



Challenge KAGGLE

Data Science eXperience

TechAcademy2018

Team « DATADEALER »



Le challenge

Le challenge Kaggle 1/2

Description du challenge :

- <https://www.kaggle.com/c/talkingdata-adtracking-fraud-detection>



En résumé :

Sur un site WEB qui collecte des données d'activité sur les "clics" :

- détecter/prédire qui a vraiment utilisé l'application mobile proposée (en la téléchargeant) pour ne pas prendre en compte ceux qui auraient juste cliqué (robots ou actions malveillantes)



Le challenge Kaggle 2/2

Les données collectées :



- **ip**: ip address of click.
- **app**: app id for marketing.
- **device**: device **type** id of user mobile phone (e.g., iphone 6 plus, iphone 7, huawei mate 7, etc...)
- **os**: os version id of user mobile phone
- **channel**: channel id of mobile ad publisher
- **click_time**: timestamp of click (UTC)
- **attributed_time**: if user download the app for after clicking an ad, this is the time of the app download

Données à prédire :

- **is_attributed**: the target that is to be predicted, indicating the app was downloaded



$y = f(x)$



La méthode

Méthodologie

Utilisation de Watson Studio (DataScience eXperience):

- <https://datascience.ibm.com>
- Projet : « TechData DATA DEALER »

Chargement des données extraites de Kaggle :

- Jeu d'entraînement : **train_sample.csv** – 100 000 lignes de données aléatoires
- Jeu de test : **test.csv** – 18 M rows

Exploration et raffinement des données :

- Préparation des données : Data Refinery
- Création de nouvelles colonnes (feature engineering)
- Création de rapports statistiques sur les données
- Ecriture d'un Notebook Jupyter/Python : [AdTracking Data exploration](#)

Modélisation :

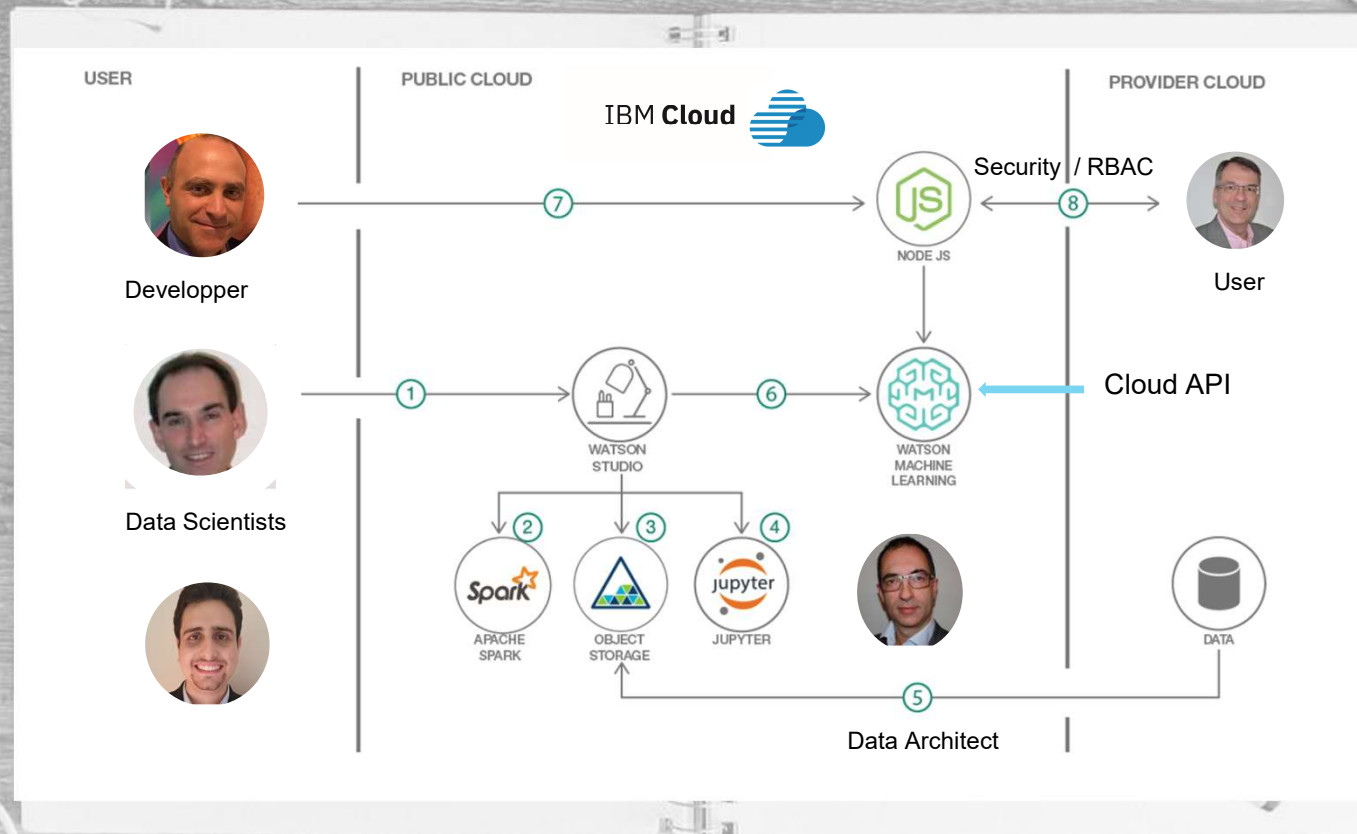
- Ecriture d'un Notebook Jupyter/Python : [AdTracking Data Modeling](#)
- Choix de l'algorithme de modélisation : xgboost
- Mesure des indicateurs

