YEE FAY LIM (508)6560622 | yeefay.lim@gmail.com | https://github.com/yflim | Vancouver, BC

SUMMARY ML engineer and M.Sc. with 5+ years' software development experience and 3+ years' varied data project ownership. Deep understanding of statistics and ML; comfortable with data projects end to end, from reading research papers, to data wrangling and analysis, to model building, to writing production-level code and deployment.

EDUCATION

Applied Quantitative Methods (AQM), Vancouver, BC
Data science workshop for graduate students; selected for Lululemon capstone project. Jan. – Nov. 2020.

The University of British Columbia, Vancouver, BC
M.Sc. Bioinformatics, March 2021; GPA: 3.33/4.0
Coursework included: Stat. Methods for High Dimensional Bio., Stochastic Processes, Topics in Probability

The University of Chicago, Chicago, IL
M.S. Computer Science, June 2011; GPA: 3.69/4.0
B.A. Economics, B.A. Mathematics, June 2008; GPA: 3.19/4.0

WORK EXPERIENCE

Versatile Media, Vancouver, BC
Machine Learning Research Engineer, Jan 2021 – present
- Adapt, implement, deploy CNN pipeline based on published research for inferring 3D output from video input
- Created faster solution than published, with comparable accuracy on larger problem size and fewer samples

Industrial Data Associates, Chicago, IL (fully remote)
Software Developer, August – September 2012, September 2014 – May 2017
- Full-stack maintenance, development, testing of taxonomy and reference content management web platform.

HP Autonomy, Chicago, IL
Software Engineer, April 2012 – July 2014

RESEARCH AND PROJECT EXPERIENCE

Bioinformatics Technology Lab, Michael Smith Genome Sciences Centre, Vancouver, BC
Graduate research assistant, August 2017 – December 2020
- Created unsupervised classifier for genome sizes from bacterial to human, and wide range of heterozygosity levels
- Performance: binary F1 > 0.99, multiclass accuracy ~ 0.9. 8 to 435% improvement over existing alternatives.

AQM projects: TransLink, April 2020; **Lululemon capstone**, August – November 2020
- Investigated complex transit and survey datasets with ~50 to ~200 features; reported to Lululemon stakeholders
- EDA, feature selection & engineering, PCA, clustering (hierarchical, *k*-means, OPTICS, affinity propagation)

Shmulevich Lab, Institute for Systems Biology, Seattle, WA
Intern, June – August 2018
- Preliminary hypotheses validation using cancer miRNA and mRNA expression dataset with 20,000+ variables

SKILLS

Statistics & ML: model selection; regression (linear, GLMs, nonlinear); random forests; dimensionality reduction; regularization; EM; neural networks and deep learning; hyperparameter tuning; etc. Quick learner; stays current on ML research and industry practices, including in MLOps, through regular ML reading group participation and other means.
Data analysis & visualisation: lmfit, NumPy, SciPy, Pandas, Scikit-learn, Keras, TensorFlow, Matplotlib, Seaborn, Jupyter. Fluent in Python and associated tools as listed.
Other: SQL (3 years' prior experience); Unix CLI; Git; AWS; Docker

AWARDS

PyData New York City diversity scholarship, November 2019
NSERC-CREATE bioinformatics graduate studies scholarship, 2017 – 2018
Public Services Agency of Malaysia undergraduate scholarship, 2004 – 2008

PORTFOLIO

Data science samples | Parallel tridiagonal linear solver | Writing and presentation | Other