

Démystifier le Machine Learning à travers AzureML

Pierre LEROY – Clément ROMAC

2017
Global **Azure**
BOOTCAMP

Speakers



Pierre LEROY

Alternant Data Scientist, Entrepreneur

Co-fondateur d'Enimia
Big Data, Machine Learning



Clément ROMAC

Alternant Machine Learning & IA

Co-directeur d'Enimia
Machine Learning, Deep Learning





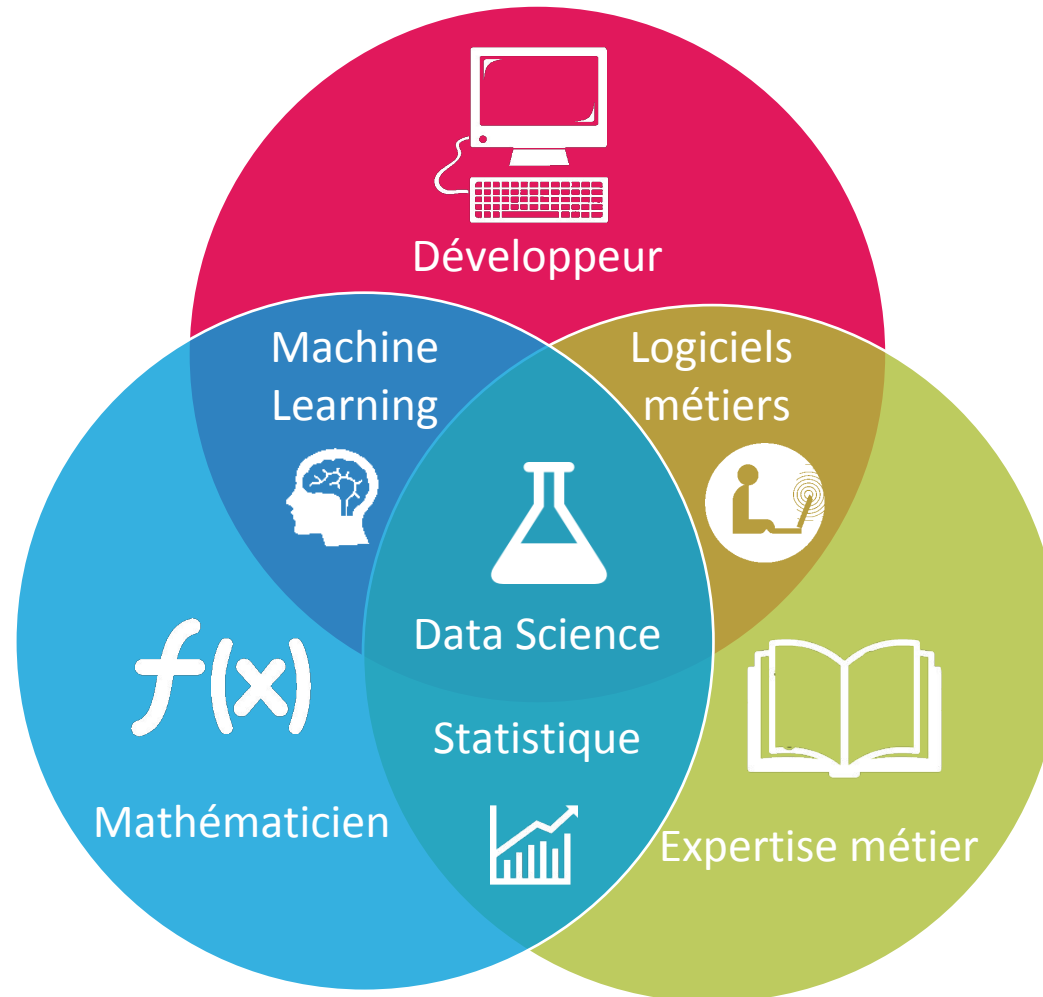
MACHINE LEARNING

