add_column_purpose('body-style', 'CategoricalHash')
#default strategy mean, add transformer param for for 3 columns
featurization_config.add_transformer_params('Imputer', ['engine-size'], {"strategy": "median"})
featurization_config.add_transformer_params('Imputer', ['city-mpg'], {"strategy": "median"})
featurization_config.add_transformer_params('Imputer', ['bore'], {"strategy": "most_frequent"})
featurization_config.add_transformer_params('HashOneHotEncoder', [], {"number_of_bits": 3})
```

Featurization in SDK needs the enterprise licence of Azure Machine Learning.

Local vs remote target (only with the SDSK)

Local compute: Training occurs on your local laptop or VM compute.

Remote compute: Training occurs on Machine Learning compute clusters.

```
import logging

automl_settings = {
    "iteration_timeout_minutes": 1,
    "iterations": 10,
    "primary_metric": 'norm_macro_recall', #precision_score_weighted, average_precision_score_weighted
    "featurization": featurization_config, #off, auto, FeaturizationConfig
    "verbosity": logging.INFO,
    "enable_local_managed": True # to submit a local conda or local docker run
}
```

```
#https://docs.microsoft.com/en-us/python/api/azureml-train-automl-client/azureml.train.automl.automlconfig
from azureml.train.automl import AutoMLConfig

automl_config = AutoMLConfig(task='classification',
                             debug_log='automated_ml_errors.log',
                             #compute_target = compute_target,
                             training_data=train_data,
                             validation_data=test_data,
                             label_column_name=label,
                             blocked_models=['ExtremeRandomTrees', 'SVM'], # https://docs.microsoft.com/en-
**automl_settings)
```

	Pros (Advantages)	Cons (Handicaps)
Local compute target	<ul style="list-style-type: none"> No environment start up time 	<ul style="list-style-type: none"> Subset of features Can't parallelize runs Worse for large data. No data streaming while training No DNN-based featurization Python SDK only
Remote ML compute clusters	<ul style="list-style-type: none"> Full set of features Parallelize child runs Large data support DNN-based featurization Dynamic scalability of compute cluster on demand No-code experience (web UI) also available 	<ul style="list-style-type: none"> Start up time for cluster nodes Start up time for each child run

Automated ML with azureml SDK

```
[23]: import azureml.train.automl.runtime
local_run = experiment.submit(automl_config, show_output=True)

Running on local machine
Parent Run ID: AutoML_36f1217c-24f4-42e8-b5da-4b502d13b222

Current status: DatasetEvaluation. Gathering dataset statistics.
Current status: FeaturesGeneration. Generating features for the dataset.
Current status: DatasetFeaturization. Beginning to fit featurizers and featurize the dataset.
Current status: DatasetFeaturizationCompleted. Completed fit featurizers and featurizing the dataset.

*****
DATA GUARDRAILS:

TYPE: Class balancing detection
STATUS: PASSED
DESCRIPTION: Your inputs were analyzed, and all classes are balanced in your training data.
Learn more about imbalanced data: https://aka.ms/AutomatedMLImbalancedData

*****
TYPE: Missing feature values imputation
STATUS: DONE
DESCRIPTION: If the missing values are expected, let the run complete. Otherwise cancel the current run and use a script to customize the handling of missing feature values that may be more appropriate based on the data type and business requirement.
Learn more about missing value imputation: https://aka.ms/AutomatedMLFeaturization

DETAILS:
+-----+-----+-----+
|Column name |Missing value count |Imputation type |
+=====+=====+=====+
|Culmen Length (mm) |1 |mean |
|Culmen Depth (mm) |1 |mean |
|Flipper Length (mm) |1 |mean |
|Body Mass (g) |1 |mean |
|Sex |8 |
|Delta 15 N (o/oo) |10 |mean |
|Delta 13 C (o/oo) |9 |mean |
+-----+-----+-----+
*****
TYPE: High cardinality feature detection
STATUS: PASSED
DESCRIPTION: Your inputs were analyzed, and no high cardinality features were detected.
Learn more about high cardinality feature handling: https://aka.ms/AutomatedMLFeaturization

*****
Current status: ModelSelection. Beginning model selection.
```

```
*****
Current status: ModelSelection. Beginning model selection.

*****
ITERATION: The iteration being evaluated.
PIPELINE: A summary description of the pipeline being evaluated.
DURATION: Time taken for the current iteration.
METRIC: The result of computing score on the fitted pipeline.
BEST: The best observed score thus far.

