

## CRUTAIN Yves

17 Henchman Street  
W12 OWB London  
Tel : (+44) 07546605818  
[yves.crutain@gmail.com](mailto:yves.crutain@gmail.com)

## Data Scientist

**SQL, R, Machine Learning**  
Six années d'expérience  
[data-laborer.eu](http://data-laborer.eu)

### PROFIL

Je suis issu d'un master statistique et n'ai eu cesse depuis d'aiguiser mes compétences en machine learning, programmation et communication. Je suis force de proposition dans le cadre technique. Je maîtrise plusieurs langages de programmation. Je suis capable de rendre des analyses de sujets complexes compréhensibles par le plus grand nombre, en restant data-driven. Je suis accroc aux meetups de data science et à Github.

Mon expérience englobe les domaines de la recherche, du marketing et de l'économétrie. Je suis autonome, curieux et intéressé par les nouvelles technologies.

J'utilise couramment les outils de data visualisation, avec un penchant pour l'open source.

### EXPÉRIENCES SIGNIFICATIVES

#### MODEL CITIZENS:

- Création de datamart, réalisation de dashboards et implémentation de modèles.
- Analyses Ad hoc et conseil client

#### GDF-SUEZ Département stratégie - CEEME :

- Mise en place d'un datamart et automatisation d'un modèle économétrique.

#### EDF Branche Commerce - Département Analyse et Connaissance Client.

- Analyse de campagnes marketing
- Analyses Ad hoc et réalisation de scores

#### EDF R&D - Icame :

- Analyse sur des problématiques de séries temporelles.

### DOMAINE D'EXPERTISE

- **Statistiques, probabilités, séries temporelles**
- **Machine learning**
- R (ggplot2, tidyverse, shiny), SQL, SAS, python(scikit-learn/numpy)
- Data visualisation

### FORMATION

- **2012** : Coursera : cours d'économétrie et de machine learning
- **2010** : [Master 2 Mathématiques pour l'Informatique Graphique et la Statistique](#)
- **2008** : Licence option mathématiques (Université de Bourgogne – Dijon)

### COMPÉTENCES DIVERSES

- **Suite Microsoft office, access et vba**
- Anglais : Bilingue ; Espagnol: Lire et écrire
- python(scikit-learn), NLP, yaml, AWS(EC2, S3), js, HTML, CSS, markdown, jekyll

## MISSIONS – PROJETS, DÉTAILS

Octobre 2014 - En poste : Data scientist

### MODEL CITIZENS

- Réalisation de dashboards pour grands comptes: BT, Regus(shiny), daily mail, ..
- Réalisation et mise en production de modèles “machine learning” (churn modeling, recommendation engine, scoring model).
- Études marketings Ad-hocs

**Mots clés** : R, shiny, python, h2o, SQL-SERVER, docker, http Rest, jupyter, scoring, deep learning, boosting

Septembre 2013 - Octobre 2014 : Développeur Bâle II

### CETELEM / BNP PARIBAS

- Développement “big data” avec SAS, en charge des analyses statistiques.
- Optimisation de requêtes oracles.
- “Management” de trois consultants SAS.

**Mots clés** : SAS (SEG, base, stat, graph, macro, SQL, connect to oracle, SAS Stored Processes), hash code, Microsoft office add-on, Excel, Bale II, oracle, tests statistiques.

Mai 2013, Keyrus : **Réalisation d'un indicateur d'incivilité pour la SNCF**

Février 2013, Keyrus : Formateur, logiciel R

Octobre 2012 - avril 2013 : Modélisation du marché de l'énergie à 2030

GDF - SUEZ Centre d'Expertise en Etudes et Modélisations Economiques (CEEME)

- **Création d'un datamart sous SAS.**
- Automatisation de la prévision à 2030 du marché Européen de l'énergie, en symbiose avec le datamart.

