Colin Troisemaine

PhD, Data Scientist

colin.troisemaine@gmail.com | linkedin.com/in/colin-troisemaine | github.com/ColinTr

Education

IMT Atlantique — Brest, France
Doctor of Philosophy in Computer Science — *Oct. 2021 – Sept. 2024*

Sherbrooke University — Sherbrooke, Canada
Master's Degree in Computer Science — *Sept. 2020 – Aug. 2021*

Polytech Tours — Tours, France
Engineering Degree in Computer Science — *Sept. 2016 – Aug. 2021*

Experience

Data Scientist — Jan. 2025 – Today
AMIAD (Agence Ministérielle pour l'IA de Défense) — *Bruz, France*

PhD in Machine Learning — Oct. 2021 – Sept. 2024
IMT Atlantique / Orange Innovation — *Lannion, France*

- Developed 4 original techniques to discover novel classes in an unlabeled set of tabular data
- Presented complex concepts to diverse audiences through accessible presentations at conferences
- Realized an extensive state of the art on novel class discovery techniques, which received 15 citations in one year
- Collected and curated a dataset of 500,000 real internet access faults and engineered 700 new features

End-Of-Study Research Internship — May 2021 – Sept. 2021
Orange Innovation — *Lannion, France*

- Designed a new way to create descriptive features to improve regression performance
- Developed a parametric framework that can make use of 4 different classification models to generate features to improve the performance of 5 regression models
- Published an analysis of the results along a positioning to the SOTA in a 12-page conference paper

Projects

Interactive Interface for Novel Class Discovery | *Python, Flask, React, Plotly* — June 2021

- Developed a web application using Flask to serve a REST API with React as the frontend
- Implemented various clustering and novel class discovery algorithms
- Created data visualization functionalities with Plotly and PDF decision trees
- Featured this work in a paper at the ECML PKDD 2023 Demo Track

PracticalNCD | *Python, PyTorch, Git, Wandb, Pandas* — Dec. 2023

- Designed a new state-of-the-art deep learning model for the discovery of novel classes
- Optimized the hyperparameters of 3 complex models on 7 datasets with a custom experiment tracking framework
- Compared 11 scores to estimate the number of clusters in an unlabeled set in different scenarios

Publications

"A Practical Approach to Novel Class Discovery in Tabular Data", Data Mining and Knowledge Discovery journal — 2024

"Novel Class Discovery: an Introduction and Key Concepts", under review at Neural Computing and Applications (NCAA) journal — 2023

"A Method for Discovering Novel Classes in Tabular Data", IEEE International Conference on Knowledge Graph (ICKG) — 2022

"Construction de variables à l'aide de classifieurs comme aide à la régression", Extraction et Gestion des Connaissances (EGC) — 2022

Technical Skills

Foreign languages: French (native language), English (TOEIC level C1 - Expert, 965/990 points, 2018)
Programming languages: Python, SQL (Postgres, Bigquery), JavaScript, HTML/CSS, LaTeX
Frameworks: PyTorch, Pandas, Numpy, Scikit-learn, Jupyter, Matplotlib, Weights&Biases, React, Node.js, Express.js
Developer Tools: Git, Docker, Google Cloud Platform, PyCharm, IntelliJ, PhpStorm, Eclipse
Hobbies: Chess, guitar, 3D printing, FPV drone flight