*****
```

ITERATION	PIPELINE	DURATION	METRIC	BEST
0	MaxAbsScaler LightGBM	0:00:17	0.9800	0.9800
1	StandardScalerWrapper SGD	0:00:16	0.9800	0.9800
2	StandardScalerWrapper RandomForest	0:00:17	0.9522	0.9800
3	MinMaxScaler SGD	0:00:16	0.9444	0.9800
4	StandardScalerWrapper SGD	0:00:17	0.9800	0.9800
5	MinMaxScaler RandomForest	0:00:17	0.9800	0.9800
6	MinMaxScaler SGD	0:00:18	1.0000	1.0000
7	StandardScalerWrapper SGD	0:00:17	0.9800	1.0000
8	StandardScalerWrapper SGD	0:00:12	0.9800	1.0000
9	MinMaxScaler RandomForest	0:00:18	0.9800	1.0000
10	StandardScalerWrapper SGD	0:00:12	0.9600	1.0000

```
[25]: local_run.wait_for_completion()
```

```
...
```

```
[26]: best_run, fitted_model = local_run.get_output()
```

```
print(best_run)  
print(fitted_model)
```

```
Run(Experiment: penguins-automatedML,  
Id: AutoML_36f1217c-24f4-42e8-b5da-4b502d13b222_6,  
Type: None,  
Status: Completed)  
Pipeline(memory=None,  
    steps=[('datatransformer', DataTransformer(enable_dnn=None, enable_feature_sweeping=None,  
        feature_sweeping_config=None, feature_sweeping_timeout=None,  
        featurization_config=None, force_text_dnn=None,  
        is_cross_validation=None, is_onnx_compatible=None, logger=None,  
        obser... n_jobs=1, penalty='none', power_t=0.2222222222222222,  
        random_state=None, tol=0.001))])  
Y_transformer(['LabelEncoder', LabelEncoder()])
```

```
[27]: print(best_run.get_file_names())
```

```
['accuracy_table', 'confusion_matrix', 'explanation/b2f7213f/classes.interpret.json', 'explanation/b2f7213f/eval_data_viz.interpret.json', 'explanation/b2f7213f/expected_values.interpret.json', 'explanation/b2f7213f/features.interpret.json', 'explanation/b2f7213f/global_names/0.interpret.json', 'explanation/b2f7213f/global_rank/0.interpret.json', 'explanation/b2f7213f/global_values/0.interpret.json', 'explanation/b2f7213f/local_importance_values.interpret.json', 'explanation/b2f7213f/per_class_names/0.interpret.json', 'explanation/b2f7213f/per_class_rank/0.interpret.json', 'explanation/b2f7213f/per_class_values/0.interpret.json', 'explanation/b2f7213f/rich_metadata.interpret.json', 'explanation/b2f7213f/visualization_dict.interpret.json', 'explanation/b2f7213f/ys_pred_proba_viz.interpret.json', 'explanation/b2f7213f/ys_pred_viz.interpret.json', 'outputs/conda_env_v_1_0_0.yml', 'outputs/env_dependencies.json', 'outputs/model.pkl', 'outputs/pipeline_graph.json', 'outputs/scoring_file_v_1_0_0.py']
```

```
[25]: # register best model  
from azureml.core.model import Model
```

```
model = best_run.register_model(model_name='penguins_class_best_model', model_path='outputs/model.pkl')
```

