

Fan Gao  
fangao\_gf@outlook.com | (416)-668-6461  
Personal Website: <https://flamingohue.github.io/>

---

## Qualifications & Skills

- Highly proficient in **python** (**pyspark**, **scipy**, **scikit-learn**, **tensorflow**, **dash**, **plotly**, etc), **R**, **SAS**, **SQL**, **unix**, **git**, **Latex**, **hadoop**
- Solid background in data mining and machine learning algorithms, such as linear regression, random forest, neural network, decision tree, KNN, Kmeans, NLP and XGBoost
- Strong report writing and presentation skills

## Certificates

- SAS certified base/Advanced programmer
  - AWS Certified Cloud Practitioner
  - Certified Financial Risk Manager (GARP FRM)
- 

## Professional Experience

### Manager, AML/ATF Model Analytic, Scotiabank, ON, Canada

Feb 2021 - Present

- Developed Client Risk Rating models on Retail (2M entries) using logistic regression and Machine learning algorithms (Random Forest, SVM, etc), which provides the key insights for stakeholders on catching money laundering using **python**, **pyspark**, **scikit-learn** and **Oracle DB**.
- Integrated a series of client risk rating models for multiple countries, created a standard unified interface which accepts various data sources and countries requirements for previously separated models, improving the monthly ad hoc efficiency and facilitating the monitoring and analysis of existing models
- Contributed to a Web Platform development, which focuses on the dynamic data visualization toolkit for data impact output for existing models for over 850 end users using **python**, **dash** and **plotly**.
- Automated the repetitive routine maintenance tasks into timely based manner and in the background running automatically on an on-going basis using **shell scripts** and **crontab**
- In charge of the Global Banking Management Client Risk Rating model and communicating with the Model validation Team for monitoring report validation.

### Modelling/ Forecasting Modelling Analyst, TD bank, ON, Canada

Nov 2019 - Feb 2021

- Collaborating with Model Implementation team and assist in the development of a centralized and automated data creation and wrangling API with **SAS** and **shell script**
- Design an automated data integrity testing for data governance with **SAS** and **bitbucket** for code version control and develop data visualization with **VBA**
- Use **Hadoop** to pull and pre-process the datasets and assist the senior quantitative analyst to develop the PD model to predict the default probability with **SAS** and **python**
- Validated classic credit card PD model's predictability check on reperforming accounts with **SAS**

### Research Analyst, UHN Hospital, Toronto, ON, Canada

Jan 2018 - Jun 2019

- Design a new classification method targeting microbiome data and validate the new classification method with out-of-sample analytics with **R**
- Compare the developed method with existing classification methods on simulated datasets and real datasets, such as **linear regression**, **Linear discriminant analysis**, **KNN**, **K-means**, **decision trees**
- Present statistical results to committee groups, write manuscript for thesis and publication with **Latex**

## Research Analyst, CAMH, Toronto, ON

Sept 2017 - April 2018

- Conducted data cleaning and quality control check on large dataset (2.1 million events)
  - Applied a **time series model** and **autoregressive integrated moving average (ARIMA) model** to assess the birthday effect on the traffic-related mortality using **R**
  - Validate the accuracy and correlation analysis on the final model with **R**
- 

## Machine Learning Project Experience

NLP with Disaster Tweets at Kaggle.com

Oct 2020 - Dec 2020

- Conducted data cleaning for stopwords, lemmatization, punctuation, and missing values
  - Performed word2vec algorithm to obtain a score as a variable
  - Conduct KNN, Logistic regression with Lasso regulation , PCA algorithm, XGBoost, random forest to build a predictive model and perform cross validation to validate the model with F1 score with **python**
- 

## Education

MSc in Bio Statistics, Thesis stream, University of Toronto, Canada

Fall 2017 - Summer 2019

BSc in Biological Science, University of California Irvine, USA

Fall 2012 - Summer 2016

---

## Leadership Experience

CSSA (Chinese Student and Scholar Association)

Fall 2014 – Fall 2015

- Assisted the VP of the career department to organize big events, such as career fairs.

BUGS (Biostatistics Union of Graduate Students)

Fall 2017 – Summer 2019

- Worked as a webmaster to keep the official website updated