

# HAOYU GU

✉ h46gu@uwaterloo.ca  
🌐 ForKaori.github.io  
☎ 5197296185  
📍 Waterloo, ON, Canada  
in linkedin.com/in/stephen-gu-92b669163/  
🔗 ForKaori

Third-year Data Science student who loves coding and analyzing hidden information behind the scenes.

## Skills

### PROGRAMMING LANGUAGES

Python  
C++  
C  
Javascript  
HTML  
CSS  
Scheme

### FRAMEWORKS

Pandas  
Numpy  
Plotly  
Dash  
Tensorflow  
Keras  
Scikit-learn  
Node.JS  
Flask

### TOOLS

Git  
Bitbucket  
Apache Kafka  
Jekyll  
Bootstrap  
GCP

## Education

### University of Waterloo Data Science

Sept. 2017 to Apr. 2022

Related Courses: Data structure and data management, Object-oriented software development, Introduction to database management

## Employment

### Enlightened Inc Data Scientist / ML Engineering

Waterloo, ON

Jan. 2020 to Apr. 2020

- Implemented dashboard for monitoring real-time streaming data using Dash framework from scratch. Reduced on-site testing human cost by helping engineering team to quickly find inaccurate sensors.
- Developed data pre-processing automation script using pandas, numpy, requests library and built an algorithm based on Kalman Filter to remove the noise on raw data sets.
- Analyzed location data and implemented different accuracy metrics based on confusion matrix and cumulative distribution function; operated as part of quality assurance process.

### Scotiabank Pricing Scientist

Toronto, ON

May 2019 to Aug. 2019

- Developed data pipeline on Google Cloud Platform using AutoML APIs and Composer to schedule monthly-run machine learning models. Lowered the human resource cost.
- Developed machine learning models using xgboost and keras library to predict the probability customers will renew a financial product based on history data.
- Analyzed data for resetting the threshold of high value customers using libraries pandas, numpy, seaborn. Presented the insightful report to business team for decision making.

### PureLiving Monitoring Associate

Shanghai, China

Sept. 2018 to Dec. 2018

- Implemented Air quality index converter program helps compare air quality standard from various countries; developed internal testing process for monitors and reduced clients' negative feedback.
- Responsible for analyzing clients' indoor air quality data, delivering suggestions to engineering team, taking onsite visit to help clients solve technique problems.

## Projects

### Data Automation script for GCP AutoML

- Performed as a pipeline for data processing on Google Cloud Platform for monthly-run machine learning tasks. Integrated with AutoML APIs and Composer (Apache Airflow DAGs).

### Real-time streaming data dashboard

- Developed based on Dash library and Apache Kafka framework to handle real-time streaming data; giving concise analysis to video website.

### Compiler for C-like languages

- Implemented based on C++ from scratch which translates high-level simplified C language to assembly language and then machine code.