```
print(model.name, model.version, sep='\t')
```

Widget in Jupyter notebooks

```
[24]: from azureml.widgets import RunDetails  
  
RunDetails(local_run).show()
```

AutoML_36f1217c-24f4-42e8-b5da-4b502d13b222:

Status: Completed

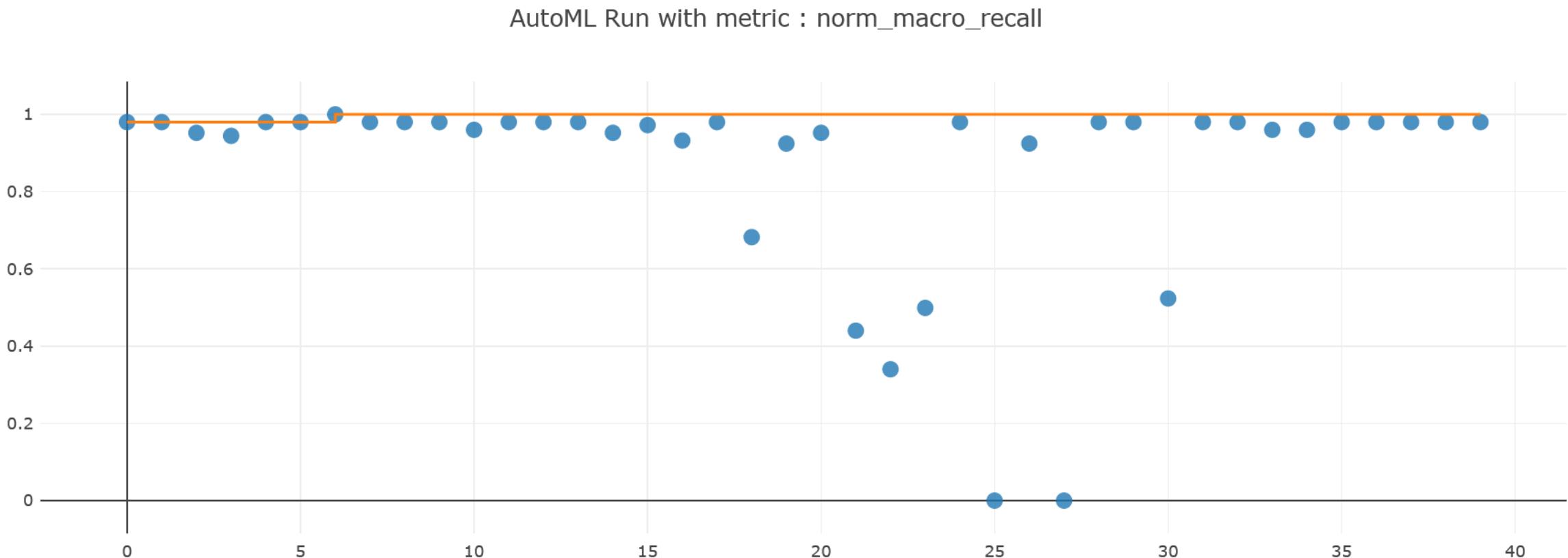


Iteration	Pipeline	Iteration metric	Best metric	Status	Duration	Started	Run Id
6	MinMaxScaler, SGD	1	1	Completed	0:00:17	Jul 28, 2020 7:46 AM	
0	MaxAbsScaler, LightGBM	0.98	0.98	Completed	0:00:16	Jul 28, 2020 7:44 AM	
1	StandardScalerWrapper, SGD	0.98	0.98	Completed	0:00:15	Jul 28, 2020 7:44 AM	
4	StandardScalerWrapper, SGD	0.98	0.98	Completed	0:00:16	Jul 28, 2020 7:45 AM	
5	MinMaxScaler, RandomForest	0.98	0.98	Completed	0:00:16	Jul 28, 2020 7:45 AM	

Pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [Next](#) [Last](#) per page



Widget in Jupyter notebooks



[Click here to see the run in Azure Machine Learning studio](#)

DEMO 3 – Explanation and deployment

Explanations (dataset exploration)

Microsoft Azure Machine Learning

eacbmilservice > Experiments > penguins-automatedML > Run 205 > Run 212

Run 212 Completed

Refresh Deploy Download Explain model Cancel

Details Model Explanations (preview) Metrics Outputs + logs Images Child runs Snapshot

Model explanations are used to understand what features are directly impacting the model and why. [Learn more](#)

Select Explanation

tabular | mimic.lightgbm | engineered | classification | b2f7213f-cf1b-4380-a510-570e0a24d9fe | 7/28/2020, 7:58:35 AM