Exploitation :

- Publication du modèle via Watson ML
- Création d'un client test qui se connecte via API REST



Architecture





Démonstration

Watson DataScience eXperience

$$y = f(x)$$

Environnement Watson Data Studio

The screenshot displays the IBM Watson Data Studio web application. The browser address bar shows the URL: <https://dataplatform.cloud.ibm.com/projects/050d8b01-44ed-4>. The interface includes a top navigation bar with tabs for 'My Projects', 'Tools', 'Catalog', 'Community', 'Services', 'Manage', 'Support', and 'Docs'. The main content area is titled 'My Projects / TechAcademy DATA DEALER' and features a table of data assets and sections for various models.

Asset Name	Asset Type	Project	Owner	Last Modified
test.csv	Data Asset	Project	OLIVIER BARROT	25 May 2018, 9:30:05 am
train_sample.flowS1	Data Asset	Project	Jean-François Rieu	24 Apr 2018, 3:52:54 pm
train_sample.csv	Data Asset	Project	Jean-François Rieu	9 Apr 2018, 9:19:59 am

Models

Natural Language Classifier models [+ New Natural Language Classifier model](#)

NAME	MODEL ID	SERVICE INSTANCE	LAST MODIFIED	ACTIONS
You don't have any Natural Language Classifier models yet.				

Visual Recognition models [+ New Visual Recognition model](#)

NAME	MODEL ID	SERVICE INSTANCE	LAST MODIFIED	ACTIONS
You don't have any Visual Recognition models yet.				