DEFINITION DES CONCEPTS DE BASE

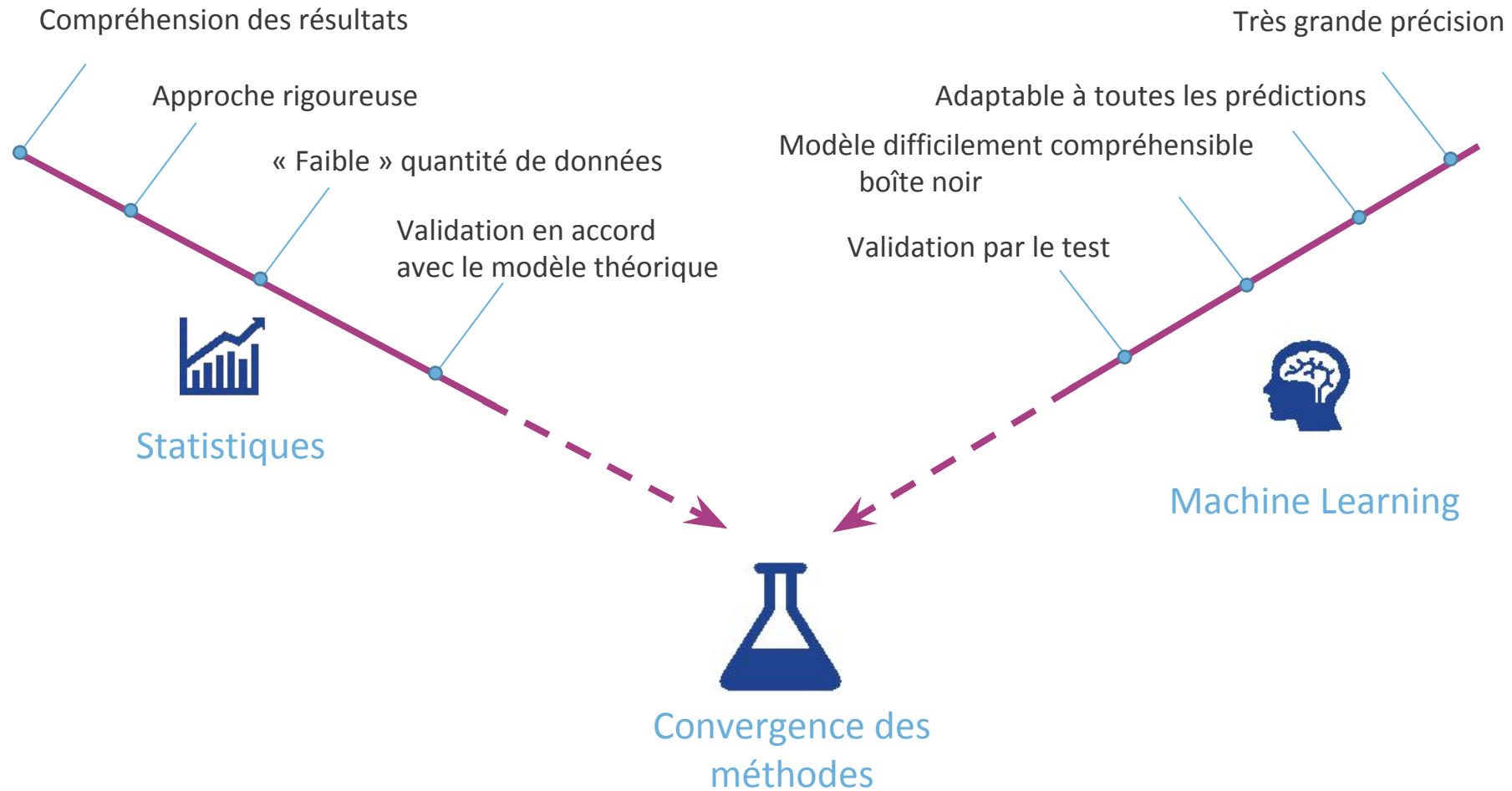
Qu'est ce que le machine learning ?

«L'apprentissage automatique est la science permettant aux ordinateurs d'accomplir des tâches sans avoir été explicitement programmé dans ce sens.» Andrew Ng

Qu'est ce que le machine learning ?

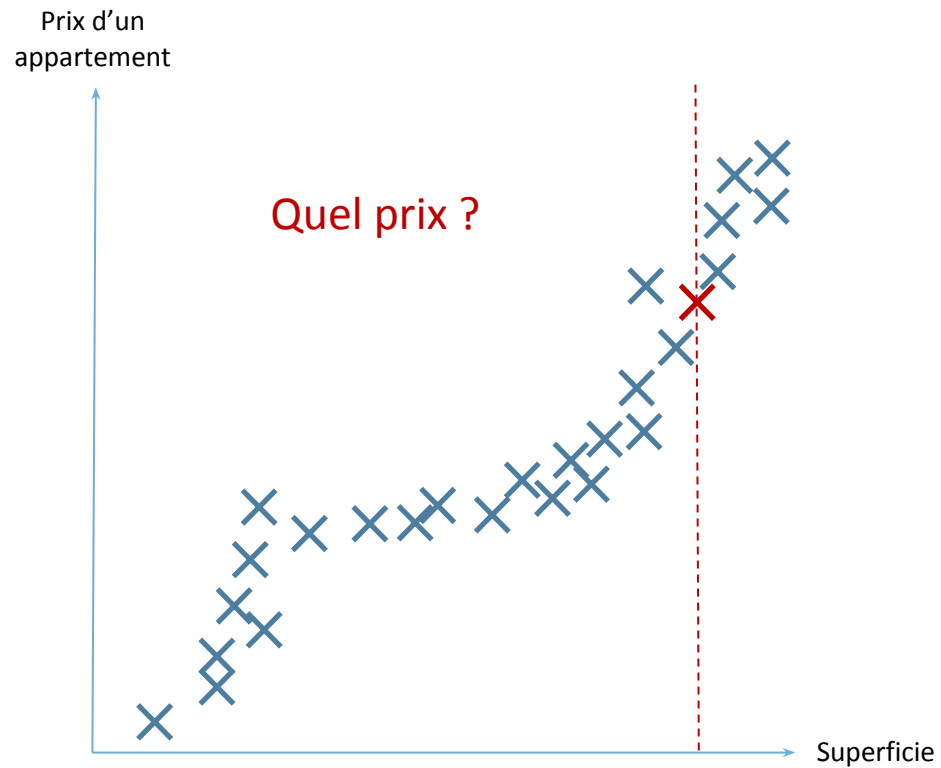


Le machine Learning, des statistiques faites par les développeurs ?

