ZHOURONG LI

Phone: (+86)18926269705 E-mail: zl2977@cumc.columbia.edu

EDUCATION INFORMATION

Columbia University, Mailman School of Public Health **New York, USA**
Master of Science in Biostatistics **09/2020-05/2022**

- *Main Coursework: Probability, Principles of Epidemiology, Data Science I, Statistical Inference, Biostatistical Methods I, Biostatistical Methods II*

McGill University, School of Science **Quebec, Canada**
Double Major, Bachelor of Science in Statistics and Computer Science **09/2015-08/2019**

- Main Coursework: *Regression & Analysis of Variance, Probability, Statistics, Introduction to Stochastic Processes, Theory of Computation, Advanced Calculus, Matrix Numerical Analysis, Applied Machine Learning, Artificial Intelligence, Algorithm Design*

PERSONAL SKILL

- Programing Language: Java, python, C, F#, shell, SQL, R, MIPS, Visual Basic, MATLAB
- Programming Utility: R Studio, Visual Studio, Eclipse, MATLAB, Logism-Evolution, PyCharm
- Design: ALU Circuit Design (Full adder, Multiplexor), CPU design
- Microsoft Utility: Microsoft Word, Excel, Power Point Presentation, OneNote
- Adobe Utility: Adobe Photoshop CS6, Adobe Premiere Pro, Adobe After Effects
- Apple Utility: Numbers, Pages, Keynotes, Logic Pro X, Final Cut Pro
- Systems: Microsoft Windows, Unix, Linux, Macintosh OS X
- Languages: Mandarin (Native), Cantonese (Native), English (Fluent), French (Conversational)
- Others: Piano (Level 8 Certify)

PROFESSIONAL EXPERIENCE

Investment Banking, Bank of China **Guangzhou, China**
Investment Assistant **11/2019-12/2019**

- Had a thorough understanding of the function of the bond underwriting group as well as the underwriting process.
- Completed qualitative analysis for a local Pharmaceutical Company and conducted due diligence investigation by reviewing the audited financial report in recent twelve months
- Implemented the Moving Average Model, Exponential Trend Model and Multiple Linear Regression Model in R-Markdown to predict monthly sales revenue for the local Pharmaceutical Company

Portfolio Management and Quantitative Strategies Investment, E Fund Management Co., Ltd. **Guangzhou, China**
Quantitative Analyst **10/2019-11/2019**

- Developed program with Python that monitoring real time portfolio rate of return compared to benchmark via Tushare library and implemented data visualization with Plotly library, verified results with Wind database
- Built Python program that classify the portfolio rate of return and benchmark rate of return by industries, and construct single-period and multi-period Brinson models for performance attribution. Results were ranked by Allocation Return

North American Headquarters, Synced **Toronto, Canada**
Data Analyst **05/2018-07/2018**

- Implemented WebCrawler with python, collected Author/Topic/Organization information from machine learning meetings (i.e. CVPR, NIPS, ICML etc.)
- Used collected topics to derive the most popular topic through different evaluation methods. (topic frequency, number of references, Google H5-index)
- Constructed a Wikipedia Entry of machine learning knowledge for users to study and search for relevant resources

EXTRACURRICULUM ACTIVITY

***Online Course Lecturer*, Central China Normal University (Montreal, Canada)** **07/2019-08/2019**

- Responsible for teaching the course COMP 271: Algorithms & Data Structures, introduced algorithms and data structures in computer science, i.e. AVL Trees, Red Black Trees, Greedy Algorithms, Minimum Spanning Trees, Dynamic Programming. Algorithms in Artificial Intelligence (Mini-Max, Monte-Carlo) were also being introduced
- Distributed assignments and exams to students, provided feedbacks and assessments for better understanding the course

***Technical Consultant*, Computer Task Force (Quebec, Canada)** **01/2017-04/2017**

- Dealt with frequently asked questions from students about computer science problems concerning with system, hardware, software development and optimization
- Assembled hardware, including constructed customized computer and repaired facility of Computer Lab

