Yi Tang

7511 Afton Drive, Richmond, BC V7A 1A2
+1-2508842029 | tangyi2014@gmail.com

Education

University of British Columbia *Kelowna, BC, Canada*

Master of Data Science *Sep. 2019 – Jun. 2020*

- **Selected Courses:** Bayesian Statistics, Machine Learning, Modelling & Simulation, Privacy & Ethics, Resampling & Regularization

University of Victoria *Victoria, BC, Canada*

Bachelor of Science in Statistics *Jan. 2015 – Jan. 2019*

- **Selected Courses:** Design & Analyze of Experiment, Stochastic Processes, Sampling, Regression Analysis

Relevant Skills

Programming R, Python, MySQL, Spark SQL, CQL, Julia
Software Microsoft Office, Tableau, Markdown, SPSS, EViews, AWS
Language Mandarin Chinese (Native), English (Proficient)

Relevant Work Experience

Lausanne Business Solutions *Philadelphia, PA, United States*

Junior Data Analyst - Part time *Feb. 2021 –Present*

- Sourcing and collecting data, recording and transcripting audio, and validating transcriptions carefully
- Simulating written and audio conversations to train chatbot backstage efficiently
- Offering Mandarin Chinese consulting including authoring and reviewing grammars accurately

Zhejiang Unitech Information Technology co.,ltd *Ningbo, Zhejiang, China*

Statistician *Jun. 2016 –Jul. 2016*

- Analyzing and testing for public opinion analysis system backstage efficiently
- Observing web crawler operational status and recording the number of news grabbed from various websites in spreadsheets by Excel
- Updating daily top information by summarizing news to a public media platform in time

Volunteer Experience

LKKER Conference 2016 *Ningbo, Zhejiang, China*

VIP Guest Assistant *May 2016*

- Connected and balancing one of VIP guest's requirements efficiently
- Adjusted one of VIP guest's schedule, agenda for taking part in the conference till feasible
- Given response in accidental situations rapidly and flexibly
- Learned communication and time management skills

Chengguan No.3 Primary School *Taijiang, Guizhou, China*

Volunteer Teacher *June 2013*

- Generated various lesson plans to account for students' individual variations
- Introduced mathematic knowledge within a relaxing teaching environment to motivate learners
- Interviewed several students' families and filling in information forms to figure out problems and suggestions for students in learning process

Selected Projects

HIV-1 NRTIs drug class resistance mutations prediction *Oct. 2018 – Nov. 2018*

- **Goal**: Performing model selection among various models to predict HIV-1 NRTIs drug class resistance mutations
- **Method**: Fitting several types of models including linear regression, penalized linear regression, logistic regression, regression tree, penalized logistic regression, Linear Discriminant Analysis(LDA), K-Nearest Neighbours (KNN) and classification tree
- **Tool**: Manipulating R Studio to complete the project

Player Unknown's Battlegrounds data prediction *Nov. 2018 – Dec. 2018*

- **Goal**: Determining the best model for predicting final placement from in-game stats and initial player rating
- **Method**: Visualizing the distribution of percentile winning placement and correlation between 29 covariates, fitting models for each type respectively where the type split by the match type including linear model, penalized linear regression model and regression tree model
- **Tool**: Manipulating R Studio to complete the project

App reviews procession and classification *Dec. 2019 – Feb. 2020*

- **Goal**: handling huge reviews which are mobile applications text with a total of 2715303 by natural language processing (NLP), and reproducing the program from report On the Automatic Classification of App Reviews with randomly selected App reviews to classify them into four categories.
- **Method**: Processing the reviews including removing non-English reviews, non-ASCII characters, punctuations, adjacent duplicate characters and reviews that contain two or fewer words by langid, string and regular expression for the original dataset, counting future tense words, past tense words, present simple tense words and present continuous tense words for each review by nltk and calculating sentiment rating for App reviews by sentistrength, selecting 384 rows App review text from the processed dataset with manually labelled four categories to test the accuracy
- **Tool**: Manipulating Python to complete the project

Central Okanagan real estate sales evaluation *Mar. 2020 – Mar. 2020*

- **Goal**: Evaluating housing investment in the Central Okanagan Region, BC and helping company managers make decisions
- **Method**: Estimating the number of aggregate housing sale volume, the average housing price and the total units in the Central Okanagan Region in 2020 using a weighted formula with housing selling data about this Region from January 2017 to February 2020, visualizing and analyzing the housing selling data about the Central Okanagan Region in the past and future
- **Tool**: Manipulating Excel and Tableau to complete the project

COVID-19 virus infection forecast *Apr. 2020 – Apr. 2020*

- **Goal**: Forecasting COVID-19 Virus Infection until May 31, 2020, in Canada with data contained the number of confirmed cases and fatalities of 120 provinces and 2 territories from January 22, 2020, to April 11, 2020
- **Method**: Presenting model selection with the lowest mean-square error (MSE) from Quasi-Poisson, linear and ARIMA models with different forms of parameters and degrees, predicting confirmed cases and fatalities for the next 50 days, visualizing and analyzing the trends of confirmed cases and fatalities in Canada for each province or territory respectively
- **Tool**: Manipulating R and Tableau to complete the project

Demographics for Interior BC cancer patient population analysis *Apr. 2020 – Jun. 2020*

- **Goal**: Exploring the geographical disparity for cancer outcomes across Health Service Delivery Areas(HSDAs) and Local Health Areas(LHAs) in the Interior Health region from dataset of Cancer Agency Information System (CAIS) at BC Cancer from 2012 to 2016
- **Method**: Examining differences in incidence and mortality regards to the distribution of cancer types by visualizations, detecting differences in cancer incidence and mortality by cancer type amongst HSDAs and LHAs and investigating how incidence and mortality change within ten cancer types across each HSDA and LHA considering age groups or not by Chi-square test and ANOVA tests
- **Tool**: Manipulating R and Tableau to complete the project

Reference

Available upon request