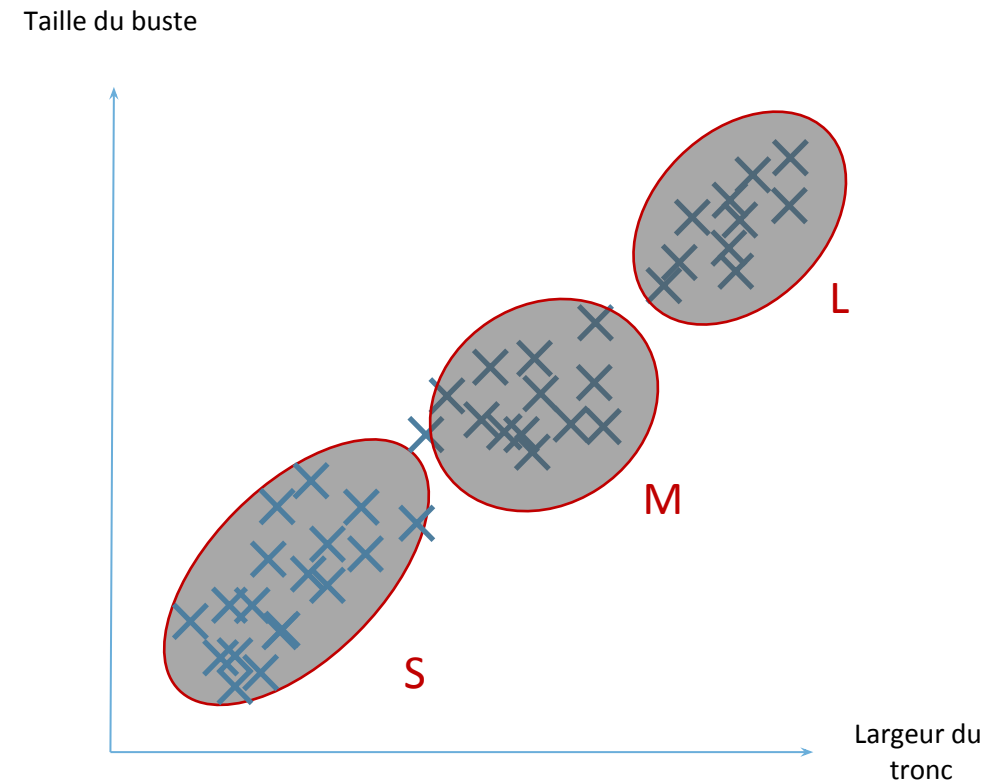


Deux familles d'algorithmes

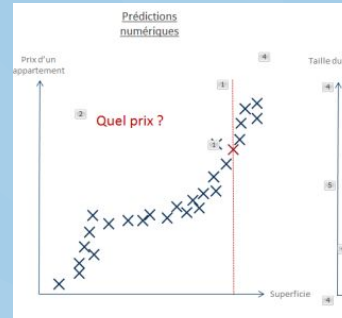
Supervisé
(données labellisées)



Non Supervisé
(données non labellisées)



