

Linh Hoang

linhhlp@gmail.com | Calgary, AB (relocate-able) | LinkedIn.com/in/linhhlp | LinhHLP.com

A leading Data Scientist who has translated novel discoveries into innovative solutions for 7+ years with experience analyzing and solving problems. Proficient in data structures and algorithms with Python, Java, C/C++, and MATLAB. Ability to communicate complex issues to different levels of the business. Full of curiosity and eagerness to explore and learn distributed large-scale systems in Data Science, Big Data, Machine Learning, and Predictive Analytics.

WORK EXPERIENCE

Data Scientist | York University, Toronto, ON

01/2016 – 01/2021

- Built automated Python tools to process and analyze the raw data from 1000+ datasets (academic & experimental).
- Invented 2 predictive models for high-expansion-ratio foaming with a super large scale of simulation to forecast the electrical properties at 20x to 100x. Improved other 10+ predictive models.
- Utilized C/C++ to develop simulation algorithms; Python & MATLAB to conduct statistical analysis and visualize data.
- Applied 3D mathematics to generate and transform; deployed on distributed computing clusters of 32+ nodes.
- Collaborated with other departments to fabricate and characterize samples; recommended infrastructure in the lab.

Project Owner | Tinh Van Group, Hanoi, Vietnam

03/2015 – 12/2015

- Project Poetry Machine: Started part-time in 2012 and scaled up to 10x QPS at 100x load more than the precedent.
- Utilized JavaScript to build responsive web apps; PHP for the core engine; and Python for background services.
- Designed the scalable microservice architecture including HTTP REST API XML (for web and mobile apps), deployed background NLP services, defined databases, and created ETL pipelines for management.
- Collaborated with testing, mobile, and marketing teams to publish Android and iOS apps to the market.
- Analyzed product and user behaviors and conducted A/B testing for experimenting with new NLP features.

R&D Researcher | Tinh Van Group, Hanoi, Vietnam

03/2015 – 12/2015

- Deployed a Big Data platform (Hadoop, Cassandra, Spark, Kafka) and ran experiments on the company's network.

Software Engineer and Data Analyst | Tinh Van Group, Hanoi, Vietnam

06/2011 – 09/2012

- Processed 10K+ daily SMS (MT & MO messaging) to execute business logic jobs (digital item purchases) by PHP.
- Conducted business analytics on products and user behavior data for the sale department and increased 25+% revenue by improving customer retention with accurate recommendations of digital products.
- Designed the scalable analytic tool with RabbitMQ, NoSQL Redis, and MySQL to show data in real time.

MACHINE LEARNING SKILLS

- Scalable Model Pipelines, Predictive Modeling, Recommendation System, Time Series Forecasting, Translator.
- Linear Regression, XGBoost, hybrid & layered/funnel-based models, seq2seq (TensorFlow), ChatGPT 3 (OpenAI API).

PROFESSIONAL SKILLS

- **7 YoEs:** Research & Development, Engineering Physics, Data Structures and Algorithms, Statistical Analysis.
- **5 YoEs:** Python, MATLAB, Debugging, System Design, MS Office, Data Preparation and Visualization.
- **3 YoEs:** MATLAB, PHP, SQL/MySQL, REST API, C/C++, Git, Linux, Agile/Scrum.
- **Solid:** NoSQL/Cassandra, TensorFlow, Anaconda, Google Cloud, Tableau, HTML, CSS, Java.

LEADERSHIP EXPERIENCE

- Led a master student in external company projects; assisted their experts with new materials (05/2017 – 01/2018).
- Trained students and mentored a student, won 1st award for Summer Research conference (05/2018 – 08/2018).
- Led incubator teams (20+ young talented students) as technology mentor and advisor, Tinh Van Group (2012, 2015).

EDUCATION

- Ph.D. in Physics cooperated to Mechanical Engineering | **York University**, Toronto, ON. 2021
- Master of Science in Computational Physics | **Sungkyunkwan University**, Suwon, Korea. 2015
- Bachelor's in physics | **Vietnam National University, HUS**, Hanoi, Vietnam. 2011

- Build Scalable Model Pipelines for Machine Learning with Data Cloud Services:
 - Built and deployed model in Virtual Machine (Amazon EC2) and Serverless Service (Lambda Functions).
 - Containerized model in managed clusters in Google Cloud Kubernetes Engine (Docker images).
 - Scheduled batch modeling with GCP Composer and Dataflow from/to BigQuery, Datastore, and Storage.
- Set up Big Data platforms with Kubernetes and Helm to cluster Spark, Cassandra, and PostgreSQL in Docker.
- Demonstrated seq2seq framework in TensorFlow using LSTM neural networks for both encoder and decoder to build Vietnamese Translator. Presented usages of APIs provided by OpenAI (ChatGPT 3) and Cohere AI.
- Built Recommenders and tuned predictors with the memory-based k-NN algorithm, Matrix Factorization-based algorithms, and layered/funnel-based models. The accuracy increased by 17% when using 4x detail of data.
- Utilized TensorFlow.Keras and fine-tuned models using optimization algorithms including Adam, SGD to predict electrical conductivity of nanomaterials with APLE Error below 10%.
- Predicted store retailers' sales (Time Series Forecasting) using Linear Regression and Hybrid Model with XGBoost with an accuracy of 87% for training and 85% for testing data sets. Analyzed seasonality, trend, and correlation.
- Analyzed tabular data and visualized it(Python, Pandas, Seaborn, Matplotlib, Scikit Learn).