Explainer: mimic.lightgbm

Dataset Exploration

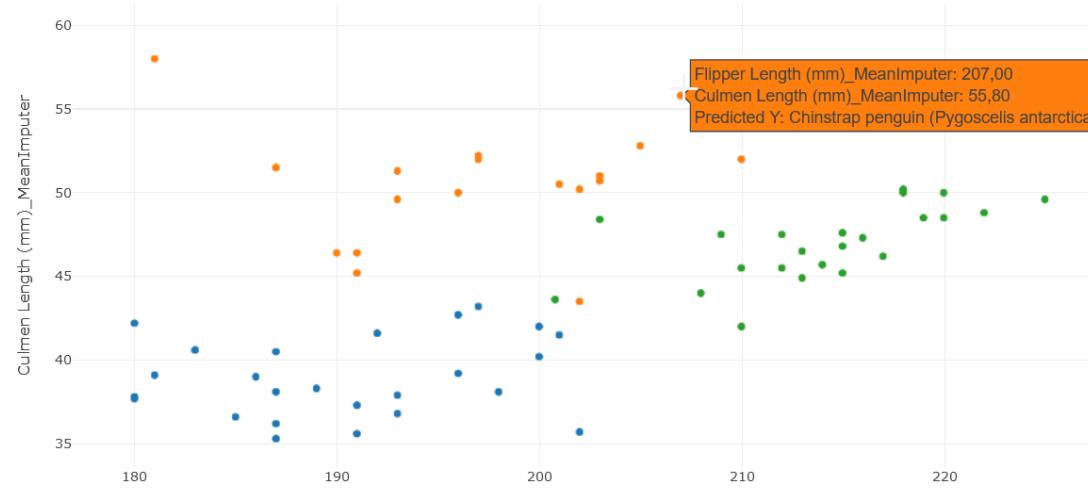
Global Importance

Explanation Exploration

Summary Importance

X value:

Color:



Flipper Length (mm)_MeanImputer: 207,00
Culmen Length (mm)_MeanImputer: 55,80
Predicted Y: Chinstrap penguin (Pygoscelis antarctica)

Explanations (global importance)

Microsoft Azure Machine Learning

eacbmilservicews > Experiments > penguins-automatedML > Run 205 > Run 212

Run 212 Completed

Refresh Deploy Download Explain model Cancel

Details Model Explanations (preview) Metrics Outputs + logs Images Child runs Snapshot

Model explanations are used to understand what features are directly impacting the model and why. [Learn more](#)

Select Explanation

tabular | mimic.lightgbm | engineered | classification | b2f7213f-cf1b-4380-a510-570e0a24d9fe | 7/28/2020, 7:58:35 AM

Explainer: mimic.lightgbm

Dataset Exploration

Top K Features: ⓘ

Sort by: Absolute global

Global Importance

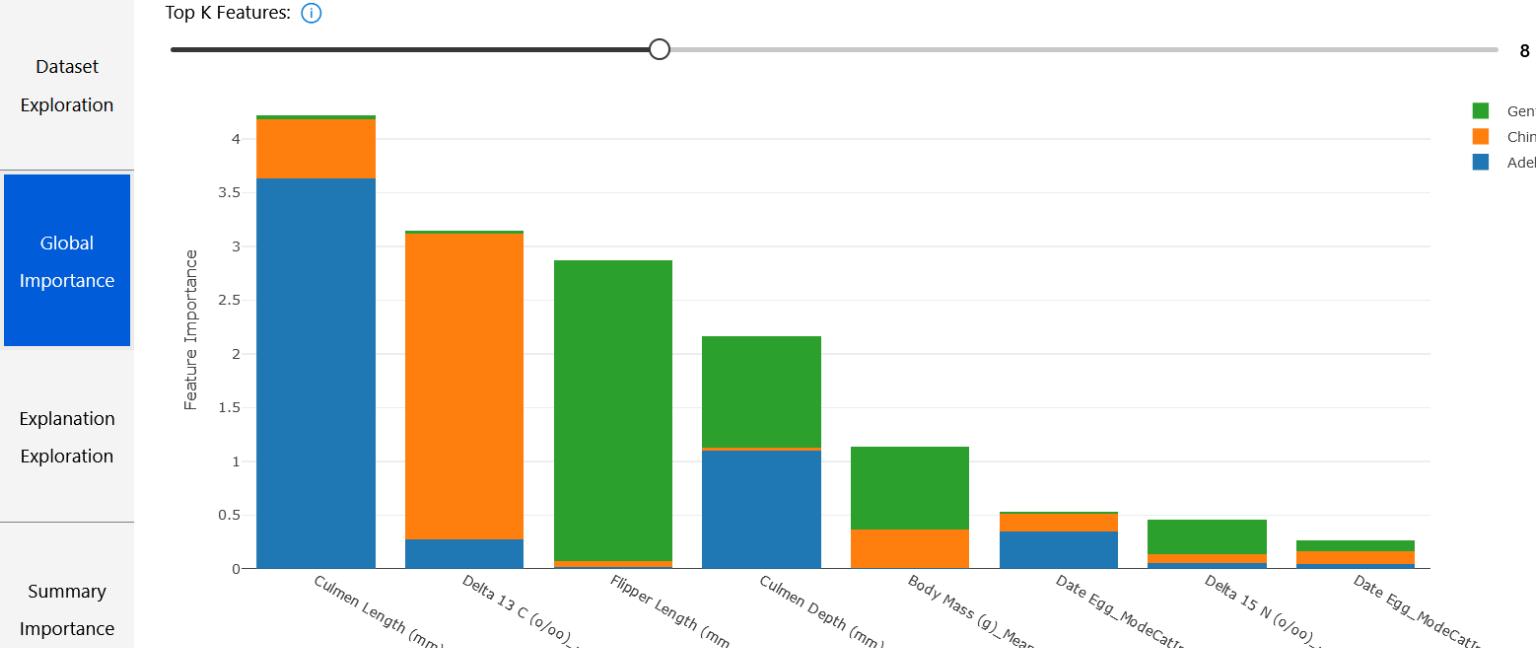
Explanation Exploration

Summary Importance

Feature Importance

Gentoo penguin (Pygoscelis papua)
Chinstrap penguin (Pygoscelis antarctica)
Adelie Penguin (Pygoscelis adeliae)

Feature	Adelie Penguin (Pygoscelis adeliae)	Chinstrap penguin (Pygoscelis antarctica)	Gentoo penguin (Pygoscelis papua)
Culmen Length (mm)	~3.6	~0.1	~0.3
Delta 13 C (‰)	~0.2	~3.1	~0.1
Flipper Length (mm)	~0.1	~0.1	~2.9