**Mots clés** : SAS(IML, macro), data management, données AIE, Enerdata et GDF-SUEZ  
Recommandation : [Daniel.grenouilleau@gdfsuez.com](mailto:Daniel.grenouilleau@gdfsuez.com) (chef de projet, macro-économiste senior)

Octobre 2012, Keyrus : Réalisation d'un support de formation sur les sondages

Février 2012 – Septembre 2012 : Étude sur le projet OPTimisation de COnsommation (OPCO)

EDF R&D Innovation Commerciale et Analyse du Marché de l'Énergie

- **Étude des influences matérielles, sociologiques et météorologiques sur la consommation électrique lors d'un effacement.**
- Étude sur les méthodes d'estimations pour le traitement de valeurs manquantes et de plages horaires effacées dans une série temporelle.
- Automatisation sous SAS du data management.

**Mots clés** : SAS(Stat,macro, connect to DB2), R, enquête logement insee, régression, séries temporelles, forte volumétrie

Recommandation : [Philippe.charpentier@edf.com](mailto:Philippe.charpentier@edf.com) (Ingénieur chercheur senior)

Janvier 2012, Keyrus : Réalisation d'un support de formation R

Juillet 2011 – Décembre 2011 : Étude de la qualité des index estimés par SIMM

EDF Branche Commerce Département Analyse et Connaissance Client

- Programmation des méthodes d'estimations SIMM/QE sous SAS à partir des spécifications fonctionnelles.
- Analyse de l'influence d'une migration de SI sur les estimations.

**Mots clés : SAS(Base, macro, connect to oracle), data management, SIMM, QE, statistique descriptive, forte volumétrie**

Mars 2011 – Juin 2011 : Étude des clients multi-sites dans les bases de données EDF

EDF Branche Commerce Département Analyse et Connaissance Client

- Échantillonnage et statistiques inférentielles.
- **Classification / segmentation sur d'importants volumes de données.**
- Définition de vocabulaire spécifique et d'indicateurs statistiques appropriés à l'étude.

**Mots clés : SAS(Base, Stat), SIG, données INSEE, QE, statistiques inférentielles, forte volumétrie**

Novembre 2010 – février 2011 : Étude de l'influence des stratégies marketing sur le profil des clients gaz

EDF Branche Commerce Département Analyse et Connaissance Client

- Profilage d'une population et rédaction d'un document résumant l'étude.

**Mots clés : SAS, excel, statistiques descriptives, QE**

Octobre 2010 : Regroupement de modalités bivariés par la méthode MODL

EDF R&D Innovation Commerciale et Analyse du Marché de l'énergie

- Segmentation de la France et création de variables pour utilisation dans un SIG

**Mots clés : probabilités bayésiennes, SIG, segmentation, Khiops, Khiphen**

Mai 2010 - Septembre 2010 : Reconstruction de courbes de charges à partir d'index semestriels (stage)

EDF R&D Innovation Commerciale et Analyse du Marché de l'énergie

- Qualification et exploitation de données clientèle (particuliers, professionnels).
- Classification et segmentation, régression non paramétrique.
- Développement d'un algorithme expérimental de reconstruction de courbe de charge.
- Rédaction de rapport de stage sous latex.

**Mots clés : SAS(Base, macro, stat), R, data management, latex, courboscope**

Décembre 2009 – Avril 2010 : Classification de données génomiques anonymes (Projet de fin d'étude en collaboration avec l'INRA)

Université de bourgogne

- Synthèse et rédaction d'un mémoire sur les Support Vecteurs Machines(SVM)
- Utilisation de Matlab pour l'implémentation et la restitution graphique

**Mots clés : Matlab, SVM, latex**

**ET AUSSI**

**Mon profil linked In:**

<https://uk.linkedin.com/in/yves-crutain-0b430324>

**Mon blog sur la data science (en anglais):**

[data-laborer.eu](http://data-laborer.eu)

**Mon profil Github:**

<https://github.com/YvesCR>