



# LEJEUNE Arnaud

Data Scientist - Machine Learning Engineer

[www.alejeune.com](http://www.alejeune.com)

arnaudlejeune44@gmail.com | 07 85 41 15 68

## Description

Passionné par les problématiques autour de la gestion des données et du machine / deep learning.

A la recherche d'une opportunité pour l'été 2019 à Montréal, Canada

## Compétences

### Outils :

/ Pandas / TensorFlow  
/ Scikit-Learn / Keras  
/ Spark / AWS / GCloud  
/ MongoDB / Ubuntu / UNIX

### Langages :

/ Python / Java  
/ Scala / WEB

### Langues :

/ Français : langue maternelle  
/ Anglais : Courant (TOEIC 965)

## Divers

/ Président du B.D.E d'IMT Atlantique campus de Nantes 2016 – 2017.

/ Gardien de handball en championnat universitaire et 10 ans de tennis

## Formation

### Mastère Spécialisé Big Data – Télécom Paristech 2018 - 2019

/ Machine Learning (Sup. et non-sup.) / Statistiques avancées  
/ Data Engineering (Spark & Hadoop) / Systèmes NoSQL (MongoDB, Neo4J, ...)  
/ Deep Learning (CNN, RNN, ...) / Reinforcement Learning (QL, SARSA, ...)

### Ecole d'ingénieur généraliste – IMT Atlantique 2015 - 2018

**Majeure :** Organisation et Management des Technologies de l'Information

/ Big Data / Business Intelligence / Architecture applicative  
/ Architecture Logicielle / Conduite du changement / Gestion de projet

**Tronc Commun :** / Statistiques / Programmation (Impérative, POO ...)  
/ Sciences sociales / Gestion des bases de données

### Classes préparatoires MPSI/MP – Lycée Montaigne 2013 - 2015

## Expériences

### Projet Fil Rouge – BPCE / Télécom ParisTech Oct 2018 – Juin 2019

Projet de MS avec le groupe bancaire BPCE, au sein d'un groupe de 4 étudiants. L'objectif est de créer une base de labels de dénomination usuelle d'entreprises françaises pour clarifier les libellés bancaires, en partant des données SIRET de l'INSEE. Travail sur des données sparses et de faible qualité, étude de l'état de l'art sur les techniques de NLP, et mise en pratique de ces dernières (TF-IDF, Glove, RNN, LSTM, Embedding...)

### Projet GDELT – Télécom ParisTech Dec 2018 - Janv 2019

Mise en place d'une architecture NoSQL cloud (AWS) pour traiter un grand volume de données (3.5 TB). En groupe de 5, nous avons mis en place un cluster MongoDB comprenant 12 machines (2 shards & replication factor : 3). Utilisation de Spark sur un EMR Amazon pour prétraiter la donnée. J'ai été le responsable de la mise en place du cluster sur AWS et de sa configuration.

### Consultant Alternant – Wavestone Sept 2017 – Août 2018

/ Etude des problématiques liées au Big Data, aux bases de données Graphes et aux bases NoSQL. Outre une cartographie des différents cas d'usage, j'ai réalisé un PoC sur Neo4J (en python) et un benchmark des différentes solutions Graphe.

/ Participation à diverses missions de conseil pour des grands comptes français (SNCF, ENEDIS, Ministère des Armées) sur des sujets autour du Big Data, du Machine Learning et de l'architecture de leur SI.

### Projet Reinforcement Learning – IMT Atlantique Février – Juin 2017

Projet de 6 mois à IMT Atlantique, j'ai participé à la réalisation agile de 2 IA sur le jeu Snake, utilisant 2 techniques distinctes de Reinforcement Learning : Breeding pour l'un, Q-Learning pour l'autre. Utilisation de Keras / Tensorflow.