Quentin-Gabriel Thurier

New Zealand | quentin.thurier@gmail.com |

I thrive in cross functional teams working at the interface between industry and academic research in order to craft and develop great Machine Learning based products.

Experience

Senior ML Engineer — **Decoded Health** — *June 2023 – Present*

- Reduced human-in-the-loop costs by rolling-out a patient conversation automation summarisation service with FastAPI.
- Established standards for data versioning and machine learning services observability on a Google Cloud Platform stack.
- Improved clinical named entity recognition F1-score by 23% using OpenAI API.

Senior Applied Scientist — **Xero** — *November 2020 – May 2023*

- Saved 1.5M small businesses 10,000+ hours of manual data entry per month by rolling-out a machine learning service to make bank reconciliation recommendations (4M+ requests per day).
- Led the research (50+ models experiments) on deep learning applied to bank statements using TensorFlow and AWS.
- Simplified a PySpark data pipeline by halving the codebase and driving its migration to Snowflake and Prefect.
- Streamlined and aligned models documentation across 4 cross-functional pods by integrating Jira & Confluence.
- Patented 3 inventions in collaboration with the intellectual property team.
- Supervised 1 applied scientist and 3 graduates among which 2 became ML engineers.

NLP Engineer — **PredictHQ** — *May 2019 – November 2020*

- Reduced the overall records duplication rate to 3% by rolling-out a random forest classifier for event entity resolution with scikit-learn, giving the company a competitive edge in the events data providers space.
- Improved by a factor of 10 the level of granularity of events categorisation by rolling-out an ontology-based classifier using spaCy, improving downstream machine learning models performance.
- Led a 3-member team to deliver a gradient boosting regression model for concert events attendance prediction that decreased the mean absolute error down to 1000 attendees.

Data Scientist — **Orion Health** — *2017 – April 2019*

- Rolled-out an ensemble classifier (scikit-Learn) to predict post stroke outcomes into Auckland North Shore Hospital, achieving an area under the ROC curve of 94% on inpatient death outcomes.
- Managed a $300,000+ research budget and led a cross-functional 8-member team to deliver 2 interpretable machine learning models and 2 peer-reviewed publications.
- Started off 3 product features by delivering deep learning prototypes with TensorFlow and Keras.
- Enabled deep learning research in the team by building a GPU workstation with a $8,000 budget.
- Presented and shared my work at 3 health informatics conferences and 2 ministries.
- Mentored 3 data scientists and supervised 4 interns among which one was hired.

Data Scientist — **Qrious** — *2016*

- Delivered geospatial insights using mobile phone activity and public Wi-Fi data with Python & Spark.

Data Manager — **NetBooster** — *2015*

- Implemented digital marketing solutions with Python leveraging web server logs data on Google Cloud Platform.

Statistician — **Société Générale** — *2011 – 2013*

- Implemented a network analysis and unsupervised machine learning software to prevent unauthorized trading.

Skills

Languages	Python, Bash, SQL, R
Technologies and Tools	TensorFlow, scikit-learn, spaCy, Snowflake, Spark, AWS, Google Cloud Platform, Git, Docker
Science	Deep Learning, ML, NLP, Statistics, Probability, Linear Algebra

Education

Post Master's Degree (ML)	Telecom ParisTech	*2014*
Master's Degree (Statistics)	ENSAI	*2005 – 2010*

Patents

Method, Program, and Apparatus for Processing Sensitive Data *2023*

Showed how an ensemble classifier trained on 5 years of stroke hospital records can predict inpatient health coutcomes blabla blaba blaaa.

Methods and Systems for Training Attribute Prediction Models *2023*

Showed how an ensemble classifier trained on 5 years of stroke hospital records can predict inpatient health coutcomes.

Transaction Data Processing Systems and Methods *2022*

Showed how an ensemble classifier trained on 5 years of stroke hospital records can predict inpatient health coutcomes.

Research

Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator (*JAMIA*) *2019*

Surveyed 1000+ physicians to investigate the association between physician understanding of machine learning outputs, their ability to explain these to patients, and their willingness to trust these, using various explainability methods. This article has been cited 98 times.

Smart MedRec: Using machine learning for reading dose instructions and incorporating this in a software (*HINZ*) *2019*

Introduced a promising avenue to improve medication reconciliation, hence patient safety, by achieving 81% accuracy with a deep learning model trained to automatically parse medication dosage instructions.

Inspecting a machine learning based clinical risk calculator: a practical perspective (*IEEE*) *2019*

Suggested a modelling workflow to facilitate the use of model agnostic explainability methods and applied this to a machine learning mortality risk model.

Interpretable machine learning for healthcare (*HINZ*) *2018*

Showed how the Python data science stack can be used to help clinicicans to audit black-box machine learning models, using the example of a hospital risk readmission model.

Improving clinical named entity recognition with transfer learning (*HIC*) *2018*

Demonstrated the effectiveness of transfer learning for clinical named entity recognition by training a deep learning model on a heuristically labelled large dataset then fine-tuning it with a manually annotated dataset and comparing its performance with clinical concept extraction softwares.

Health outcomes prediction engine for stroke (*HINZ*) *2017*

Presented how machine learning can aid the communication of prognosis by training an ensemble classifier on 5 years of stroke hospital records that achieved an area under the ROC curve of 94% when predicting inpatient death outcomes.

New Zealand health data review (*HINZ*) *2017*

Made health data easier to access for research by reviewing publicly available datasets and their relationships.

Interests