Kevin Yang
https://keviny2.github.io/
Email : kevinyang10@gmail.com
Mobile : +1-604-710-7454

Summary

Aspiring data scientist equipped with extensive knowledge of machine learning methods and exceptional programming skills. Experienced in building algorithms and developing models to answer meaningful questions in cancer research. Capable of constructing pipelines to process raw data for data analysis. Effective in collaborating with other developers on large-scale projects.

Key Skills

• Machine Learning • Computer Vision • Bayesian Modeling • Model Training • Benchmarking • Data Analytics • Data Visualization • Data Wrangling • Data Preprocessing • Pipeline Development • Cloud Computing • Communication

Technical Skills

- **Tools:** Python, R, Linux, SQL, SLURM, Azure, PyQt
- **Packages:** NumPy, Pandas, Matplotlib, PyTorch, Pyro, Snakemake, GATK, Statsmodels, Scikit-Learn, Jupyter Notebook, Vim
- **Statistics/Machine Learning:** Bayesian Modeling, Sampling Methods, Deep Learning, Linear Models, Decision Trees, SVM

Experience

BC Cancer
Jan. 2021 – Present
Master's Student, Supervisor: Dr. Andrew Roth — *Vancouver, BC*

Model Development & Implementation
- Building **Bayesian ML models** for inference of cancer population structure and spatial tissue analysis
- Developing **deep learning** solutions in **PyTorch** for image reconstruction

Analysis Pipeline Design
- Constructing **pipelines** for bioinformatics experiments using snakemake
- Generating realistic synthetic biological datasets using samtools and GATK for large scale **benchmarking** experiments
- Preprocessing complex sequencing data from real cancer patients with **pandas**, **dplyr** and **tidyr**

Cloud Computing & Optimization
- Utilizing **HPC clusters** to gain over an order of magnitude speed up on snakemake experiments
- Operating **Microsoft Azure** to store and load sensitive patient sequencing data

Data Visualization & Communication
- Generating plots in Jupyter Lab with **matplotlib** and **ggplot** for data visualization and EDA
- Composing a manuscript for submission to popular scientific journals

Internship

Statistics Canada
Jan. 2019 – Aug. 2019
Application Developer — *Ottawa, ON*

Full Stack Development
- Developed web apps for storing, managing and visualizing Property Value data in **JavaScript** using **Ajax** and **jQuery**
- Generated **SQL** stored procedures for **ETL** processes on large databases

Education

University of British Columbia
Jan. 2021 – Present
MSc. Bioinformatics - GPA: 4.0/4.0 — *Vancouver, BC*

University of British Columbia
Sep. 2015 – Dec. 2020
BSc. Computer Science and Statistics - GPA: 3.9/4.0 — *Vancouver, BC*

Projects

LiquidBayes
- Implementing a machine learning model to infer cancer population structure using Bayesian inference techniques
- **Key Achievement:** Gained a 75% reduction in error rate over current state-of-the-art methods

Regression App
- Created a GUI Application to perform various types of regressions on user inputted datasets

stpt2imc
- Built a deep learning model using PyTorch to reconstruct IMC images from STPT images