Yann Hoffmann

Email : contact@yannhoffmann.com
http://www.linkedin.com/in/yann-hoffmaaann
Mobile : (+352) 621 486 472

EXPERIENCE

- **SnT Laboratory** — Luxembourg, Luxembourg
  PhD Candidate — *Mar. 2023 – Present (6. mo)*
  - **Master-Grade Class Giving**: Giving a Master-level class in Applied Machine Learning for the Data Science master. The course covers feature engineering, Machine Learning (ML) optimization, unsupervised learning, and more.
  - **Data Augmentation for Fraud Detection**: Improving bank models by implementing state-of-the-art data augmentation techniques from academia.
  - **Infrastructure Building**: Managing Kubernetes cluster for MLOps.

- **Talkwalker** — Luxembourg, Luxembourg
  Data Scientist — *Sept. 2021 – Feb. 2023 (17 mo.)*
  - **Multilingual Email Content Extraction**: Was tasked with developing an end-to-end ML solution in Python to extract multilingual content from emails. Created an annotated email dataset, then did feature engineering, built a model, tuned it, and deployed it as a gRPC service. The final model is an autoregressive XGBoost ensemble with 95 AUC running in production.
  - **Audience Segmentation**: Auto-segmentation of Twitter profiles based on demographics and user interests using umap, hdbscan, and sbert transformer. Interests were further expanded with the help of GPT-3.
  - **Word Clustering**: Keyword clustering using FastText embeddings, hdbscan clustering, and KDTree. This has application in finding general concepts in word clusters and simplifying our user interface.

- **Talkwalker** — Luxembourg, Luxembourg
  Data Science Intern — *Feb. 2021 – June 2021 (6 mo.)*
  - **Infrastructure Optimization**: Prototyped reinforcement learning (RL) methods to optimize resources of a cluster in Python. Built a simulator and implemented the agents in Java with ZeroMQ to evaluate my methods. The RL agents used a variant of Q-learning and XGBoost as its state-action value approximator.
  - **End-of-Master Thesis**: Described the research process, the architecture, and defended orally my results. (18.5/20)

- **University of Luxembourg** — Luxembourg, Luxembourg
  DevOps Research Assistant — *Nov. 2020 – Feb. 2021 (4 mo.)*
  - **Data Collection**: Created a solution to scrape data from university GitLab projects using Prometheus, Grafana, then created a dashboard with Plotly dash. Co-wrote manuscript on best practices for DevOps.

- **Paris-Créteil University** — Paris, France
  Machine Learning Research Assistant — *Summer 2019 (2 mo.)*
  - **Multi-Modal Emotion Recognition**: Implemented RNN variants in TensorFlow to predict human emotions. Co-wrote manuscript detailing an approach to speech processing.

EDUCATION

- **University of Luxembourg** — Luxembourg, Luxembourg
  Master of Science in Computer Science; Data Science Specialization — *Aug. 2019 – Sept. 2021*

- **Sorbonne University** — Paris, France
  Bachelor of Science in Computer Science; Mathematics Minor — *Sept. 2017 – June 2019*

- **CPGE MPSI, Pierre de Fermat** — Toulouse, France
  Mathematics, Physics, and Computer Science — *Sept. 2016 – July 2017*

Projects

- **AWS Tuner**: MLOps tool available on my github to automate model evaluation and hyperparameter tuning in AWS using Python, Wandb, and Docker. Authored a companion article on TowardsDataScience promoting my work
- **Paperless Hoster**: SaaS project to improve an open-source document management application by simplifying the interface and hosting it in the cloud.
- **University Projects**: Big Data framework benchmarking, Neural Architecture Search, CNN filter visualization.

Other

- **Technologies**: Python, PyTorch, Java, certified AWS, SQL, Scala, NLP, DevOps, Docker

- **Natural Languages**: EN (Fluent), FR (Native), GE (C1), LU (B2)