



# DAMIEN ISAI

## Data Analyst // Scientist

### Me Contacter

Tél : 07 70 32 42 50

Email : [damien236@gmail.com](mailto:damien236@gmail.com)

Linkedin : [damien-isai](#)

Github : [natsunami](#)

Site : [natsunami.github.io/website/](https://natsunami.github.io/website/)

### Présentation

Data Analyst à l'Agence Nationale de la recherche (ANR) rattaché à la Direction des Grands Investissements de l'Etat (DGPIE). Chargé d'analyser, mesurer l'activité, les performances et impacts des Sociétés d'Accélération de Transferts Technologiques (SATT).

Diplômé en ingénierie biomédicale ayant des compétences en analyse et science de données. Toujours prêt à contribuer à de nouveaux projets Data novateurs.

### Compétences

#### • Hard skills:

Python Programming – Data Collection – Data Analysis – Data Visualization – Supervised/Unsupervised Machine Learning – NLP – Deep Learning – Statistics

#### • Tools skills:

Python – SQL – Numpy – Pandas – Tableau | Matplotlib | Seaborn – Scikit learn – Tensorflow | Keras – AWS – Dask – Big Query - Git | Github – Airflow – Docker – FastApi - MS Office -

### Certifications

- Tableau Desktop I & II certification
- SQL Fundamentals (DataCamp)
- Anglais (Avancé – TOEIC 920 – C2)

## Projets

### Missions (Freelance)

- Prédire les mots-clés des requêtes Google générant une marge négative TF-IDF | Over/Under Sampling | Random Forest Classifier | F1-score: 0.76 | Permutation Importance

### Portofolio

#### • Microsoft Malware Prediction

Travailler dans un environnement Big Data (10 M rows & 83 columns) avec Dask et un cloud cluster

#### • Airbnb price prediction

Prédire le prix des airbnb parisiens | MSE: 0.14 | Tuned XGBoost Regressor

#### • Healthcare cost prediction

Prédire les 10 % de patients ayant les coûts de santé les plus élevés et modéliser l'explicabilité | Accuracy: 0.72 | Tuned XGBoost Classifier

### Articles

- French to SQL query API using transformers | FastApi | Docker
- Parallel computing with Dask
- Price prediction using AutoML with AWS SageMaker Autopilot
- How to explain your ML model with SHAP

## Expériences professionnelles

### Data Analyst

Agence Nationale de la Recherche | Février 2021 – Aujourd'hui

Tools : Tableau Desktop, Python, Apache Airflow, Gitlab

- Conception, mise à jour et optimisation de dashboards de suivi de performances, reportings pour l'aide à la prise de décision ; Monitoring des indicateurs et KPI (Tableau Software)
- Traitement des données (ETL pipelines) et automatisation (Python, Apache Airflow, Tableau Prep Builder)
- Analyses statistiques et exploratoires des données
- Démonstration faisabilité (POC) : Traduction et classification binaire de texte avec un hyper-tuned RoBERTa transformers

### Ingénieur de recherche en neuroimagerie

Centre d'Energie Atomique - Neurospin | Février - Juin 2020

Tools : Python ( Numpy, Pandas, Scikit-learn, Matplotlib), AWS

- Développer des scripts python pour la segmentation de régions cérébrales
- Collecte, nettoyage et structuration des données obtenues par IRM
- Contrôler la qualité des données, détecter les outliers
- Mise en place et évaluation de différentes méthodes de Machine Learning et de biostatistiques (régression multiple linéaire et non linéaire)
- Utilisation du Cloud computing pour accélérer le traitement de données ( AWS EC2, S3, SageMaker)
- Data Visualization & reporting

### Analyste de données en neurosciences

INSERM | Décembre 2018 - Juin 2019 ( 400 heures)

Tools : Python (Numpy, Pandas, Statsmodels, Matplotlib), R

- Collecte, nettoyage et structuration de données comportementales
- Mise en place et évaluation de différentes méthodes de biostatistiques
- Data Visualization & reporting

## Etudes

### Data Science Bootcamp

Vivadata | Juin – Août 2019 ( 400 heures)

Programmation Python – Collecte et analyse de données – Machine Learning Supervisé & Non Supervisé – Data Engineering – Deep Learning

### Master en ingénierie biomédicale

Université de Paris – Arts et Métiers – Université PSL | 2018 – 2020

Spécialité: Bio ingénierie et innovation en neurosciences

Mention Assez Bien

Ingénierie – Statistiques – Programmation Python – Neurosciences