Liam Toran

liam.toran@gmail.com • San Francisco, CA • liamtoran.com • US Citizen • LinkedIn

WORK EXPERIENCE

Nav *San Francisco, CA*

Staff Data Scientist **June 2023 - Present**

- Designed a time-series forecasting model to predict users' financial balance. The model uses state-of-the-art Transformer encoding/decoding techniques and is deployed in real-time with errors averaging less than 300$.
- Built end-to-end an industry classification system using natural language processing, unlocking key attributes for >93% of our users, leading to improved targeting, personalization, and revenue from referral funnels.
- Applied phased rollout and A/B testing to 4 new features and models to boost business impact and revenue.
- Enhanced existing models with confidence intervals and explainability features, improving trust and model interpretability.
- Provided data-driven analysis and insights that influenced key cross-functional business decisions for our product, collaborating effectively with product managers and lead stakeholders.

Flowcast *San Francisco, CA*

Senior Data Scientist **March 2022 - June 2023**

- Led the development of a bank transaction classification system using weak supervision, natural language processing, and deep neural networks, accuracies surpassing the market-leading provider (Plaid API) by 21% [arXiv paper].
- Trained, fine-tuned and validated time series and text NLP RNN embeddings for unstructured transactional data utilizing Torch and FastText, 13% F1 improvement for 9 classification tasks compared to off-the-shelf pre-trained embeddings.
- Generated weak-labels to supervise DNNs, bypassing expensive labeling and annotation.
- Designed pipeline architecture, with emphasis on reproducibility and performance, accomplishing >10x speedups.
- Interviewed and selected data science and machine learning engineer talent who contributed to Flowcast's success.

Data Scientist **January 2020 - March 2022**

- Launched a large-scale anomaly detection model for ING bank, currently scoring three million clients/month in production efficiently with Spark and XGBoost for the last four years uninterrupted.
- Achieved 47% error reduction in production of an interpretable regression model by Nike to control a multi-million dollar disputes funnel, through rework of preprocessing, feature engineering and bayesian hyper-parameter tuning.

UCSD - Biomedical Research Institute - Knight Lab *San Diego, CA*

Machine Learning Researcher, Internship **May 2017 - September 2017**

- Applied UMAP, PCA, UniFrac distance and phylogenetic trees to analyze three real world microbiological datasets.
- Conducted research to solve bias in dimensionality reduction, resulting in co-authoring a peer-reviewed [research article] (cited by 75) and presenting findings to an audience of 55 research scientists at a conference hosted by UCSD.

EDUCATION

École Normale Supérieure de Lyon *Lyon, France*

Masters degree in Mathematics, BS Math with Computer Science specialization **2015 - 2019**

- Succeeded a top 0.5% ranking in nationwide scientific exams to enter ENSL (french ivy-league) with a scholarship.
- Studied ML, Statistics, Linear Algebra, Computer Science, Stochastic Calculus, PDEs, Numerical Simulation, and more.
- Taught undergraduate math and physics to help 6 students succeed in competitive entrance exams.

SKILLS

Machine Learning: Deep Learning, NLP, Weak Supervision, Unsupervised learning, Generative Models, Interpretability and Explainability, Uncertainty, Multi Armed Bandits, Bayesian Optimization, Time Series, Classification, Clustering, Metrics, Transformers, LLMs, Large data sets, Imbalance, Statistical Modeling, Neural Networks, Gradient Boosting.
Software Development: Python (pandas, numpy, scikit-learn, matplotlib), PyTorch, PySpark, Ray, Git, GitHub, DVC, SQL, Distributed Computing, Linux, AWS, S3, Azure, GCP, TensorFlow, ETL, Airflow, Docker, Snowflake, Redshift, Databricks.
Current Projects: flippers, an open source library for weak supervision. See liamtoran.com.