

Robin Li

✉ x977li@uwaterloo.ca

☎ +1 2269789309

Toronto, ON

LinkedIn: [linkedin.com/in/xueqianli61](https://www.linkedin.com/in/xueqianli61); GitHub: <https://github.com/RobinL0601>

► EDUCATION

University of Waterloo

Sept. 2019 - June 2023

Bachelor Mathematics

Double Major in **Statistics** and **Computational Mathematics**, Minor in **Actuarial Science**

Received the University of Waterloo **President's Scholarship** with **excellent** academic standing (**GPA 3.6/4**)

► SKILLS

Programming Languages: Python, R, SQL, C, MATLAB, Scheme, SAS, C++, JAVA

Frameworks: QRM, Pandas, NumPy, Teradata SQL, Scikit-learn, Matplotlib, TensorFlow, XGBoost, MASS

Technologies: Git, IBKR, GCP, Teradata, Big Query, Vertex AI, Docker, Linux, Excel, AWS, Hadoop, Kubernetes, Airflow

► Professional Experience

Travelers

Aug. 2023 – Current

Associate Data Scientist

Toronto, ON

- * Retrieved data from AWS S3 and executed run jobs on **AWS EC2** using SDK for Python (Boto3)
- * Developed **regression models** in R for sustainability ranking correlation analysis with a 13% increase in accuracy.
- * Constructed SQL queries, utilizing subqueries, correlated queries, functions and procedures, optimizing extraction time by 30%
- * Designed an **inferential test** to assess independent correlation differences, transforming the ranking correlation test between policyholder demographics and renewal probability.

Research Assistant (Information Retrieval Cross Section LLM)

Jan. 2023 - Apr. 2023

University of Waterloo

Waterloo, ON

- * Developed a Streamlit web app to evaluate the **performance** of pre-trained models utilizing Kemeny rank aggregation.
- * Built ranking models for text retrieval tasks without dependency on large labeled dataset.
- * Implemented bagging and **Gradient Boosting methods** for 10-fold CV in model training
- * Applied advanced **NLP techniques**, including tokenization, Stemming, Lemmatization

Willis Towers Watson

Sept 2021 - Dec. 2021

P&C Data Analyst

Toronto, ON

- * Developed a **Python-based UI** dashboard to present the statistics with daily extracted data from **AWS S3**.
- * Conducted preliminary data analysis using **advanced SQL** (table join, subqueries, T-SQL, window functions, aggregation ...)
- * Implemented **ETL** for datasets from APIs, created end-user data sources and dataflows, to perform data cleansing,
- * Built and deployed machine learning models using Python libraries such as TensorFlow and PyTorch to predict claim likelihood and fraud detection, resulting in a 20% reduction in false positives.

Shenwan Hongyuan Securities

Jan. 2021 - Apr. 2022

Data Analyst

Shanghai, China

- * Conducted **hypothesis testing** and visualized S&T sustainability survey data with over 300 variables
- * Assisted the registered investment advisors with daily analytical work

► PROJECTS

LLM-based platform for job applications

Feb 2024 - Current

Extracted meta data from LinkedIn via APIs, loaded into S3 with **Apache Airflow**, embedded JD and resume into Vector DB using **Langchain**, utilized a Hugging Face 7B LLM for matching resumes and generating CV for each position.

Machine Learning Project: Auto Insurance Fraudulent Pattern Detection

June 2022 - Sept. 2022

Built a **ML model** that can detect auto insurance fraud by **Sklearn** and **XGboost** library (XGBClassifier, RandomForest, Decision Tree, GLM, Logistic Regression...), achieving **over 90% accuracy** in identifying fraudulent claims.

Monte Carlo Simulation in Delta Hedged Portfolio and Quantitative Risk Management

Jan 2023 - Mar 2023

Conducted portfolio risk analysis by minimizing VaR and CVaR in a Delta Hedged portfolio through Monte Carlo simulations across diverse t-distribution price paths

Kaggle Competition Gold

Sept. 2022 - Dec 2022

Performed classification on both texts and images of e-commerce dataset by training a **transformer** and **multi-layer** perceptron classifiers, combined with supervised models, such as **KNN**, Random Forest, and SVM using Python scikit-learn