

JILL WANG

Data Scientist & Software Developer

📞 (647) 936-5468 @ jill.ycwang98@gmail.com 🔗 LinkedIn: linkedin.com/in/jill-ycwang98 | Portfolio: jillwang87.github.io

EXPERIENCE

Data Analytics Developer (via Adastra)

TD Bank

📅 05/2021 - Ongoing

Developed projects in the Data Governance and Analysis department of Business Banking

- Single-handedly built a Python-based data lifecycle management (DLM) application that uses **Computational Linguistics** and **Natural Language Processing (NLP)** to extract features for a **machine learning pipeline** which **predicts security classifications and retention periods**
- Built a **user-friendly interface** with tkinter for business users to provide a location (either local files or a Hadoop database) to scan and automatically runs the ML pipeline, emitting the predictions in an Excel file
- The newly automated process delivers predictions in **minutes** that previously took months of manual work from data stewards
- **Technology: Impala, Python, pandas, NumPy, scikit-learn, NLP, tkinter, spaCy**
- ETL from Hadoop using Impala and developing Tableau dashboards
- Some examples of the Tableau dashboards include a turnover report and a business process study report
- **Technology: Impala, Python, pandas, NumPy, Tableau, Oozie**

Data Science Consultant

Adastra North America

📅 04/2021 - Ongoing

Data and AI consultancy

Data Scientist & Software Developer

Graphen (based in New York, USA)

📅 06/2019 - 08/2019

Organization building machine learning platforms leveraging their graph database to mimic human brain functionality

- Developed an unsupervised model on time-series data and a supervised model on multivariate data, both for the purpose of performing **financial behaviour anomaly detection/ fraud detection**
- Applied **statistical methods** to design a systematic way to calculate a risk (anomaly) score
- Applied a **numerical analysis method** to improve the accuracy of a time series model
- Advised company leadership on strategies to combine the company's core product, a fast and efficient graph database, with machine learning models that helped **land a proof-of-concept project** with one of the largest banks in Taiwan
- Developed the front end of a React-based **anti-money-laundering platform**
- **Technologies: Python, NumPy, pandas, scikit-learn, XGBoost, Git, JavaScript, React, GraphQL, Elasticsearch, SASS, CSS3**

Software Developer

Entrepreneurship Society at University of Waterloo

📅 01/2019 - 04/2020

University of Waterloo club connecting students to the Kitchener-Waterloo startup culture

- Transitioned the React-based static desktop website to a fully responsive, mobile friendly experience
- **Technologies: JavaScript, React, SASS, CSS3, Git**

PROGRAMMING LANGUAGES & TOOLS

Python	SQL
Tableau	Impala, Hive, Oozie (Hadoop)
R (Statistical Computing)	JavaScript, React
MATLAB	C, C++
Bash	Git

EDUCATION

Honours Bachelor of Mathematics in

Computational Mathematics

University of Waterloo, 2021

Honours Bachelor of Mathematics in

Statistics, Computing Technology Option

University of Waterloo, 2021

AWARDS & PROJECTS

Disaster Tweets Prediction (In Progress)

- Predict which Tweets are about real disasters and which ones are not
- Using spaCy to clean up and analyze texts for the extraction of features for a classification model
- **Technologies: Python, spaCy, NLP, sklearn**

🔗 <https://www.kaggle.com/code/jillwang87/analyze-and-extract-features-with-spacy/notebook>

Humpback Whale Identification

- Trained using a very challenging, highly imbalanced dataset with over 25,000 images and about 5000 classes
- Use a VGG19-inspired CNN and achieved a ranking of 1722/2121 on Kaggle
- **Technologies: TensorFlow, scikit-learn, Deep Learning**

🔗 <https://www.kaggle.com/jillwang87/stat-441-fin-al-project-humpback-whale-vgg/notebook>

Jewellery Shop Prediction Top 4%

- Predicting whether an online jewellery shop visitor will purchase an item from the shop using XGBoost with 93.33% accuracy
- **Technologies: Python, scikit-learn, XGBoost**

🔗 <https://www.kaggle.com/code/jillwang87/xgboost/notebook>

NOVA - MLH StarterHacks 2018 Best Pitch (Finalist) Award

An Android app with an AI chat-bot designed to connect university students to on-campus mental health resources

- **Technologies: Python, Java, Android, Facebook Messenger API, NLP**

🔗 <https://devpost.com/software/nova-29i4zx>

VOLUNTEERING

Events Coordinator

StarterHacks

📅 08/2018 - 02/2019

Beginner-focused hackathon for over 1200 hackers