Minh Duc Hoang

dhoang301@gmail.com | https://github.com/minhduchoang301| linkedin.com/in/minhduc-hoang

EDUCATION

University of Toronto **Toronto, ON, Canada**

Honour B.Sc in Computer Science (4-year program + 1 year co-op) *Graduated May 2023*

- **Grade: Distinction. Annual GPA: 3.7/4.0**
- **Awards:** J.S Mclean Scholarship, ProjectX - International AI Research Competition Winner (awarded $20,000), 1500/1600 SAT (99th percentile)

PROFESSIONAL SKILLS

- **Programming Languages:** Python, Java, SQL, C, C++, R, PHP
- **Technologies Stack:** FLASK RestAPI, PostgreSQL, Pytorch, Tensorflow, Airflow, Dataswarm, Airflow, Spark, Presto, Git, Docker, Kubernetes

WORK EXPERIENCE

Meta Platforms (Facebook) **London, United Kingdom**

Data Engineer Intern *Jun 2022 – Sep 2022*

- Developed petabyte-scale ETL pipelines to enable cross-functional teams to extract insights from big data using Dataswarm.
- Optimized integrity enforcement on Meta Business platforms by conducting multiple A/B tests and consolidating relevant metrics, resulting in a 15% expected retention rate increase.
- Improved data analysis practices by implementing batch-processing features for complex data infrastructures and pipelines in Python and PHP.

Laboratory of Innovation Science at Harvard University **Remote, United States**

Quantitative Research Associate *Dec 2021 – Aug 2022*

- Conducted empirical research in quantitative and computational economics and management, resulting in novel discoveries in the field
- Analyzed the impact of consumer sentiment on innovation within organizations
- Researched and implemented multifactor models to extract alphas on financial assets from alternative data

Huawei Technologies Canada **Toronto, ON, Canada**

Machine Learning Engineer Intern *May 2021 – Apr 2022*

- Developed and implemented deep reinforcement learning models and supporting infrastructure to optimize big-data engines, resulting in a 33% reduction in latency.
- Built and deployed extraction, categorization, and prediction features to Hive clusters to collect historical data and recommend pre-computed analytical cubes, resulting in a 56% reduction in latency with minimal storage space cost.
- Automated testing data pipelines for feature integration using Docker and Kubernetes

RESEARCH PUBLICATIONS

Time Series Forecasting of Black Sigatoka in Banana Crops Using Neural ODE

International Conference on Machine Learning 2021

- Introduced a state-of-the-art model for infectious disease risk forecasting using deep learning based ordinary differential equations

PERSONAL PROJECT

Book Smart Search Engine API (Language: Python)

- Developed a search system that uses matrix factorization by combining rating histories of the user, the search query, and the existing reviews to generate the most appropriate search results from the database

Virtual Simple File System (Language: C)

- Implemented a Virtual Simple File System (VSFS) in C focusing on inode management, file/directory operations, and block allocation/deallocation while optimizing memory usage and system performance, enabling users to efficiently store, manage, access, modify, and delete files within the custom-built system