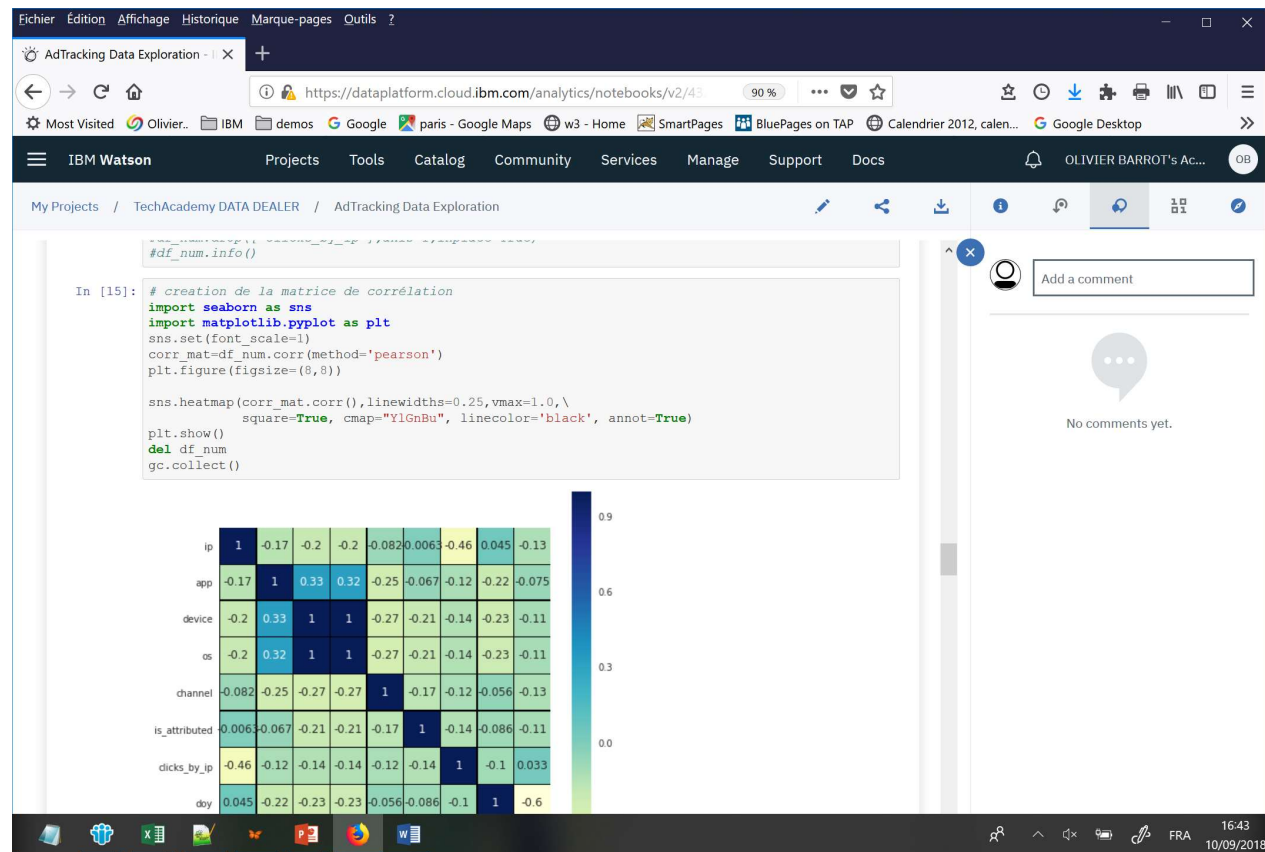
Watson Machine Learning models [+ New Watson Machine Learning model](#)

NAME	STATUS	TYPE	RUNTIME	LAST MODIFIED	ACTIONS
XGBoost model for AdTracking	trained	xgboost-0.6	python-3.5	29 Jun 2018	

The right sidebar contains a 'Load' tab with a file upload area that says 'Drop files here or browse for files to upload.'

