(604) 710-7454
Vancouver, BC
kevinyang10@gmail.com

# Kevin Yang

Junior Data Scientist

Portfolio: keviny2.github.io/
github.com/keviny2
linkedin.com/in/keyang2

## Skills

| | |
|---|---|
| **Tools and Languages** | Python, R, SQL, Linux, Git, Jupyter Notebook, Visual Studio, PyCharm, RStudio, Vim, SLURM, Azure |
| **Packages** | PyTorch, NumPy, pandas, scikit-learn, matplotlib, Pyro, NumPyro, PyMC, PyQt |
| **Machine Learning** | Neural Networks/Deep Learning, Regression, Clustering & Classification, Decision Trees, Time Series, SVM, t-SNE, K-means, DBSCAN, Hypothesis Testing, ANOVA |

## Technical Experience

**MSc. Student / LiquidBayes** **01/2021 — Present**
*BC Cancer Research Centre* *Vancouver, BC*

- Slash error rate 75% over state-of-the-art methods by designing machine learning models to infer cancer population structure
- Exploit HPC clusters to gain 10x speed up on pipeline executions
- Generate publication quality plots with matplotlib to visualize L1 losses for over 5 experiments
- Coordinate with a team of over 10 researches from two countries for deploying bioinformatic analyses to the cloud

**MSc. Student / Cancer Tumour Image Reconstruction** **01/2021 — Present**
*BC Cancer Research Centre* *Vancouver, BC*

- Decrease validation error by 60% for spatial tissue analysis by implementing two cutting-edge deep learning models
- Accelerate neural network training over 100x using GPUs with CUDA in PyTorch

**Undergraduate Teaching Assistant** **09/2019 — 08/2020**
*The University of British Columbia* *Vancouver, BC*

- Communicated key concepts related to data structures and basic algorithms, receiving 100% positive feedback from students
- Coached students through coding assignments and exam preparation for up to 50 students during weekly office hours

**Application Developer / Property Value Website** **01/2019 — 08/2019**
*Statistics Canada* *Ottawa, ON*

- Stored, managed and visualized Property Value data by writing 10+ SQL stored procedures for a relational database
- Achieved over 80% function coverage in unit testing on a program to process web scraping data
- Collaborated with 4-5 senior programmers to develop creative solutions and optimized an $O(n^2)$ algorithm to $O(n)$

## Projects

**MethylationDPGMM**
- Extended a machine learning clustering algorithm to perform multivariate analysis
- Cut memory usage 75% by creating an R package to process and wrangle biological big data

**Vancouver Temperature Forecasting**
- Forecasted Vancouver's temperature in 2020 given temperature data for Vancouver from 2003-2019 using time series analysis

**Rental Vehicle Statistical Dashboard**
- Assembled a Java GUI App that performs SQL queries to retrieve vehicle statistics for a car rental company

**Data Science Regression Application**
- Created a GUI Application to perform 6 different types of statistical regressions on user inputted data sets

## Education

**MSc. Bioinformatics** - ***GPA: 4.0***, *The University of British Columbia* 2021-2022
**BSc. Computer Science & Statistics** - ***GPA: 3.9***, *The University of British Columbia* 2015-2020
*Canadian Institutes of Health Research (CIHR) CGS+M ($17 500 over 1 year)* 2022

## Presentations

- CAIDA 2020 - *Bayesian Non-Parametric Model for Pan-Cancer Analysis (Poster Presentation)*

## Conferences

- RECOMB-Seq 2021
- Cancer Grand Challenge 2021
- CAIDA 2020