QUIZZ



Supervisé



Non supervisé

Connaître l'âge d'un utilisateur



Détecter d'anomalies



Prévoir si un programme est malveillant



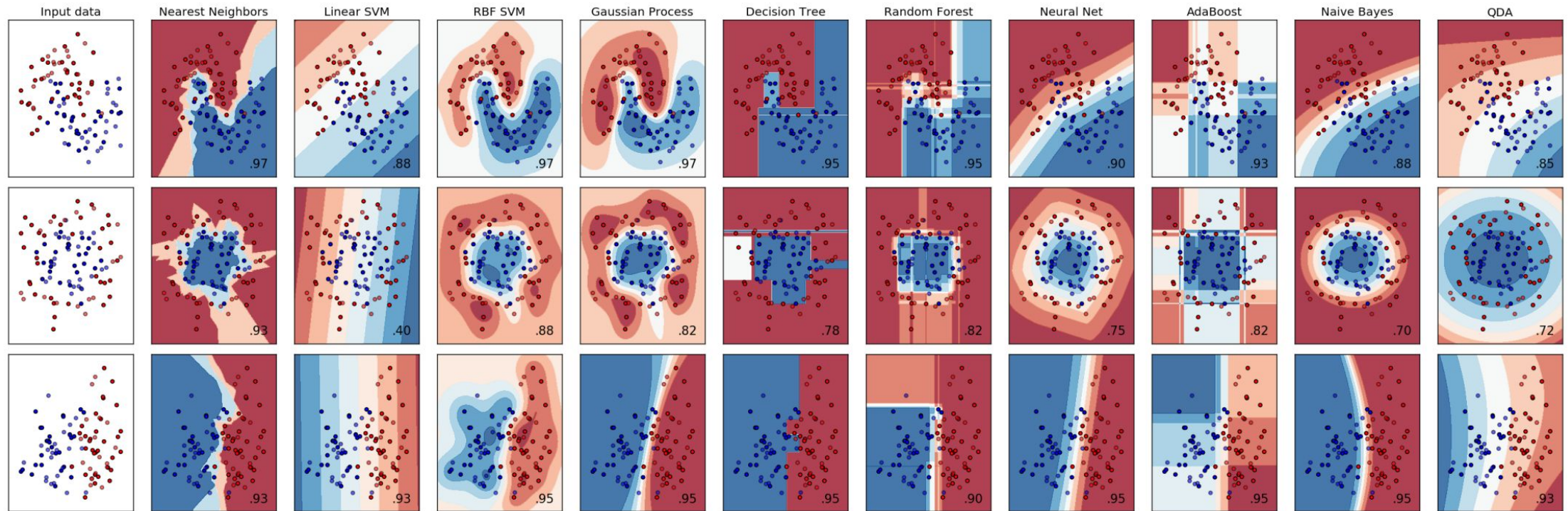
Reconnaître un visage sur une image



Prévoir le temps de rentabilité
d'une action en bourse



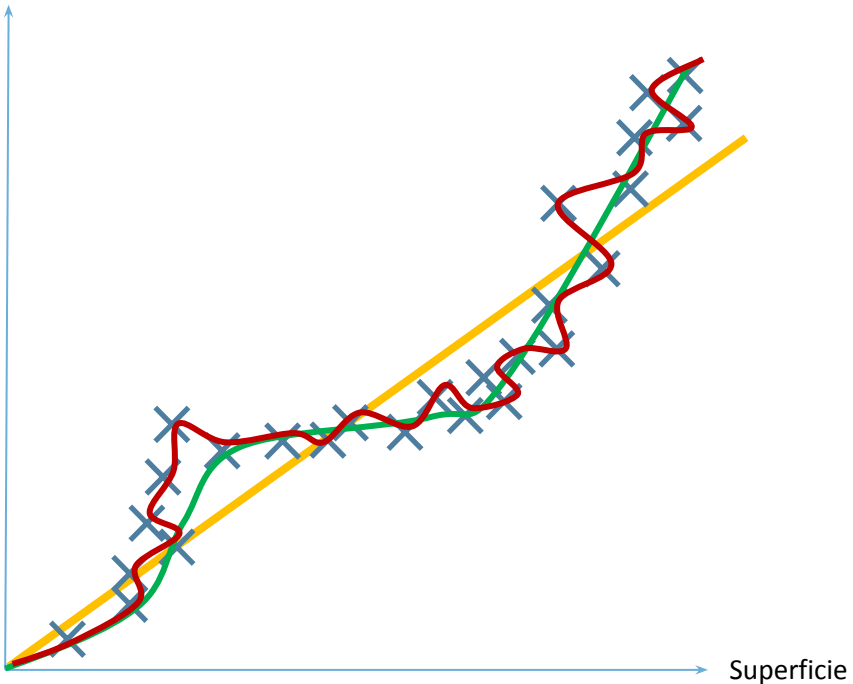
Comment choisir le « bon » algorithme ?



Quelles métriques pour valider mon algorithme ?

Données connues en
base de données

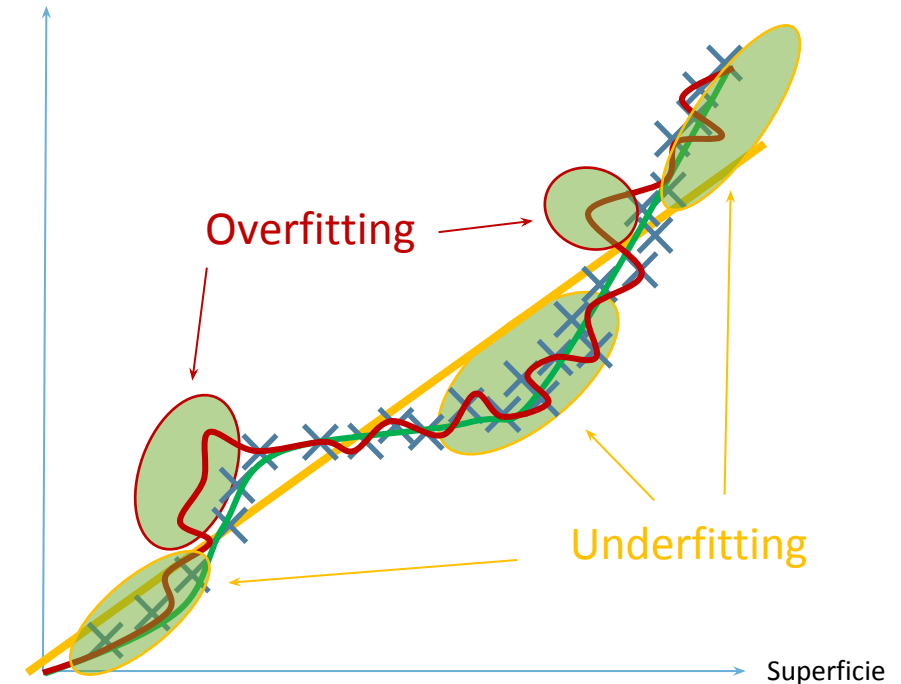
Prix d'un
appartement



- Hypothèse linéaire
- Hypothèse non linéaire
- Hypothèse au plus proches des données

Nouvelles données
entrantes

Prix d'un
appartement

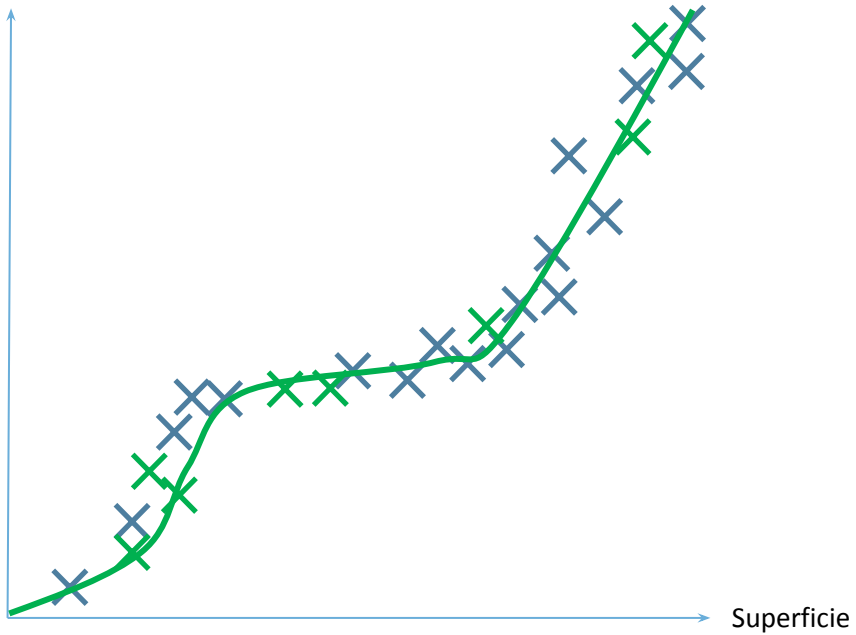


- Hypothèse linéaire
- Hypothèse non linéaire
- Hypothèse au plus proches des données