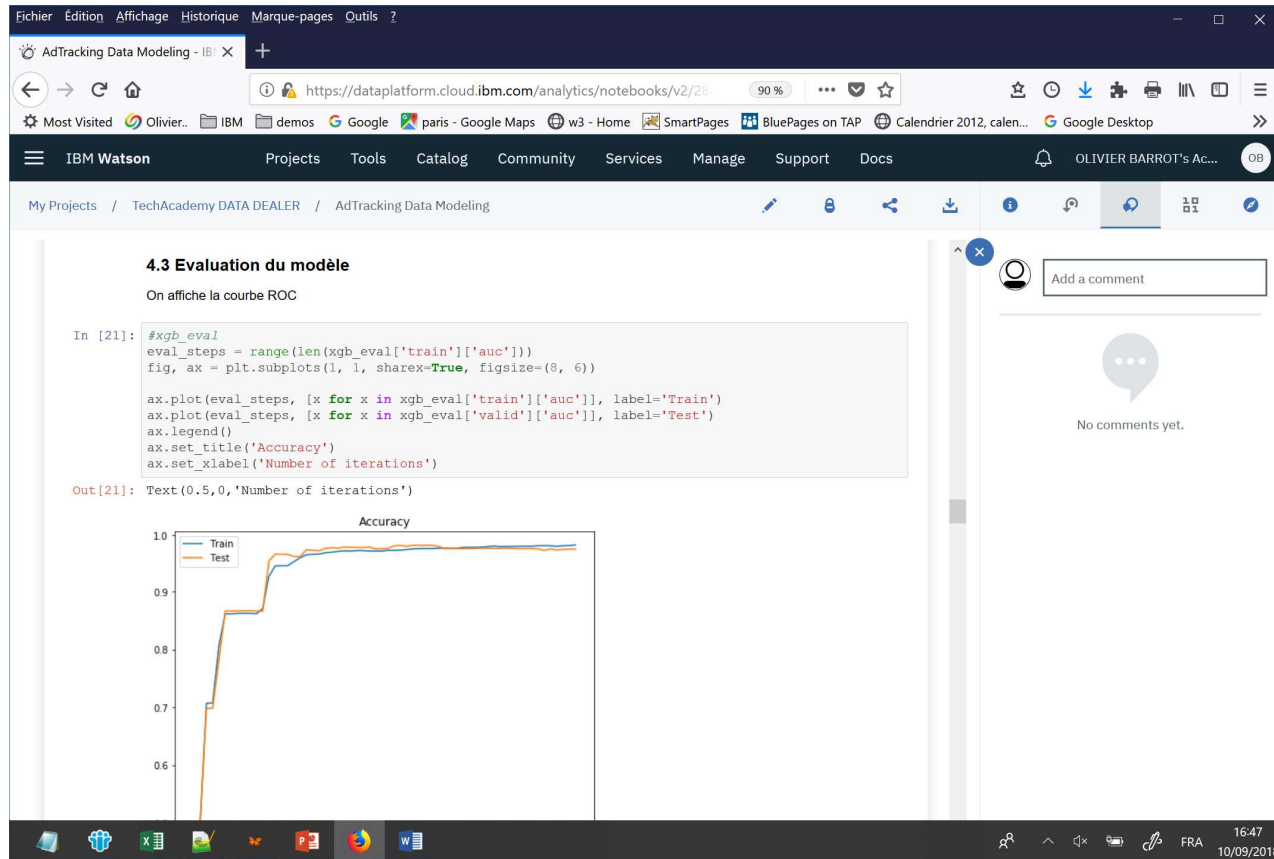
<https://www.ibm.com/cloud/watson-studio>

Notebook exploration



<https://github.ibm.com/TechAcademy2018/DataDealers/blob/master/AdTracking%20Data%20Exploration.ipynb>

Notebook modélisation



<https://github.ibm.com/TechAcademy2018/DataDealers/blob/master/AdTracking%20Data%20Modeling.ipynb>

Client Node.js / WatsonML

Fichier Édition Affichage Historique Marque-pages Outils ?

Tech Academy - Data Dealers - Wat X

https://datadealerswmltester.mybluemix.net

Most Visited Olivier.. IBM demos Google paris - Google Maps w3 - Home SmartPages BluePages on TAP Calendrier 2012, calen... Google Desktop

Challenge KAGGLE
Data Science experience
TechAcademy2018
Équipe : DATADEALER

Tech Academy - Data Dealers
Watson Machine Learning Scoring Demo

APP 9 DEVICE 1 OS 3
CHANNEL 107 CLICK_BY_IP 2 DOY 314
HOD 4

Submit

16:54
10/09/2018

Nous contacter

- ✓ Olivier Barrot "Data Scientist"
olivier.barrot@fr.ibm.com
- ✓ Valéry Debeer "Developper"
valery.debeer@fr.ibm.com
- ✓ Philippe Ponti "User"
pponti@fr.ibm.com
- ✓ Jean-François Rieu "Data Architect"
jfrieu@fr.ibm.com



IBM [A]n[a]lytics

Merci !