Thierry Jean

M.Sc., Computational Medicine – Final Year

thierry.jean@umontreal.ca | LinkedIn
(514) 771-0457 | GitHub
Montréal, Canada | tjean.me

Work

Data Science Intern (recommendation) **@MoovAI** (Montréal) **May 2022—Aug 2022**

- Worked on **dynamic external bidding** for a large ad exchange. On each bid (repeated game), the ad exchange applies a % revenue split between itself and the publisher. Our goal was to **maximize revenue by varying the % split**;
- Proposed **novel algorithms**. The first iteration was a **discrete matrix-based greedy optimization** to find optimal % split on bid bins. Then, **proposed research avenues** for continuous estimation on each axis using linear estimators and Poisson sampling.
- **Backtested** and **benchmarked** algorithms against the **current approach** and **greedy baseline (MLFlow);**
- Estimated **potential lift or drawdown** and provide outcome scenarios including business metrics and visualizations to client;
- **Reduced pipeline latency from 20mins to 5mins** by leveraging pivot tables in-database **(Snowflake, SQL, Airflow)** and **vectorizing algorithms** in inference pipeline **(Kedro, Python).**

Instructor Assistant—ML & Data Science **@Sphere** (San Francisco; remote) **Aug 2022—Present**

- Worked with instructors to **build and deliver 3 different live courses**: *NLP with transformers* (w/ Hugging Face), *Applied causal inference* (w/ Rob Donnelly from ArenaAI), *ML-powered search* (w/ Doug Turnbull from Reddit);
- Produce **code demos and practical exercises**, and provide supplementary material for learners;
- Advise on ML and data science **curriculum** (future courses, instructor sourcing, market trends).

Research

Student Research (Undergrad and Master's) **@Mental Health Research Institute** (Montréal) **Sep 2020—Present**

- Forecast **psychiatric symptoms** based on **patient smartphone sensor data** with a focus on interpretability;
- Apply sequence methods: RNN with **ordinal regression** (CORN) and **explainable predictions** (TimeSHAP);
- Apply tabular methods: **XGBoost** with **SHAP** and **Explainable Gradient Boosting Machines** (EGBM);
- Engineer features for **time series** data at various irregular sample rates from **10+ tables and millions of rows**;
- Setup **Metaflow** and deploy **batch pipelines** on **AWS** with **Weights&Biases** for scalable ML experimentation.

Projects (more on my website tjean.me/projects/)

Semantic Blog Search (repo) *Explore blog posts using the Weaviate vector search engine via a Streamlit webapp.*

Parse HTML with XML to retrieve RSS feeds and download blog posts. **Embed posts with Weaviate** and the UI allows for keyword and/or semantic search. I want to add orchestration to automate RSS querying.

Voice Assistant (repo) *Launch arbitrary Python functions using your microphone*.

Connect a client audio input stream to a listening server hosting a Whisper model from OpenAI to transcribe audio. Words following the key phrase are parsed and "fuzzily" evaluated against a library of registered commands on the client.

Development Container (repo) *Enable reproducible experimentation via a Python + Linux container*

Allows for YAML-based container config for simple version control. Multistage Docker build with mamba and conda-pack for fast build and lightweight image. Jupyter servers can be accessed remotely from outside the container.

Education

M.Sc., Computational Medicine, Université de Montréal w/ courses at MILA (GPA: A) **Sep. 2021 – Sep. 2023**
Master's Research Scholarship (2x), both Canadian and Québec's Health Research funds 2021-2023
Recruitment Scholarship, Faculty of Medicine, Université de Montréal 2021

B.Sc. Hon., Psychology, Université de Montréal (GPA: A+) **Sep. 2018 – Apr. 2021**

Other

I'm reading **Effective Data Science Infrastructure** by Ville Tuulos. This summer, I finished **Designing Machine Learning Systems** by Chip Huyen and attended the **MLx Health Summer School** by Oxford and CIFAR. I've been playing guitar since I was 11 years old and I'm currently learning Caprice no. 16 by Paganini.