Vérifier son algorithme

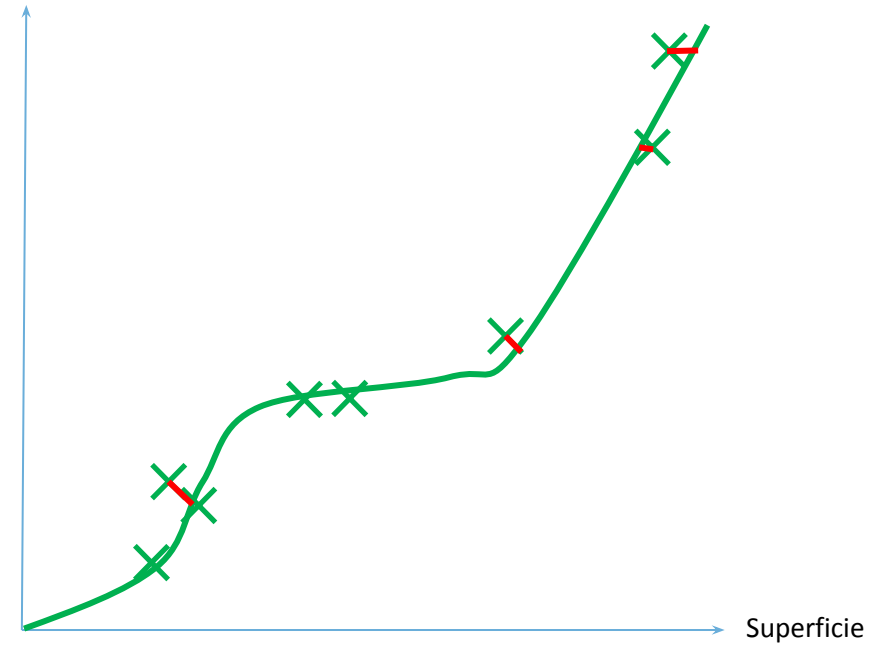
Données connues en
base de données –
Train (entraînement)

Prix d'un
appartement



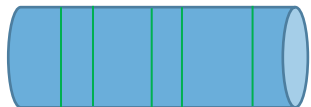
Données connues en
base de données – Test

Prix d'un
appartement



— Hypothèse non linéaire
— Mesure de la différence entre la valeur réelle et l'hypothèse

Data set



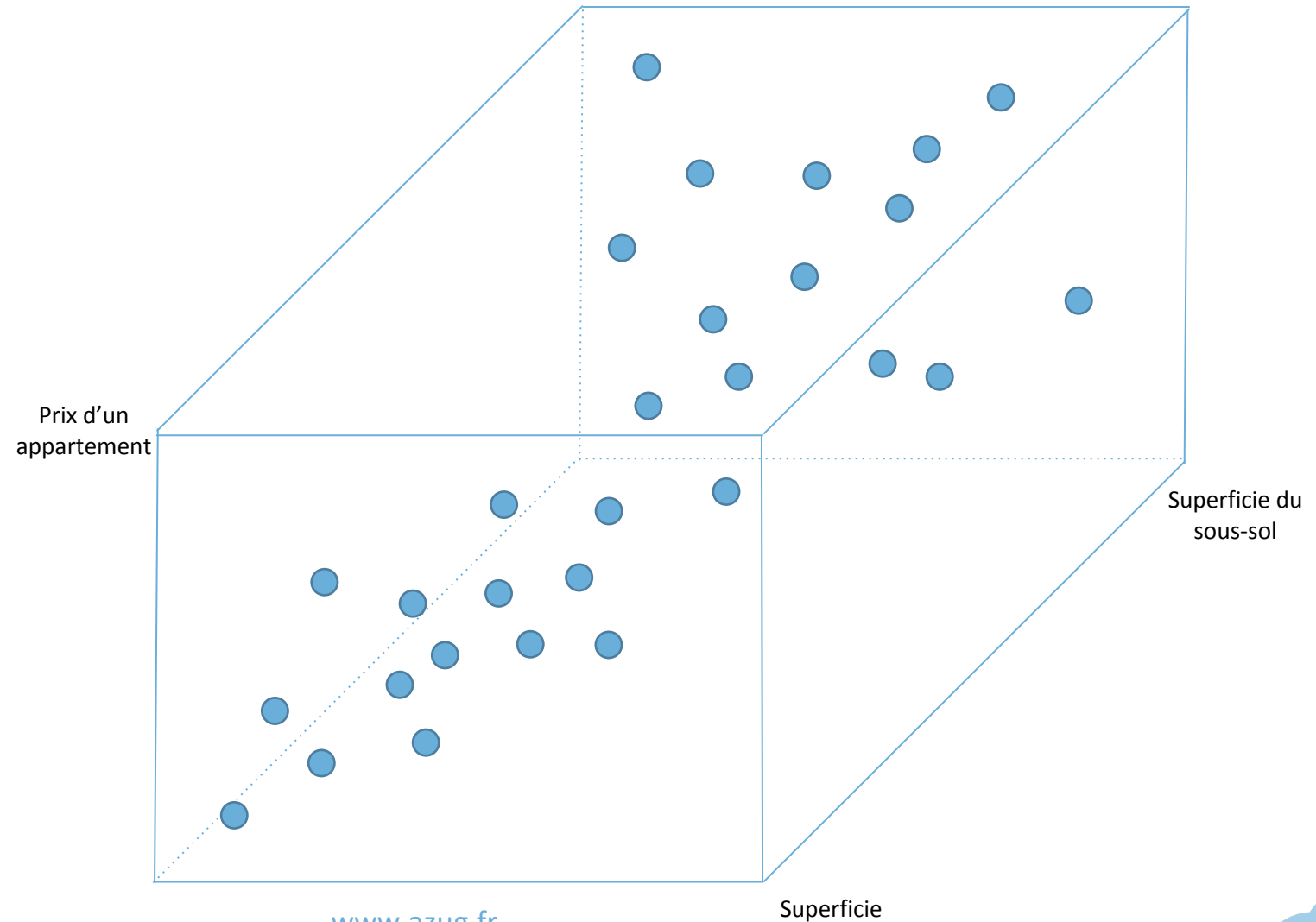
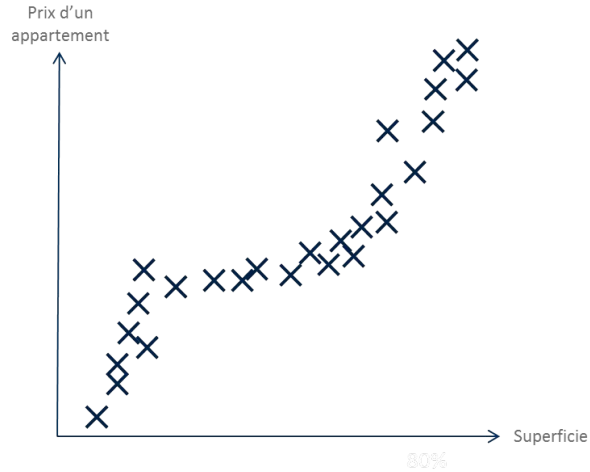
Train
set

