

Khoi Nguyen

✉ nguyentienkhoi01@gmail.com

☎ 438-938-0432

📍 Montreal, Canada

🌐 khoitiennguyen.github.io

🔗 [KhoiTienNguyen](#)

in [khoi-tien-nguyen](#)

Work Experience

Intact Insurance, Data Lab

Montreal, Canada

DATA SCIENTIST INTERN

May 2023 – Aug 2023

- Performed exploratory data analysis on different datasets to select relevant features and prepare clean data for flood risk model benchmarking.
- Investigated flood model's performance through geospatial analysis utilizing **PostGIS** and **SQL** on historic flood events.
- Evaluated claims notes classification model on recent data and assess its effectiveness on weather-related tasks.
- Wrote well-documented, reproducible **R** and **Python Jupyter** notebooks with visualizations to offer actionable insights.

McGill University, Distributed Digital Music Archives & Libraries Lab

Montreal, Canada

SOFTWARE DEVELOPER

May 2022 – Apr 2023

- Built workflow allowing users to label image layers and train optical recognition models to separate them.
- Developed essential image processing functionalities using **Python OpenCV** to prepare training data for the models.
- Worked closely with researchers to apply iterative changes to the modelling process and perform experiments.
- Designed and implemented a dynamic data loading scheduler to improve image loading efficiency and speed up model training by three times.

Accreon Inc.

Fredericton, Canada

CO-OP STUDENT

Sep 2018 – Jan 2019

- Implemented new functionalities in the ETL pipeline using **Java** for more flexible data processing and retrieval.
- Extracted client information from a central database and processed it as compressed JSON files to improve data storage and data sharing.

Education

McGill University

Montreal, Canada

B.Sc. COMPUTER SCIENCE, GPA: 3.53/4.0

Sep 2019 – May 2023

- Awards: Major Entrance Scholarship
- Courses: Data Science, Applied Machine Learning, Deep Learning, Artificial Intelligence, Intelligent Software Systems, Natural Language Processing with Deep Learning, Algorithms and Data Structures

Skills

Programming Python · SQL · R · Java · C · Bash · OCaml · JavaScript · HTML · CSS · Assembly

Libraries NumPy · Pandas · Scikit-learn · TensorFlow · PyTorch · HuggingFace · XGBoost · Django · Flask

Tools Git · Jupyter · PostgreSQL · Docker · AWS · Jira · Confluence · QGIS · PostGIS · GitHub Actions

Projects

Automatic Music Transcription

- Converted audio to Mel spectrograms and tokenized music symbols for model training using **NumPy**.
- Experimented with various custom **CRNN** model architectures in **Keras** using CTC loss for automatic transcription.
- Trained model across seven instruments, achieving less than 10% character error rate on four instruments.

Movie Recommendation System

- Developed a pipeline to collect data from a **Kafka** stream, pre-process the data using **Pandas**, and retrain models.
- Trained an **XGBoost** model with hand-crafted features and **SVD** collaborative filtering, achieving 0.15 RMSE.
- Trained a **K-Means clustering** model with user profiles using **scikit-learn** to mitigate cold start problem for new users.
- Implemented CI/CD with unit-testing, offline and online evaluations using **GitHub Actions**. Performed load-balancing with canary releases using **Docker** and **Nginx**. Monitored the system using **Prometheus** with **Grafana**.