



# Lucas POTIN

## Data Scientist | Graphes et Modélisation

Ingénieur et doctorant en data science, je recherche un poste de Data Scientist pour mobiliser mes compétences en modélisation explicable et robuste sur des données complexes, relationnelles ou tabulaires.  
Disponibilité : Septembre 2025.

🏠 Rouen  
📄 Permis B  
✉ lucas.potin.data@gmail.com  
☎ +33 6 51 71 01 81  
🔗 LinkedIn : lucas-potin

## Projets

<https://lucaspotin98.github.io/>

### Projets :

- **DramaPang** : Classification supervisée de pièces de théâtre.
- **API de sirétisation** : Service FastAPI pour retrouver des SIRET.
- **Dashboard de ventes** : Application Streamlit de prévision des ventes.

Technologies : Python, NetworkX, Scikit-learn, XGBoost, FastAPI, Streamlit, CI/CD.

## Compétences

### Mathématiques & Modélisation

Statistiques, Probabilités, Optimisation (linéaire, combinatoire), Théorie des graphes.

### Langages

Python, SQL, C++, CUDA.

### Data Science

Pandas, Scikit-learn, XGBoost, PyTorch, NetworkX.

### Environnement Technique

Git, CI/CD (GitHub Actions), Streamlit, FastAPI, Bash.

### Savoir-être

Adaptabilité, Autonomie, Créativité, Travail d'équipe, Vulgarisation.

### Langues

Français (natif),  
Anglais (C1, TOEIC 970/990).

## Centres d'intérêt

### Tennis

Classement 15/2

### Bridge

Classement 3ème série trèfle

## Expériences professionnelles

### Thèse de Doctorat

Depuis 2021 Laboratoire d'Informatique d'Avignon (Avignon)

- Technologies : Python (Pandas, NetworkX, Scikit-learn, XGBoost), SQL, Git.
- Sujet : Détection de corruption dans les marchés publics à l'aide de réseaux complexes.
- Conception de deux bases open data sur les marchés publics français (FOPPA et BeauAMP), intégrant plus de **300 000** entreprises sirétisées avec **80 % de réussite**.
- Développement d'un framework de classification de graphes basé sur la fouille de motifs fréquents, atteignant un **F1-score de 95 %** sur des cas réels de détection de fraude.
- Obtention du **Prix Science Ouverte 2024** pour la base BeauAMP.
- Valorisation scientifique : **3 publications internationales**.

### Stage Ingénieur

De mars 2020 à septembre 2020 Aubay (Boulogne-Billancourt)

- Technologies : Python (Pandas, Keras), méthodologie agile.
- Amélioration d'un générateur de sessions web synthétiques, en modélisant **3 types d'agents** utilisateurs (aléatoire, rationnel, récurrent) et **2 profils** (client, candidat).
- Conception et entraînement des modèles de classification et de prédiction via des réseaux LSTM ; obtention d'une **précision de 94,5 %** pour l'identification du profil utilisateur.
- Développement d'un dashboard web pour visualiser les données.

### Assistant Chercheur

De juin 2018 à septembre 2018

James Hutton Institute (Dundee, Écosse)

- Technologies : C++, CUDA, Linux.
- Implémentation d'un algorithme de modélisation cellulaire en C++/CUDA pour simuler des processus biologiques sur GPU.
- Réduction du temps d'exécution par un **facteur de 2 à 50**.
- Collaboration au sein d'une équipe **internationale** et **interdisciplinaire**.

## Diplômes et Formations

### Doctorat en Informatique

De 2021 à 2025 Laboratoire Informatique d'Avignon (Avignon)

### Master Sciences et Ingénierie des Données

De 2019 à 2020 Université de Rouen (Saint-Étienne-du-Rouvray)  
Double cursus avec l'INSA

### Diplôme Ingénieur option Génie Mathématique

De 2015 à 2020 INSA Rouen (Saint-Étienne-du-Rouvray)

## Encadrement de projets

### Encadrement de projets de fin d'études.

De 2021 à 2025 INSA Rouen (Saint-Étienne-du-Rouvray)

Apprentissage par renforcement multi-agents pour 6 étudiants.  
Cadrage, suivi technique et validation des livrables.