Culmen Depth (mm)	~1.1	~0.1	~1.1
Body Mass (g)_Mean...	~0.4	~0.3	~0.5
Date Egg_ModeCatIm...	~0.4	~0.1	~0.0
Delta 15 N (‰)_Mode...	~0.1	~0.1	~0.3
Date Egg_ModeCatIm...	~0.1	~0.1	~0.2



Explanations (explanation exploration)

Microsoft Azure Machine Learning

eacbmlservicews > Experiments > penguins-automatedML > Run 272 > Run 281

Run 281 Completed

↻ Refresh ▷ Deploy ⬇️ Download 🔍 Explain model ✖️ Cancel

Details Model Explanations (preview) Metrics Outputs + logs Images Child runs Snapshot

Model explanations are used to understand what features are directly impacting the model and why. [Learn more](#)

Select Explanation

tabular | mimic.lightgbm | engineered | classification | c1da2a2c-a2a3-4a3c-87cf-4e21768ea063 | 7/31/2020, 2:40:26 PM

Explainer: mimic.lightgbm

X value: Importance : Body Mass (g)_MeanImputer
Y value: Importance : Culmen Length (mm)_MedianImputer

Color: Predicted Y
Cross-class weighting: Predicted class

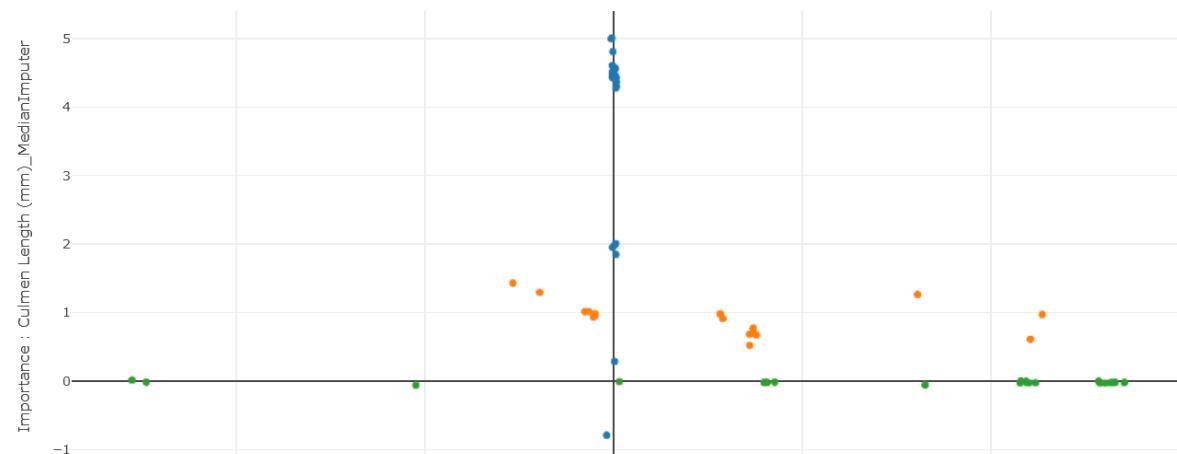
Dataset Exploration Global Importance

Explanation Exploration

Summary Importance

Importance : Culmen Length (mm)_MedianImputer

Importance : Body Mass (g)_MeanImputer



Legend:

- Adelie Penguin (*Pygoscelis adeliae*)
- Chinstrap penguin (*Pygoscelis antarctica*)
- Gentoo penguin (*Pygoscelis papua*)

Explanations (summary importance)

Microsoft Azure Machine Learning

eacbmilservicews > Experiments > penguins-automatedML > Run 205 > Run 212

Run 212 ✓ Completed

