Khoi Nguyen

nguyentienkhoi01@gmail.com · (438) 938-0432

linkedin.com/in/khoi-tien-nguyen · github.com/khoitiennguyen · khoitiennguyen.github.io

Work Experience

Intact Insurance, Data Lab   Montreal, Canada
DATA SCIENTIST INTERN   *May 2023 – Aug 2023*

- Performed comprehensive exploratory data analysis on client flood claims, geospatial maps, and satellite imagery using **Python** (Pandas, Matplotlib, NumPy) and **R**, uncovering critical insights for flood risk assessment.
- Increased flood claim identification rate by 85%, leveraging weather claims classification **LLM** and feature engineering.
- Discovered key patterns in claim distribution affecting 700,000+ customers through rigorous hypothesis testing.
- Reduced data inconsistencies by 60% through extensive data cleaning and preparation using **SQL,** ensuring data reliability and validity for model evaluation.

McGill University, Distributed Digital Music Archives & Libraries Lab   Montreal, Canada
SOFTWARE DEVELOPER   *May 2022 – Apr 2023*

- Developed a workflow for user-based image labeling and optical recognition model training using **Python** and **Docker**.
- Implemented image processing functionalities with **TensorFlow** and **OpenCV** to accelerate training data preparation efficiency, saving annotators 80+ hours per quarter.
- Designed a data loading scheduler for user-labeled datasets, increasing model training speed by 200%.
- Collaborated with researchers to execute code changes, integration testing, and model deployment on production.

Accreon   Fredericton, Canada
SOFTWARE DEVELOPER CO-OP   *Sep 2018 – Jan 2019*

- Refined **ETL** processes for improved data handling, introducing new features for enhanced flexibility and efficiency in data processing using **Java** and **SQL**.
- Streamlined client data extraction into bilingual formats and implemented compression techniques, improving data storage by 30% and increasing browser accessibility for developer use.

Projects

Movie Recommendation System

- Engineered a movie recommendation system incorporating a **Kafka** data pipeline, a **Redis** database, and advanced model training with XGBoost in **Python** for enhanced accuracy (0.15 RMSE).
- Developed a **Flask** API with WSGI server for real-time recommendations, supporting efficient backend operations.
- Integrated robust **CI/CD** practices using GitHub Actions, ensured load-balancing and system scalability through **Nginx** and **Docker**, and implemented comprehensive monitoring using Prometheus and Grafana.

Automatic Music Transcription

- Developed an automatic music transcription system, obtaining over 90% accuracy for sheet music conversion on four out of seven instruments, using a Convolutional Recurrent Neural Network (CRNN) in **TensorFlow**.
- Leveraged **Python** for audio processing and data handling, converting MIDI files to spectrogram images and tokenizing musical symbols for efficient model training and evaluation.

Skills

Programming	Python · SQL · R · Java · C · C++ · JavaScript · HTML · CSS · Bash · Unix · Linux
Libraries	NumPy · Pandas · TensorFlow · PyTorch · Scikit-learn · HuggingFace · XGBoost · Flask · Pytest
Tools	Git · Jupyter · AWS · Docker · MLOps · Jenkins · CI/CD · Airflow · Spark · Snowflake · Kafka · Redis

Education

McGill University   Montreal, Canada
B.Sc. COMPUTER SCIENCE, GPA: 3.53/4.0   *Sep 2019 – May 2023*

- Awarded the J.W. McConnell Major Entrance Scholarship for outstanding academic achievement.
- Courses: Data Science, Applied Machine Learning, Deep Learning, Natural Language Processing, Artificial Intelligence.
- Extracurriculars: McHacks Hackathon (2020), McWiCS Hackathon (2022), McGill Orientation Leader (2020 – 2022).