

NISAL PERERA

Mississauga, ON

www.linkedin.com/in/nisal-perera

nisalperera@icloud.com

Mobile: (647)-212-2177

PROFESSIONAL EXPERIENCE

Sr. Machine Learning Engineer

Aug 2025 - Dec 2024

Rakuten Kobo Inc., Toronto

- Designed a general MLOps framework for streamlining model deployment on GCP VertexAI.
- Built custom content moderation models utilizing OpenCLIP and OpenAI GPT vision models.
- Built a book metadata validation framework using Qdrant vector similarity search and FastAPI.
- Introduced AI coding agents while establishing best practices and increasing team productivity.

Lead Machine Learning Engineer

Mar 2023 - Mar 2025

Expedia Group, Seattle (Remote)

- Implemented scalable Python services on K8 that handle thousands of inferences per second.
- Built a LangChain RAG app using OpenAI LLMs to proactively flag policy violations in chat.
- Optimized RAG retrieval pipeline via evaluation-driven tuning, improving precision and recall.
- Implemented a Kafka based model request/response logging services for model monitoring.
- Integrated and optimized a DynamoDB Feature Store for sub 10ms feature look up at inference.
- Optimized live inference services, cutting cloud costs by 40% while preserving the same SLAs
- Utilized GitHub Actions and Spinnaker to automate model training and CI/CD pipelines.

Sr. Machine Learning Engineer

Jan 2021 – Feb 2023

Telus Communications Inc., Toronto

- Built a fraud model to prevent 20% of in store device fraud resulting in \$4M savings annually.
- Debugged and fixed production models using data drift analysis and model monitoring tools.
- Developed propensity models for various churn, up-sell, cross-sell Telus marketing campaigns.
- Utilized network data to optimize planning and enhance telecom services for rural customers.
- Increased team efficiency by migrating ML workflows to Vertex AI and BigQuery on GCP.
- Mentored junior engineers, established best practices, and fostered a collaborative environment
- Collaborated with business stakeholders to identify and prototype new AI initiatives.

Data/ML Engineer

Jan 2018 – Dec 2020

The Weather Network, Oakville

- Built an end-to-end predictive system to estimate ad click-through rate (CTR) using XGBoost.
- Implemented a behavioral segmentation algorithm in PySpark ML for personalized ad targeting.
- Developed ETL pipelines in Dataflow (Java) to extract insights from sparse geospatial data.
- Modelled data in Elasticsearch and Redshift databases to meet real-time querying requirements.
- Implemented Python Flask back-end microservices using AWS Lambda/API Gateway/Zappa.
- Maintained an Apache Airflow server for coordinating and scheduling ETL pipeline jobs.

Data Science Engineer

May 2017 - Dec 2017

Rubikloud Technologies Inc., Toronto

- Worked with data scientists to build a personalized retail product recommendations model.
- Developed scalable machine learning and data processing pipelines in PySpark on AWS EMR.
- Productized machine learning models into robust production-grade systems using Luigi.
- Deployed machine learning solutions on AWS cloud services using Docker containerization.
- Built an automated testing and validation tools for verifying the quality of model predictions.

EDUCATION

MSc in Applied Computing University of Toronto Thesis Awards: Mitacs Accelerate Fellowship Grant (\$15,000)	Sep 2016 – Dec 2017
BASc Computer Engineering, Honours University of Toronto Awards: Dean's Honours List, University Entrance Scholarship	Sep 2011 – May 2016