Refresh Deploy Download Explain model Cancel

Details Model Explanations (preview) Metrics Outputs + logs Images Child runs Snapshot

Model explanations are used to understand what features are directly impacting the model and why. [Learn more](#)

Select Explanation

tabular | mimic.lightgbm | engineered | classification | b2f7213f-cf1b-4380-a510-570e0a24d9fe | 7/28/2020, 7:58:35 AM

Explainer: mimic.lightgbm

Dataset Exploration Chart type: Swarm Color: Scaled Feature ... Top K Features: 8 Cross-class weighting: Predicted class

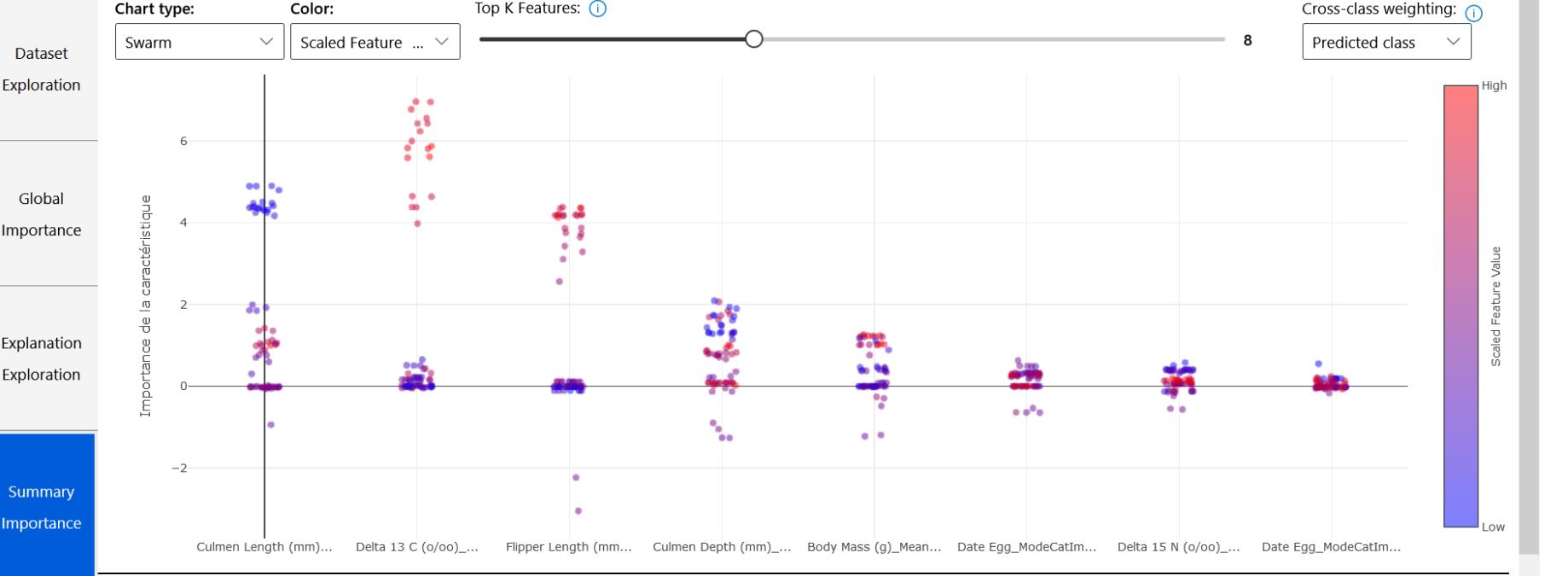
Global Importance

Explanation Exploration

Summary Importance

Importance de la caractéristique

Culmen Length (mm)... Delta 13 C (o/oo)... Flippers Length (mm)... Culmen Depth (mm)... Body Mass (g)_Mean... Date Egg_ModeCatIm... Delta 15 N (o/oo)... Date Egg_ModeCatIm...



Save the (best) model (GUI versus code)

eacbmilservicews > Experiments > penguins_GUI_autoML > Run 1 > Run 5

Run 5 ✓ Completed

Refresh Deploy Download Explain model Cancel

Details Model Explanations (preview) Metrics Outputs + logs Images Child runs Snapshot

Model summary

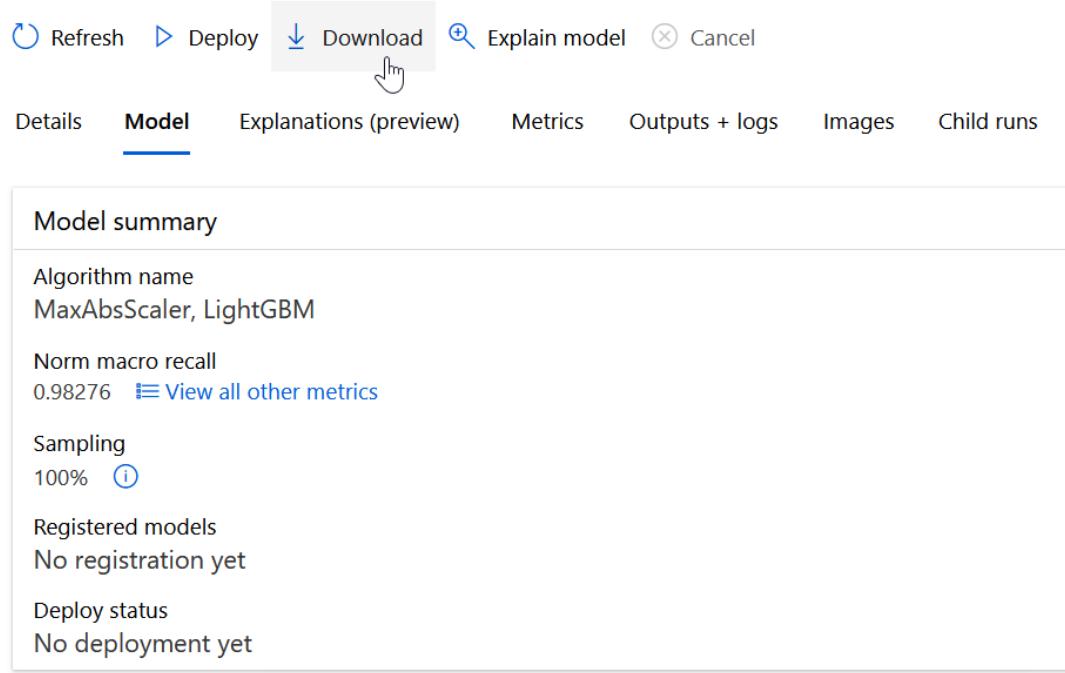
Algorithm name
MaxAbsScaler, LightGBM

Norm macro recall
0.98276 [View all other metrics](#)

Sampling
100% [ⓘ](#)

Registered models
No registration yet

Deploy status
No deployment yet



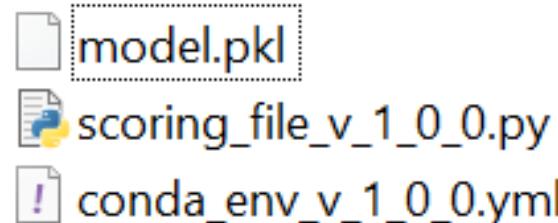
```
best_run, fitted_model = run.get_output()
```

