Mathilde **BOLTENHAGEN**, PhD

mathilde.boltenhagen@outlook.com | +33.6.80.96.51.40 | @BoltMaud

EDUCATION

Nov. 2018 – Oct. 2021 — **PhD student – Process Mining**
Paris-Saclay University, ENS, LMF

Process Instance Clustering based on Conformance Checking Artefacts
Process Mining is a recent field that aims at mining process models representing systems of organization. My thesis joined Verification techniques and clustering approaches to address two broad problems of Process Mining:

1) how to relate a process model (represented as a Petri net) and a set of log sequences?
2) how to identify meaningful groups (clusters) of log sequences based on similarity of behavior?

My contributions generate novel insights by addressing these problems thanks to variants of alignment artefact, known in the field, like the multi-alignment artefact. For each introduced method, I presented a SAT encoding given the optimal solution and an A* algorithm to deal with the complexity of the problems.

Sept. 2016 – June 2018 — **Data Science Master**
Lyon 1 Claude Bernard University

Sept. 2012 – June 2016 — **Computer Science Bachelor**
2-year degree at Strasbourg University, then learning distance at Aix-Marseille University

WORK EXPERIENCE

Sept.2018 – June 2021 — **Teaching Assistant and Substitute Teacher**
Paris Dauphine University and Sorbonne Universities

- Business Process Analytics to Ms. students, created all course supports from scratch (lectures, tutorials, labs), teached the field during two years which allows me to improve the content.
- Relational Databases to BSc. students and PhD students, assisted professors (tutorials, labs)

Python – SQL – Process Mining

Mars – Sept.2018 — **Big Data researcher**
Internship at Trimane

- Associated in the intern laboratory CBI², bridged the relationship gap with the consultants by proposing and organising a presentation on Intelligence Artificial (IA)
- Addressing the problem of schema extraction from NoSQL databases, proposed to use Natural Language Processing (NLP) techniques to revelant relations between attribute names
- Weekly presented some related scientific papers, my progress and future works

Word2Vec – MongoDB – Java – Python

Jul. – Aug. 2017 — **Data scientist**
Optional internship at Actinvision

- Assessed correlations between variables using linear regressions in a health application
- Extracted meaningful visualizations of datasets
- Integrated data from SQL to PostgreSQL

Python – Tableau software – SQL/PostgreSQL

GENERAL SKILLS

Data Mining | Natural Language Processing | Machine Learning | Deep Learning | Big Data | Data Visualization | Agile Methods | Sequence Analytics | Process Mining

TECHNICAL SKILLS

Python (sklearn, pandas, pytorch, transformers, pm4py,…), **Java**
SQL, MongoDB, Spark/Hadoop, AWS
Linux, shell, Docker
Javascript (D3.js)

SELECTED PROJECTS

2017 – today — **Kaggle projects**

- Following the state-of-the-art IA problems on various applications, learning new techniques by trying other's contributions and reading the literature of different fields.
- Training data manipulation, EDA and ML techniques by taking part of competitions
- Satisfying my curiosity on NLP and DL problem solving

NLP – Python – Transformers + Huggingface (Bert, detoxify, biobert, bern) – Pytorch – sklearn

2019 – 2021 — **pm4py & da4py**

- Contributed to and created Python librairies for Process Mining
- Implemented my PhD works in open source tools, unit tests

Python – git – pytest

2017 — **Analyze the sky using Spark (team of 2)**

- Accessed to data on hdfs servers, created map/reduce jobs to compute distances between objects detected in the sky, used kmean algorithm to obtain clustering of these objects
- Developed a pytest tutorial to share the knowledge acquired during the project

Python – PySpark – Pytest – Hdfs (Hadoop)

2017 — **Politic Twitter Analysis (team of 2)**

- Supervised research project, studied the tweets of politicians during the previous presidential compaign
- Applied various text analysis and learned the Latent Dirichlet Allocation algorithm

Text analysis - Python/Knime – LDA – D3.js

Nov. 2015 – Aug. 2016 **Developer**
Internship + 1 month contract at A2micile

- Contributed to existing web applications, added essential automations which made the payroll department gaining days of work, added a map tool for localized assistants, listened to user need
- Developped a role security management for the growth of the company
- Improved data heterogeneity of locations and textual information
- Started a new web application

All my works are still used at the time of this application.

Git – Laravel – JEE/Zend – PHP – MySQL

Oct. 2014 – May 2015 **Au pair**
In New Zealand

- Open minded, improved my English

Jun. 2014 – Jul. 2014 **Research assistant**
Optional internship at CNRS

- Studied complier optimization

Ocaml

CERTIFICATIONS

2022 AWS / Big Data - Cegefos
2021 Sequence Models – Coursera

SIDE PROJECTS

Journaux de Famille - Notebook Creation
Bo.h.u. - Generative Art

LANGUAGES

French: native
English: good working proficiency
Spanish: currently learning

INTERESTS

Running (marathons), hicking, swimming, biking
Painting (oil based), drawing on pad
Reading books (philosophy, history)
Listening to Podcast (sport, health, psychology)

SELECTED PUBLICATIONS

My teams and I obtained 9 papers accepted in international workshops, conferences and journals. I selected two of them below.

- Mathilde Boltenhagen, Thomas Chatain, Josep Carmona, Generalized Alignment-Based Trace Clustering of Process Behavior, (ICATPN'2019), extended in Information Systems' Journals.
  - Introducing a clustering algorithm of sequences based on a Petri net, bringing key criteria of quality
- Mathilde Boltenhagen, Benjamin Chetioui, Laurine Huber, An Alignment Cost-Based Classification of Log Traces Using Machine-Learning, (ML4PM@ICPM).
  - Used Recurrent Neural Network (LSTM) in Process Mining context