70%

Test
set

30%

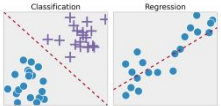
Plusieurs caractéristiques



Récapitulatif



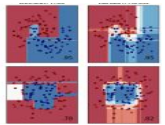
Nécessite des données



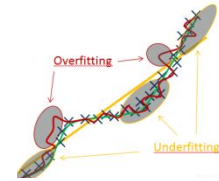
Adaptable à chaque problématique :
Supervisé ou Non Supervisé



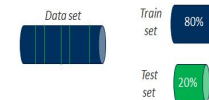
Plus ou moins une boîte noir



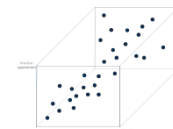
Il existe plusieurs algorithmes qui sont
plus ou moins efficaces en fonction de la
problématique



2 Phénomènes : Overfitting et Underfitting



Séparation du jeu de données en 2 parties :
Entraînement et Test



Modèle en x dimension(s), où
 x = nombre de paramètres



AZURE ML

PREDIRE LE PRIX D'UN APPARTEMENT

Un rapide tour

The screenshot shows the Microsoft Azure Machine Learning Studio interface. The browser address bar displays the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/4a5749566c98455ab218a0eb05eb0bee#Workspace/Experiments/ListExperiments>. The page title is "Microsoft Azure Machine Learning Studio". The left sidebar contains navigation links: PROJECTS, EXPERIMENTS (selected), WEB SERVICES, NOTEBOOKS, DATASETS, TRAINED MODELS, and SETTINGS. The main content area is titled "experiments" and has tabs for "MY EXPERIMENTS" and "SAMPLES". Below the tabs is a table with columns: NAME, AUTHOR, STATUS, LAST EDITED, and PROJECT. The table is empty, with the text "No experiments found" displayed. To the right of the table, it says "0 items selected". At the bottom of the page, there is a dark bar with a "+ NEW" button and two icons: "DELETE" and "ADD TO PROJECT".

Microsoft Azure Machine Learning Studio

ROMAC Clément-Free-Wo...

experiments

MY EXPERIMENTS SAMPLES

NAME	AUTHOR	STATUS	LAST EDITED	PROJECT
No experiments found				

0 items selected

+ NEW

DELETE ADD TO PROJECT

Création d'une expérience

The screenshot displays the Microsoft Azure Machine Learning Studio web interface. The browser address bar shows the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/4a5749566c98455ab218a0eb05eb0bee#Workspaces/Experiments/Experiment/Draft/ViewExperiment>. The page title is "Microsoft Azure Machine Learning Studio".

The main workspace is titled "AZUG Bordeaux 2017" and is in "In draft" status. A search bar labeled "Search experiment items" is located at the top left of the workspace. A single experiment item, "PriceHousing.csv", is visible in the workspace, marked with a circled "1".

The left sidebar contains a list of experiment modules:

- Saved Datasets
- Data Format Conversions
- Data Input and Output
- Data Transformation
- Feature Selection
- Machine Learning
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
 - Create R Model
 - Execute R Script
- Statistical Functions
- Text Analytics
- Time Series
- Web Service
- Deprecated

A "Mini Map" view is visible in the bottom left corner, showing a smaller version of the workspace with the "PriceHousing.csv" item. The bottom toolbar includes icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

Exploration des données

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The browser address bar shows the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/4a5749566c98455ab218a0eb05eb0bee#Workspaces/Experiments/Experiment/Draft/ViewExperiment>. The page title is "Microsoft Azure Machine Learning Studio".