```
from azureml.core.model import Model
```

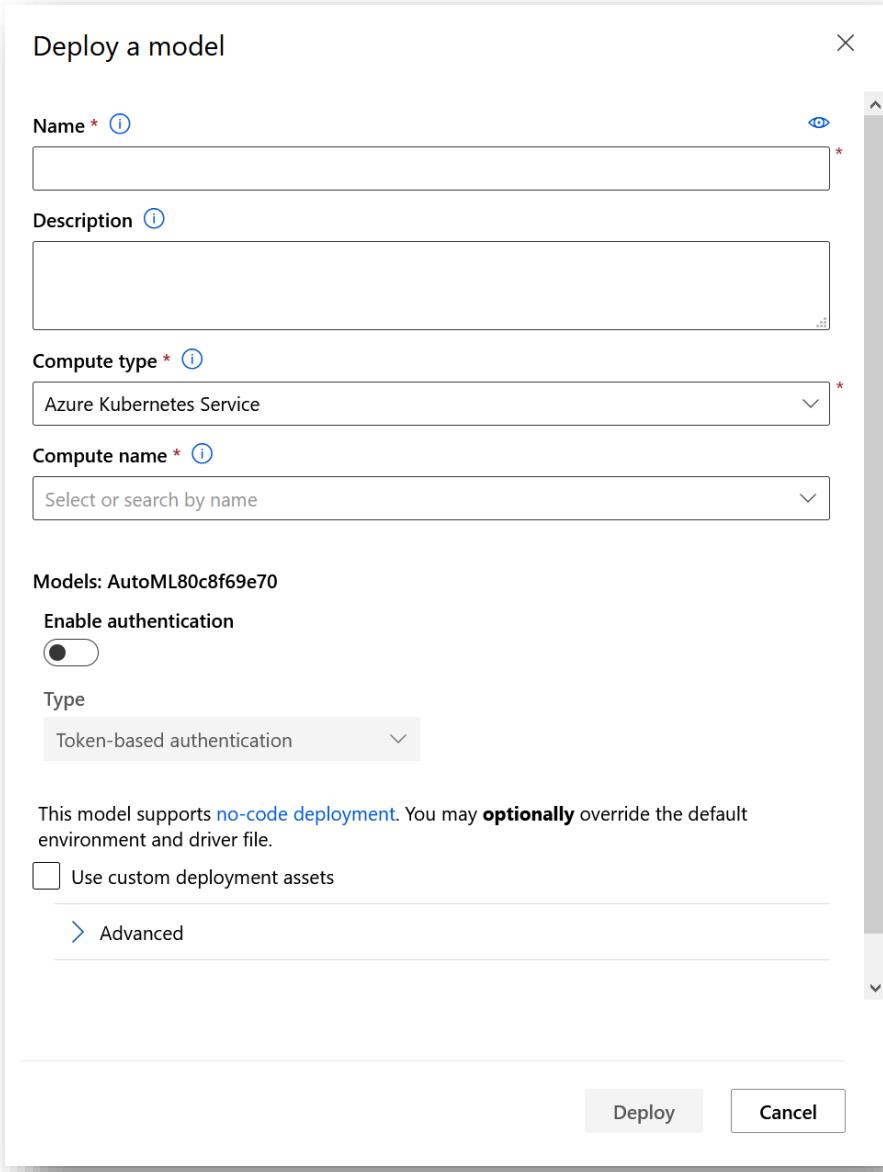
```
model =  
best_run.register_model(model_name='penguins_  
class_best_model',  
model_path='outputs/model.pkl')
```

