David Yang

davidheweiyang@gmail.com

https://www.linkedin.com/in/david-h-yang/ 1492 Panatella Blvd Calgary AB | (587)-889-2618

SKILLS

Languages/Software: Python, Java, R, UNIX Shell, LaTeX, SQL, OpenBUGS, MS Excel, Git, Minitab, Tableau

Target skills: Data Analysis (Pandas, NumPy, dplyr), Data Visualization (Matplotlib, ggplot2), Machine Learning (scikit-learn, TensorFlow), Webscraping (Selenium, BeautifulSoup), Automation Testing (Selenium, Cucumber)

WORK & RESEARCH EXPERIENCE

Web Automation Developer, *ADM Lucid Solutions Inc.*, (Part-time) April 2023 – Present

- Developed automation test scripts using Selenium with Java which validated the integrity of web applications
- Collected performance metrics with Google Lighthouse to provide constructive suggestions for web applications

Exploratory Analysis with RNA sequencing (RNAseq) and metabolomics data, *University of Calgary* May 2023

- Demonstrated advanced analytical skills by pinpointing ~50 out of >30,000 significant genomic factors in big complex biological datasets through statistical techniques such as regression modelling, PCA, and Bootstrapping
- Generated scientific figures using data visualization libraries which elucidated key research findings to external institutions leading to the receival of monetary grants valuing greater than $25,000
- Exemplified a strong ability to interpret complex datasets by employing dimensionality reduction techniques (i.e. regularization, CCA), and other data wrangling techniques (i.e. normalization, data imputation) on noisy biological datasets with high dimensionality and multi-collinearity
- Collaborated with researchers from interdisciplinary backgrounds to present findings in the form of a scientific paper

Chief Information Officer, *Canadian Organization for Undergraduate Health Research* Jun 2018 – Aug 2021

- Designed the framework for an Android mobile health tracking application with Android SDK in Android Studio
- Leveraged data analytics to exceed marketing KPI by 100% for both internal recruitment and general advertisement

Optimized Sampling for COVID-19 MCMC Evolutionary Analyses, *University of Calgary* May 2021

- Devised a weighted sampling strategy based on scientific deductions of the COVID-19 pandemic which drastically reduced sampling bias when selecting 1000 samples from a population of more than 2 million samples
- Applied MCMC algorithm on COVID-19 data to describe disease transmission consistent with current scientific models

PROJECTS

Feedforward Classification Neural Network (FNN) Sep 2022

- Developed & deployed a FNN via TensorFlow to accurately discriminate between images of healthy and diseased crop
- Utilized a model pretrained on ImageNet dataset via transfer learning to perform repurposed image classification tasks

Cryptocurrency Trading Bot (CTB) Jun 2022

- Implemented a CTB that connects to the Binance API to perform trades on high volatility currencies with Python

Bayesian Inference of Zero-Inflated Datasets Jan 2022

- Programmed custom Bayesian statistical models in R using OpenBUGS to model zero-inflated datasets, resulting in a more accurate representation of the data which was validated with Bayesian credible intervals and statistical tests
- Implemented Gibbs sampling in R to approximate parameters for Bayesian statistical models with complex distributions

EDUCATION

M.Sc. Mathematics and Statistics Mar 2024 (Expected)

University of Calgary | GPA 3.7/4.0 | Thesis project: Parallelization of MCMC Phylogenetic Analyses

Coursework: Deep Learning, Generalized Linear Models, Statistical Inference, Bayesian Statistics, Theory of Probability

B.Sc. Cellular, Molecular, and Microbial Biology Sep 2017 - May 2021

University of Calgary | GPA 3.96/4.00 | Honours project: Eliminating Sampling Bias in SARS-CoV-2 Analysis