

# Lucas Potin

+33651710181 | [lucas.potin.data@gmail.com](mailto:lucas.potin.data@gmail.com) | [linkedin.com/in/lucas-potin](https://linkedin.com/in/lucas-potin) | [lucaspotin98.github.io](https://lucaspotin98.github.io)

## SUMMARY

---

Data scientist and final-year PhD candidate with a strong engineering background, specializing in graph analytics, explainable machine learning, and complex data processing. Experienced in designing robust data pipelines, building production-ready APIs and dashboards, and delivering actionable models for fraud detection and forecasting. Available from September 2025 for data science roles in industry or applied research.

## EDUCATION

---

<b>PhD in Computer Science</b> <i>Université d'Avignon, LIA Laboratory</i> <ul style="list-style-type: none"><li>Thesis: <i>Fraud detection in public procurement using complex networks</i></li><li>Designed interpretable graph classification methods for irregularity detection</li></ul>	Sept. 2021 – Sept. 2025 (expected) <i>Avignon, France</i>
<b>Master's in Data Science</b> <i>University of Rouen (with INSA Rouen)</i>	Sept. 2019 – Sept. 2020 <i>Rouen, France</i>
<b>Engineering Degree in Applied Mathematics</b> <i>INSA Rouen – Department of Mathematical Engineering</i>	Sept. 2015 – Sept. 2020 <i>Rouen, France</i>

## EXPERIENCE

---

<b>PhD Researcher in Data Science</b> <i>Université d'Avignon, LIA Lab</i> <ul style="list-style-type: none"><li>Designed and maintained two open data graph-based datasets for public procurement analysis (FOPPA, BeauAMP), integrating over 300,000 SIRET-matched companies with 80% success rate</li><li>Developed a graph classification framework using frequent subgraph mining, achieving 95% F1-score on real-world fraud detection cases</li><li>Awarded the 2024 French Open Science Prize for the BeauAMP dataset</li><li>Wrote 4 international publications on interpretable graph learning and fraud detection</li></ul>	Sept. 2021 – Present <i>Avignon, France</i>
<b>Data Science Intern</b> <i>Aubay</i> <ul style="list-style-type: none"><li>Improved a synthetic web session generator by modeling 3 agent types (random, rational, recurrent) and 2 user profiles (client, applicant)</li><li>Built and trained LSTM-based classifiers to predict user behavior with 94.5% accuracy</li><li>Developed a web dashboard to visualize user sessions and prediction outputs</li></ul>	Mar. 2020 – Sept. 2020 <i>Boulogne-Billancourt, France</i>
<b>Research Intern (HPC Modeling)</b> <i>James Hutton Institute</i> <ul style="list-style-type: none"><li>Implemented a GPU-accelerated cellular modeling algorithm using C++ and CUDA to simulate biological processes</li><li>Reduced execution time by a factor of 2 to 50 depending on model size</li><li>Collaborated within an international and interdisciplinary research team</li></ul>	June 2018 – Sept. 2018 <i>Dundee, UK</i>

## PROJECTS

---

<b>DramaPang</b>   <i>Python, Scikit-learn, NetworkX, Streamlit</i> <ul style="list-style-type: none"><li>Built a supervised classifier to distinguish between comedies and tragedies based on character interactions in classic plays</li><li>Modeled character co-occurrence networks and extracted graph-based features</li><li>Deployed an interactive web app using Streamlit [App]</li><li>Source code available on GitHub</li></ul>
<b>Siretizator API</b>   <i>Python, FastAPI, SQLAlchemy, Docker</i> <ul style="list-style-type: none"><li>Developed a public API to retrieve the French SIRET number of public entities based on their name and city</li><li>Deployed the app with automatic testing and CI/CD on Render [App]</li><li>Source code available on GitHub</li></ul>
<b>Retail Sales Forecast Dashboard</b>   <i>Python, XGBoost, Prophet, Streamlit</i> <ul style="list-style-type: none"><li>Simulated retail transaction data and predicted future sales using time series models</li><li>Explored product affinity using a co-purchase graph and community detection</li><li>Deployed a Streamlit dashboard with dynamic forecasting and graph analysis [App]</li><li>Source code available on GitHub</li></ul>

## TECHNICAL SKILLS

---

**Languages:** Python, SQL, C++, Bash

**Libraries & Frameworks:** pandas, scikit-learn, XGBoost, PyTorch, NetworkX, FastAPI, Streamlit

**Tools:** Git, Docker, GitHub Actions, VS Code, Linux, LaTeX

**Data Science:** Graph mining, Machine learning, Explainable AI, Time series forecasting

**Concepts:** Graph theory, Optimization, Statistics, HPC (CUDA)

## LANGUAGES

---

**French:** Native proficiency

**English:** Fluent (C1 – TOEIC 970/990)

**German:** Intermediate (B1, previously B2–C1)

## SELECTED PUBLICATIONS

---

Potin, L., Figueiredo, R., Labatut, V., & Largeron, C. (2023). *Pattern mining for anomaly detection in graphs: Application to fraud in public procurement*. In *ECML PKDD 2023*.

Potin, L., Labatut, V., Morand, P.-H., & Largeron, C. (2023). *FOPPA: an open database of French public procurement award notices from 2010–2020*. *Scientific Data*, 10(1), 303.

Potin, L., Figueiredo, R., Labatut, V., & Largeron, C. (2025). *Pattern-Based Graph Classification: Comparison of Quality Measures and Importance of Preprocessing*. *ACM TKDD* (in press).

Full list available at: [scholar.google.com/citations?user=1\\_SXGLMAAAAJ&hl](https://scholar.google.com/citations?user=1_SXGLMAAAAJ&hl)

## TEACHING & SUPERVISION

---

### Capstone Project Supervisor

2021 – 2025

INSA Rouen

Rouen, France

- Supervised final-year project on multi-agent reinforcement learning for a group of 6 engineering students
- Defined the technical scope, provided regular feedback, and validated deliverables

### Undergraduate Intern Supervisor

2023

Université d'Avignon

Avignon, France

- Supervised a 3-month Bachelor's intern working on data cleaning and SIRET matching for the FOPPA procurement database
- Introduced open data workflows and reproducible research practices

## EXTRA-CURRICULAR ACTIVITIES

---

**Tennis:** Competitive player – ranked 15/2 in the French national classification system (UTR 8)

**Bridge:** Tournament player – ranked 3<sup>rd</sup> series