The main workspace is titled "AZUG Bordeaux 2017" and is in "In draft" status. A search bar labeled "Search experiment items" is located on the left. The left sidebar contains a list of modules:

- Saved Datasets
- Data Format Conversions
- Data Input and Output
- Data Transformation
- Feature Selection
- Machine Learning
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
 - Create R Model
 - Execute R Script
- Statistical Functions
- Text Analytics
- Time Series
- Web Service
- Deprecated

The main workspace area shows a single module named "PriceHousing.csv" with a small "1" icon next to it. A "Mini Map" view is visible in the bottom left corner, showing a smaller version of the workspace with the "PriceHousing.csv" module. The bottom toolbar includes icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

Entrainement d'un modèle

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The browser address bar shows the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/4a5749566c98455ab218a0eb05eb0bee#Workspaces/Experiments/Experiment/4a5749566c98455ab218a0eb05>. The workspace is titled "AZUG Bordeaux 2017".

Left Panel (Navigation):

- Search experiment items
- Saved Datasets
- Data Format Conversions
- Data Input and Output
- Data Transformation
- Feature Selection
- Machine Learning
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Time Series
- Web Service
- Deprecated

Main Canvas:

The workflow diagram shows the following steps:

- Import Data (File: Pricelisting.csv)
- Summarize Data (Σ)
- Clean Missing Data (Clean Missing Data)
- Summarize Data (Σ)
- Clean Missing Data (Clean Missing Data)
- Split Data (Split Data)

Each step is marked with a green checkmark, indicating it has been completed. A "Mini Map" view is available at the bottom left of the canvas.

Right Panel (Properties):

Experiment Properties

START TIME	4/16/201...
END TIME	4/16/201...
STATUS CODE	Finished
STATUS DETAILS	None

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

Choix du meilleur algorithme

Experiments - Microsoft

Securisé | <https://studio.azureml.net/Home/ViewWorkspaceCached/4a5749566c98455ab218a0eb05eb0bee#Workspaces/Experiments/Experiment/4a5749566c98455ab218a0eb05eb0bee>

Microsoft Azure Machine Learning Studio

ROMAC Clément-Free-Wo...

Finished running ✓

AZUG Bordeaux 2017

Search experiment items

- Saved Datasets
- Data Format Conversions
- Data Input and Output
- Data Transformation
- Feature Selection
- Machine Learning
 - Evaluate
 - Cross Validate Model
 - Evaluate Model
 - Evaluate Recommend...
 - Initialize Model
 - Score
 - Train
 - OpenCV Library Modules
 - Python Language Modules
 - R Language Modules
 - Statistical Functions
 - Text Analytics
 - Time Series

Properties Project

Experiment Properties

START TIME	4/16/201...
END TIME	4/16/201...
STATUS CODE	Finished
STATUS DETAILS	None

Prior Run

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

Expérience prédictive

Experiments - Microsoft | x

Securisé | <https://studio.azureml.net/Home/ViewWorkspaceCached/4a5749566c98455ab218a0eb05eb0bee#Workspaces/Experiments/Experiment/4a5749566c98455ab218a0eb05eb0bee>

Microsoft Azure Machine Learning Studio

ROMAC Clément-Free-Wo... ?

Finished running ✓

AZUG Bordeaux 2017

Search experiment items

- Saved Datasets
- Data Format Conversions
- Data Input and Output
- Data Transformation
- Feature Selection
- Machine Learning
 - Evaluate
 - Cross Validate Model
 - Evaluate Model
 - Evaluate Recommend...
 - Initialize Model
 - Score
 - Train
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Time Series

Properties Project

Boosted Decision Tree Regre...

Create trainer mode
Single Parameter

Maximum number of le...
20

Minimum number of sa...
10

Learning rate
0.2

Total number of trees <...
100

Random number seed

☒ Allow unknown cat...

START TIME 4/16/201...
END TIME 4/16/201...
ELAPSED TIME 0:00:00.0...
STATUS CODE Finished
STATUS DETAILS Task output

Quick Help
Creates a regression model using the Boosted Decision Tree algorithm (more help...)

+ NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

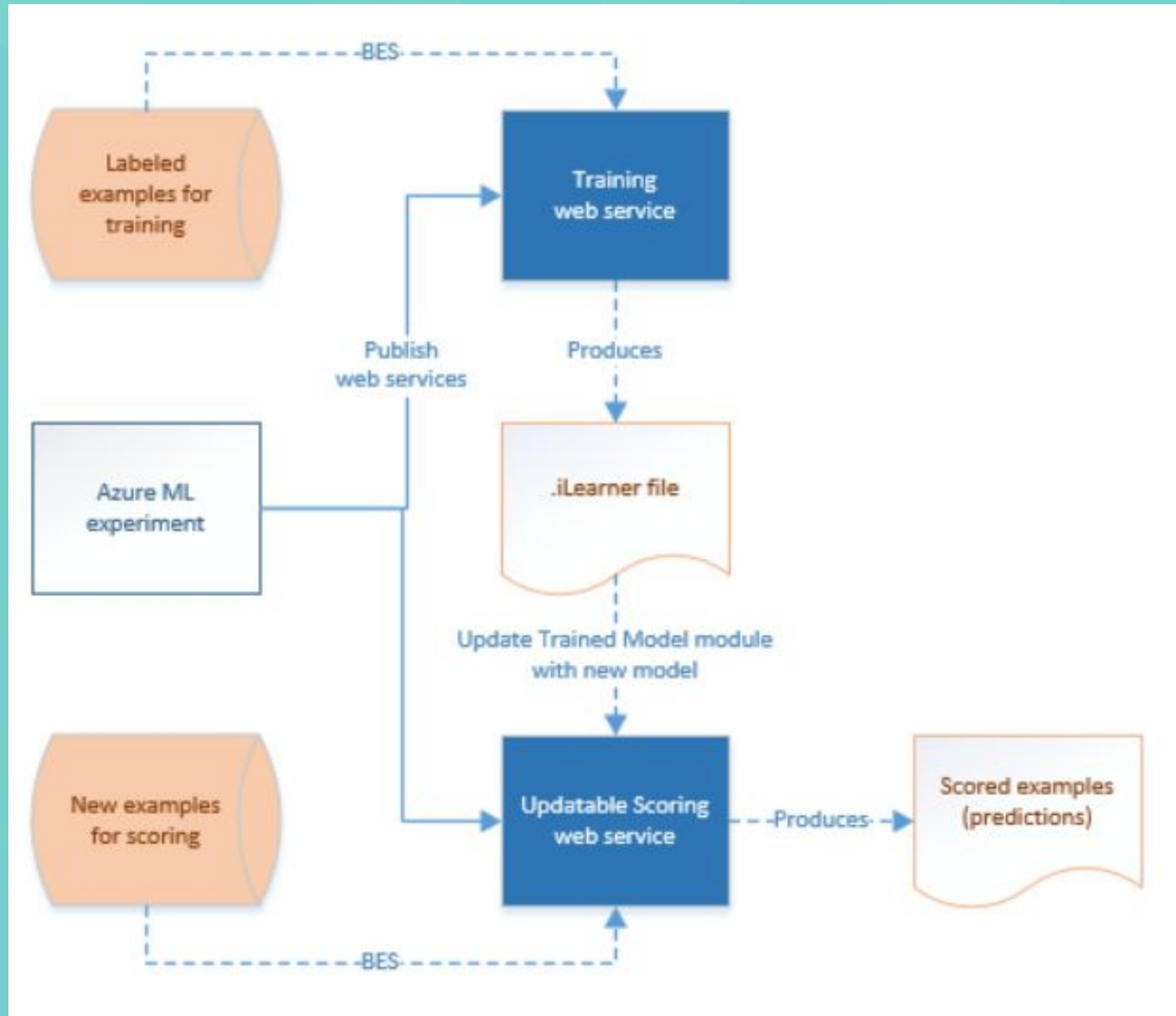
Aller plus loin

The screenshot displays the Microsoft Azure Machine Learning Studio web interface. The browser address bar shows the URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/4a5749566c98455ab218a0eb05eb0bee#Workspaces/Experiments/Experiment/4a5749566c98455ab218a0eb05eb0bee>. The page title is "Microsoft Azure Machine Learning Studio".

The interface is divided into several sections:

- Left Sidebar:** A navigation pane with a search bar "Search experiment items" and a list of categories: Saved Datasets, Trained Models, Transforms, Data Format Conversions, Data Input and Output, Data Transformation, Feature Selection, Machine Learning, OpenCV Library Modules, Python Language Modules, R Language Modules, Statistical Functions, Text Analytics, Time Series, Web Service, and Deprecated.
- Top Bar:** Includes tabs for "Training experiment" and "Predictive experiment", the workspace name "AZUG Bordeaux 2017 [Predictive Exp.]", and a status indicator "Finished running" with a green checkmark. User information "ROMAC Clément-Free-Wo..." and icons for help, users, and a profile are also present.
- Main Canvas:** A workflow diagram for a predictive experiment. It starts with a "Web service input" node, which feeds into an "Apply Transformation" node. This node also receives input from a "Dataset" node labeled "AZUG Bordeaux 2017 (Clean...)". The output of the first transformation feeds into a second "Apply Transformation" node, which also receives input from another "Dataset" node labeled "AZUG Bordeaux 2017 (Clean...)". The output of the second transformation feeds into a "Score Model" node, which also receives input from a "Dataset" node labeled "AZUG Bordeaux 2017 (train...)". The final output is a "Web service output" node.
- Right Panel:** Contains sections for "Experiment Properties" (Start Time: 4/16/201..., End Time: 4/16/201..., Status Code: Finished, Status Details: None), "Summary" (a text area for describing the experiment), and "Description" (a text area for detailed description). A "Quick Help" link is at the bottom.
- Bottom Bar:** A toolbar with icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "DEPLOY WEB SERVICE", and "PUBLISH TO GALLERY".

Aller plus loin



Vision globale d'un projet de Machine Learning

Équipes métiers :
recueil au près des différentes
équipes métiers (différentes sources)
et création d'un *entrepôt de donnée*

Data scientist :
utilisation d'algorithmes de
Machine Learning pour réaliser
une prédiction



Équipe métier :
Création d'une *intuition*
par l'équipe technique

Data cleaner :
Utilisation d'un ETL
pour nettoyer;
transformer, enrichir
les données.
Préparation des
données à une
modélisation

Data scientist :
phase d'amélioration
avec combinaisons
d'algorithmes et
techniques de
validations croisées

Merci à nos sponsors

PLATINUM



LOCAUX



PARTENAIRES MEDIA



2017
Global Azure
BOOTCAMP

Sponsors internationaux



Nous suivre



Facebook

facebook.com/groups/azugfr/



LinkedIn

Linkedin.com/inspirasign



Twitter

twitter.com/AZUGFR



Web

www.azug.fr



Meetup

meetup.com/AZUG-FR/

Nous suivre



LinkedIn

[Linkedin.com/in/pierreleroyfr/](https://www.linkedin.com/in/pierreleroyfr/)



LinkedIn

[Linkedin.com/in/clement-romac](https://www.linkedin.com/in/clement-romac)



Twitter

twitter.com/ClementRomac



Web

<http://enimia.com/>

L'enregistrement est ouvert
!

[http://www.spsevents.org/
city/Monaco/Monaco2017](http://www.spsevents.org/city/Monaco/Monaco2017)

3 juin 2017
#SPSMonaco





Merci
d'être venus

A bientôt !