Armel Soubeiga

Machine Learning Research Engineer | GenAI | Explainable AI

📱 (+33) 768249323 ✉ armelsoubeiga93@gmail.com in armel-soubeiga ⌂ armelsoubeiga
🌐 armelsoubeiga.github.io 🎓 ResearchGate 🏠 France

I am interested in real-world applications of Explainable Machine Learning, GenAI, and Data Mining.
@ I contribute to open source software. 📖 Author of Web Scraping with R.

Professional Experience

Safran Aircraft Engines   **Apr. 2025 – Present**
Machine Learning & GenAI Engineer   *Corbeil-Essonnes, France*

- Developed an AI system for supplier risk detection and prediction.
- Engineered end-to-end AI agents using multimodal LLMs to ingest and reason over multiple data sources (.pdf, .xlsx, .docx, .pptx, .msg, images).
- Implemented a advanced RAG and knowledge graph pipelines for document information retrieval.
- Built Named Entity Recognition pipeline using fine-tuned LLMs and PromptNER.
- Engineered a Neo4j graph database and dynamic supply chain visualizations.
- Deployed scalable LLM agents with LangChain on AWS.

LIMOS   **Apr. 2022 – Apr. 2025**
Research Engineer in Machine Learning   *Clermont-Ferrand, France*

- Conducted research on collaborative multi-view evidential clustering for uncertain time-series trajectory analysis.
- Applied methods to categorical, discrete, and uncertain time-series data.
- Leveraged advanced statistical and computational techniques to enhance clustering accuracy and efficiency.
- Developed XAI approaches for unsupervised model evaluation and interpretability.
- Assisted teaching in Python, Shell, SQL, data mining, and supervised interns.

RCTs   **Sept. 2019 – Mar. 2022**
Data Science Consultant   *Lyon, France*

- Developed healthcare pathway analysis tools using R, Python, and SQL:
  * Analyzed patient care trajectories with multidimensional sequential approaches.
  * Applied hard and fuzzy clustering and regression to identify trajectory typologies and patient profiles.
  * Performed dimensionality reduction and variable selection (unsupervised Random Forest, Laplacian Score, Spectral Score) and association rule analysis.
  * Implemented supervised ML algorithms to predict disease patterns and hospital admissions.
- Utilized Big Data tools (Spark SDK) for secondary health and national database analysis.
- Created interactive visualizations and reports with PowerBI, Django, Flask, RShiny; managed SQLite and Postgres databases.
- Defined client requirements and proposed analytical strategies and plans.
- Collaborated on technical documentation for tool validation and maintenance.
- Led weekly meetings and presentations to track project progress.

World Bank   **Dec. 2016 – Nov. 2017**
Data Analyst Consultant   *Burkina Faso*

- Designed data collection forms with CSPro and ensured QC.
- Automated data validation and processing scripts (Bash) for STATA, SQL, and R compatibility.
- Developed batch scripts for Windows and Android to sync data across 200+ tablets.
- Collaborated on statistical analysis plans and coded analyses in R.
- Built data visualization and interactive dashboards using RShiny.

Education

University Clermont Auvergne   **Apr. 2022 – Apr. 2025**
PhD in Computer Science & Machine Learning   *Clermont-Ferrand, France*

Université Grenoble Alpes   **Sept. 2017 – Jul. 2019**
MSc in Statistics & Data Science   *Grenoble, France*

Université Nazi Boni   **Sept. 2013 – Jul. 2016**
BSc in Statistics & Computer Science   *Burkina Faso*

Projects

NEWSWacafi | *Python, R, Web Scraping, Azure Functions, GitHub, API, PowerBI*   **Jan. 2024**
- Partnered with the Danish Refugee Council to develop NEWSWacafi, an innovative social media monitoring tool for the Sahel region.
- Implemented a real-time information gathering system across news outlets and social platforms.
- Delivered APIs and SDKs in Python and R for data access, sentiment analysis, and thematic extraction.
- Built a PowerBI dashboard for data exploitation and insights.

evclust | ⬡ | *Python Library, PyPI, GitHub*   **Sept. 2023**
- Developed evidential partitioning algorithms based on Dempster–Shafer theory for robust uncertain data analysis.
- Packaged a performant Python library for evidential clustering and published on PyPI for easy adoption.
- Authored comprehensive documentation with detailed usage examples.
- Established contribution guidelines and maintained the library for ongoing development.

Relevant Skills

- **AI/ML**: LLMs, RAG, Transformers, Deep Learning, Explainable AI
- **Data Science**: Clustering, Time Series, Predictive Modeling, Uncertainty
- **Engineering**: MLOps, Big Data, Feature Engineering, Optimization
- **Tools**: Python, R, SQL, TensorFlow, PyTorch, Spark, AWS, Neo4j
- **Domains**: NLP, NER, Trajectory Analysis, Dimensionality Reduction

Technical Skills

- **Languages**: Python, R, SQL, SAS, Bash
- **Frameworks**: Scikit-Learn, TensorFlow, PyTorch, Spark, LangChain
- **DevOps**: GitHub/GitLab, CI/CD, Docker, AWS
- **Data**: SQL/NoSQL, PowerBI, Data Mining

Involvement / Extracurricular Activities

- Activities — 🏃 ⚽ 🚴
- French National Research Agency Fellowship
- **Teaching Assistant:** [Teaching Page](#)

- **1st Place,** Best Rural Project (2020): poultry farming improvement using AI
- **Online Courses & Certifications:** [See More](#)

Selected Publications & Conferences

Armel Soubeiga, Violaine Antoine, Alice Corteval, Nicolas Kerckhove, Sylvain Moreno, Issam Falih et Jules Phalip. *Clustering and Interpretation of Time-series Trajectories of Chronic Pain using Evidential c-means*, Expert Systems With Applications, Vol 260, pp 125369, 2025.

Armel Soubeiga, Violaine Antoine et Sylvain Moreno. *Multi-View Relational Evidential C-Medoid Clustering with Adaptive Weighted*, IEEE 11th International Conference on Data Science and Advanced Analytics (DSAA), Oct 2024, San Diego, USA. pp.1-10.