Lucas POTIN

Data Scientist | Graph Analysis and Modeling

PhD-level data scientist specializing in pattern mining and graph classification. I design interpretable models to detect atypical behaviors in heterogeneous data (public procurement, web sessions, biological networks).

Rouen, France
Driving licence (B)
lpotin888@gmail.com
+33 6 51 71 01 81

Portfolio

https://lucaspotin98.github.io/

Projets :

- **PANG**: Modular framework for graph classification.
- **SIRET API**: FastAPI service to retrieve French company identifiers.
- **Sales Dashboard**: Streamlit app for sales forecasting.

Skills

Mathematics & Modeling
Statistics, Probability, Optimization (linear, combinatorial), Graph Theory.

Languages
Python, SQL, C++, CUDA.

Data Manipulation
Pandas, NumPy, SciPy.

Graphs and Classification
NetworkX, PyTorch Geometric, Pattern mining.

Machine Learning
Scikit-learn, XGBoost, PyTorch.

Deployment and APIs
Streamlit (dashboards), FastAPI (APIs).

Technical Environment
Git, LaTeX, Bash, Linux.

Soft Skills
Adaptability, Autonomy, Creativity, Teamwork, Scientific communication.

Languages

French
Native

English
Fluent (C1) – TOEIC: 970/990

Interests

Tennis
French ranking 15/2

Bridge
French ranking – 3rd series (clubs)

Work experience

PhD Thesis
Since 2021 Laboratoire d'Informatique d'Avignon Avignon

- Topic: Detecting corruption in public procurement using complex networks.
- Designed two open datasets (FOPPA and BeauAMP) on French public contracts, integrating **300,000+** SIRET-matched companies (**80% success**) from noisy, heterogeneous data.
- Developed a graph classification framework based on frequent pattern mining, reaching **95% F1-score** on public benchmarks and real-world fraud detection cases.
- Awarded the **2024 Open Science Prize** for the BeauAMP database.
- Technologies: Python (Pandas, NetworkX, Scikit-learn, XGBoost), SQL, Git.

Engineering Internship
From March 2020 to September 2020 Aubay Boulogne Billancourt

- Improved a synthetic web session generator by modeling 3 user types (random, rational, recurrent) and 2 profiles (client, applicant).
- Designed and trained behavior prediction and classification models using LSTM networks on simulated sessions, achieving **94.5% profile identification accuracy**.
- Technologies: Python (Pandas, Keras), méthodologie agile (Kanban).

Research Assistant
From June 2018 to September 2018
James Hutton Institute Dundee (Écosse)

- Implemented a parallel cellular modeling algorithm in C++/CUDA to simulate biological processes on GPU.
- Reduced runtime by a **factor of 2 to 50** through low-level optimizations.
- Collaborated in an **interdisciplinary** research team (computer science, agronomy, biology).
- Technologies: C++, CUDA, Linux.

Education

Engineering Degree – Mathematical Engineering
From 2015 to 2020 INSA Rouen Saint-Étienne-du-Rouvray

Master's Degree – Data Science & Engineering
From 2019 to 2020 University of Rouen Saint-Étienne-du-Rouvray
Double degree with INSA.

Scientific Baccalauréat (Math major), European German section
2015 Lycée Georges Brassens Neufchâtel-en-Bray

External Links

- GitHub : LucasPotin98
- Linkedin : lucas-potin
- Google Scholar : Lucas Potin