```
y_predict =  
fitted_model.predict(test_data.to_pandas_  
dataframe().drop(label, axis=1))
```

```
class_prob =  
fitted_model.predict_proba(test_data.to_panda  
s_dataframe().drop(label, axis=1))
```



Model deployment (GUI versus code)



Define an inference configuration

- Package dependencies
- Scoring script

Deploy on Azure : ACI or AKS

```
: from azureml.core import Webservice
from azureml.core.webservice import AciWebservice
from azureml.exceptions import WebserviceException

service_name = 'penguins-prediction-service'

# Remove any existing service under the same name.
try:
    Webservice(ws, service_name).delete()
except WebserviceException:
    pass

aci_config = AciWebservice.deploy_configuration(cpu_cores=1, memory_gb=1, auth_enabled=True)

service = Model.deploy(workspace=ws,
                      name=service_name,
                      models=[model],
                      inference_config=inference_config,
                      deployment_config=aci_config)

service.wait_for_deployment(show_output=True)
```

CONTEST(S)



GUI (low code) vs SDK (code first), who's best ?

GUI :

Common mistakes are easier

Can't modify a run, must redo every steps



SDK :

Real data preparation with pandas

Featurization **only with enterprise licence**

Choice between local and remote compute (including Azure Databricks)

Widget in notebooks

Choose the best model not only on the primary metric

ONNX model compatibility

Integration in a more complex pipeline (with hyperdrive for the best model and test)



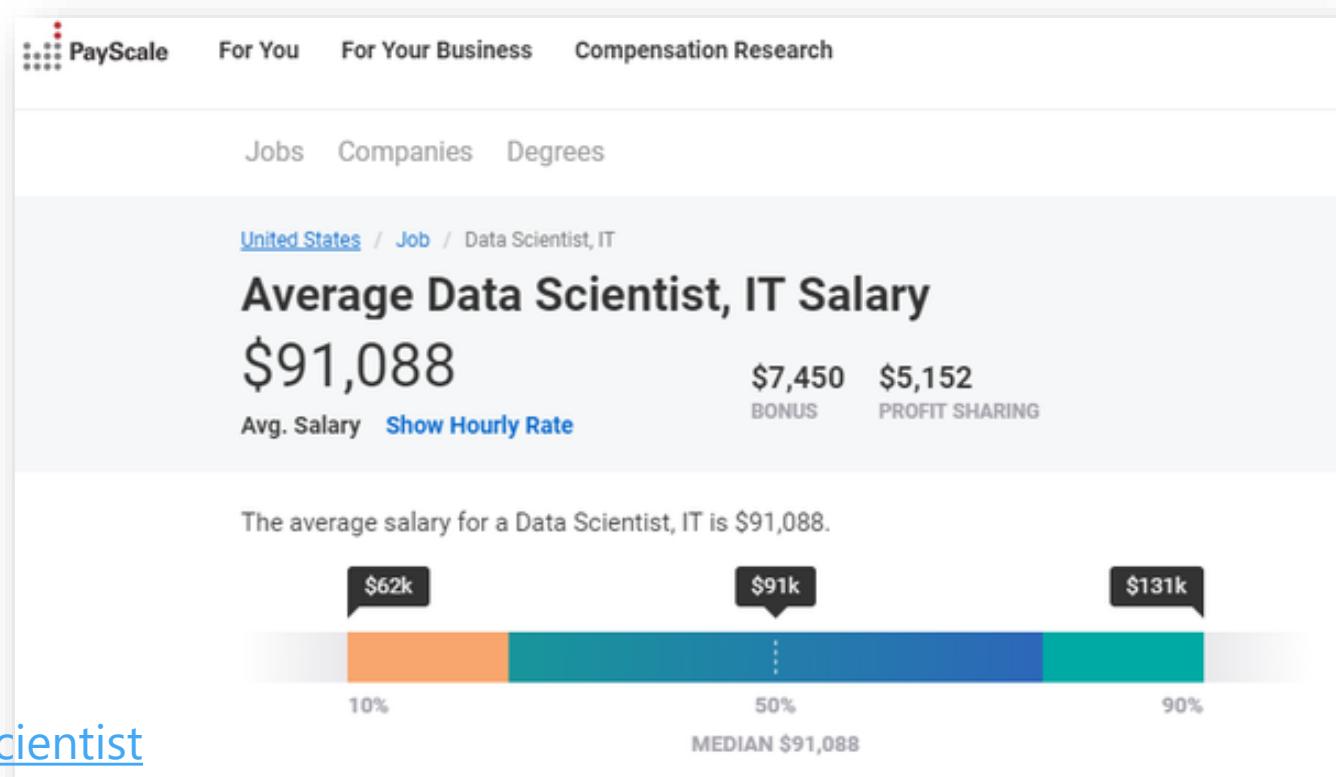
Comparative pricing

The costs of Azure Machine Learning are related to the VMs used to

- launch notebooks (calculation instance),
- train models (training cluster) or
- deploy models (inference cluster).

Three additional resources will be deployed at additional cost:

- [Azure Container Registry](#) Standard account
- [Azure Block Blob Storage](#) (general purpose v1)
- [Key Vault](#)



AutoML vs Data Scientist, that is not the good question

This is not an opposition: automated ML is a tool that must be used with discernment.

It is always essential to :

- Understand the business issues and its stakes
- Understanding the data made available
- Evaluate its quality, improve it if necessary
- Remove information that is not relevant to the model
- Transforming certain data to better exploit them
- To know which algorithm families are relevant according to the use case
 - Linear or non-linear models
 - More or less complex models
 - Ensemble (but beware of overlearning)
- Have a complete view of the evaluation metrics
- Integrate training time and cost
- Interpret the selected

So, do we still need Data Scientists ?

YES!

- To ask the right question for supervised algorithms
- For complex data preparation
- To select features in a non-automated way (based on business criteria)
- For Machine Learning other than supervised: unsupervised, reinforcement learning, transfer, etc.
- To read the interpretation of the models with a critical eye

In conclusion:

Azure ML and the Designer = the *playground* of Citizen Data Scientists

Azure ML and its (Python) SDK = the Data Engineers new best friend !

Webographie

<https://azure.microsoft.com/en-us/services/machine-learning-service/>

<https://docs.microsoft.com/en-us/azure/machine-learning/service/concept-automated-ml>

<https://azure.microsoft.com/en-us/pricing/details/machine-learning-service/>

Forum : <https://social.msdn.microsoft.com/Forums/en-US/home?forum=AzureMachineLearningService>

Samples notebooks

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

<https://github.com/Azure/MachineLearningNotebooks/blob/master/tutorials/regression-automl-nyc-taxi-data/regression-automated-ml.ipynb>

<https://github.com/Azure/MachineLearningNotebooks/blob/master/how-to-use-azureml/automated-machine-learning/forecasting-energy-demand/auto-ml-forecasting-energy-demand.ipynb>



SIÈGE SOCIAL
52 Avenue André Morizet
92100 Boulogne-Billancourt

Paris – Bordeaux – Lyon – Nantes – Toulouse



Siège Social

52, avenue André Morizet - 92100 Boulogne-Billancourt
contact@azeo.com | [01.83.62.65.54](tel:01.83.62.65.54) | www.azeo.com

AZEO
talents & technology



Paul Péton
Tech Lead Data & AI
MVP AI & Data Platform
paul.peton@azeo.com
06 11 10 22 01



Jean-Pierre Riehl
Directeur de l'Innovation
MVP Data Platform
jean-pierre.riehl@azeo.com
06 62 50 61 24

Gold
Microsoft Partner

Contacter un conseiller