David Liu

(+1) 647-674-6834 | ✉ davidsiqiliu@gmail.com | 🏠 davidsliu.com | 🔗 davidsiqiliu

Skills

Languages	Python, Java, C++, JavaScript, Scala, CSS, SQL, R
Technologies	Redis, Kafka, Spark, Docker, Databricks, Elasticsearch, React
Certifications	AWS Certified Associate Solutions Architect, Fellow of the Society of Actuary (Quantitative Finance)

Experience

Amazon   Toronto, ON
MACHINE LEARNING ENGINEER   *May 2021 - Present*

- Maintain and develop ML infrastructure for a core tier-1 service that serves **100K+ requests per second**
  - Engineer ML pipelines for key revenue-driving sponsored ads widgets (e.g., 'highly rated', 'shoppers also explored')
  - Led cross-team effort of creating an ETL pipeline that securely leverages sensitive customer data in model training
  - Optimized user clicks and elevate ads revenue (**+$13MM annualized** in the US marketplace) through A/B tests
- Designed and deployed an end-to-end ML pipeline in AWS Sagemaker for document recommendation
- Led the full-stack development of a high-performance knowledge-sharing web application in AWS
  - Architected automatic failover, load balancing and auto-scaling mechanisms that ensured optimal server load (60% - 70% CPU utilization) and high fault tolerance (99.5% availability) during traffic spikes and regional AWS outages
  - Successfully built and integrated new features into the backend (gRPC and GraphQL) and the frontend (React SPA)
- Enhanced system observability through innovative client-side user monitoring solutions
  - Surfaced metrics on system performance and user engagement to support making data-driven decisions
  - Presented the solution at a **company-wide innovation tech talk**
- Ensured security and privacy compliance by collaborating with legal teams
  - Addressed corporate and GDPR policies to establish a secure and privacy-centric ML environment

Huawei Technologies   Toronto, ON
COMPUTER VISION ENGINEER   *Sep 2020 - Apr 2021*

- Developed hand gesture-based command modules for seamless media control in Android phones and smart TVs
  - Refactored development Python codes into production Java codes to run optimally on low-CPU Android devices
  - Implemented a landmark detector leveraging Google MediaPipe, **boosting hand tracking stability by 40%**
  - Replaced resource-intensive bitmap operations with matrix-based computations, **reducing model runtime by up to 80%**

Munich Re   Toronto, ON
DATA SCIENTIST   *May 2020 - Aug 2020*

- Implemented a BERT-based text similarity matching model in PyTorch to streamline reinsurance underwriting
- Replaced a deterministic medical concept extraction tool with a novel Transformer-based, multi-hop question-answering model, resulting in a **threefold increase in the retrieval of relevant information**

Capital One   Toronto, ON
DATA SCIENTIST   *Sep 2018 - Aug 2019*

- Developed classification models leveraging credit card transactions for discerning profitable customer segments
- Enhanced model monitoring efficiency with Databricks and Spark, **reducing reporting time from 2 weeks to just 1 day**
- Transformed analytical table ETL processes using Jenkins and SQL, winning **company-level initiative awards**

Projects

Cancer Type Classification using Pathology Images 🔗   *Mar 2021 - Apr 2021*
PyTorch model with permutation-invariant layer (DeepSets) and graph convolution layer (ChebConv).

Distributed Data Cleaning Framework 🔗   *Mar 2020 - Apr 2020*
Spark framework in Scala for orchestrating parallel data deduplication tasks, ensuring optimal workload distribution.

Yelp Reviews Sentiment Analysis 🔗   *Mar 2020 - Apr 2020*
Ablation study on text pre-processing techniques and ML models to evaluate their effectiveness in predicting user sentiments.

Education

University of Waterloo
M.MATH. IN COMPUTER SCIENCE   *Sep 2019 - Apr 2021*
B.MATH. IN STATISTICS   *Sep 2011 - Apr 2016*