***Secretary*, McGill Student Chinese Brush Arts Club (Quebec, Canada)** **01/2016-05/2017**

- Responsible for scheduling time and place of weekly internal meeting, planning club activities and events, making strategies for contacting and negotiating with sponsors

- Published advertisements and posted passages in several social media platforms to update news and events of the club

***Software Developer*, LIA Engineering Club (Ontario, Canada)** **01/2015-05/2015**

- Created and customized software applications and tools for the 3D printer in school using Java/Python. Prepared presentation on the working of the developed modules and the nature of blending with the other modules
- Participated in the discussion and planning of projects or assignments. Reported progress of programs development

ACADEMIC RESEARCH PROJECT

***Individual Researcher,* Survival Analysis of Nasopharyngeal Carcinoma (directed by Dr. Caixia Li)** **11/2019-12/2019**

- Built survival models for 654 nasopharyngeal carcinoma patients, studied the effects of the prognostic factors, including age, sex, pleural invasion and tumor size
- Found that age and tumor size had a non-linear effect on the risk of recurrence, then employed a logarithmic Gaussian process (GP) for modelling smooth non-linear effects

***Individual Researcher*, State Space Model and Kalman Filters (directed by Prof. David Stephens)** **01/2019-05/2019**

- Introduced principle knowledge, mathematical representations and deductions of State Space Models and Kalman Filters
- Discussed possible improvements of numerical stability of posterior variance matrix in Kalman Filters Algorithm--used sequentially updating the singular value decomposition (SVD) during the evaluation of the variance matrix
- Employed Kalman Filters and State Space models in R-Markdown with Monthly US domestic Enplanements data (1985-1996) and UK gas consumption data (1960Q1 to 1986Q4)
- Presented 36 pages final report including detail process and R-codes in this project

***Team Leader*, Improved Baselines of Published Paper (directed by Prof William L. Hamilton)** **03/2019-04/2019**

The goal of this project was to reproduce and improve the baselines used and referenced in a paper and evaluate whether the benefits of Paragraph Vectors are truly as good as the authors claim them to be

- Reproduced the best baseline (i.e. Naive-Bayes-SVM) referenced in the published paper: *Le & al. Distributed Representations of Sentence and Documents*, the reproduced baseline achieved an accuracy of 92.096% on the test set. This is +0.876% above the baseline reported in the article (91.22%)
- Implemented new baseline using word2vec technique, with careful preprocessing setting and optimized parameters, obtained an accuracy of 94.234%, outperformed Paragraph Vectors in the article (92.58%) by 1.654%

***Team Leader*, Modified MNIST Project (directed by William L. Hamilton)** **03/2019-04/2019**

This project attempts to perform an image analysis prediction challenge on a modified MNIST dataset. The goal is to train the best Deep Learning Model to recognize the number that occupies the most space in an image

- Built and tested multiple Deep Learning Models using different combinations of convolutional layers, tested different data preprocessing techniques including Image Binarization, Segmentation and Rotation
- Submitted the final Deep Learning Model with seven Convoluted-2D Layers and two Hidden Layers, attained validation accuracy of 97.30% and validation loss of 0.1560. Ranked 14 out of 120 groups among the whole class.

***Team Leader*, IMDB Sentiment Analysis (directed by William L. Hamilton)** **02/2019-03/2019**

- Built machine learning models using Logistic Regression, Support Vector Machines and Stochastic Gradient Descent as classifier respectively. And use N-gram and TF-IDF weighting in feature extraction pipelines
- Summited the final Machine Learning Model using Logistic Regression, Tri-gram feature selection and remove 'stop-words', attained prediction accuracy with 91.088%, ranked 8 out of 120 groups among the whole class in Kaggle

***Team Leader*, Predicting the Popularity of Reddit Comments (directed by William L. Hamilton)** **01/2019-02/2019**

This study used a dataset of 12000 instances which was split into training, validation and test sets. Available features included text comments, the number of replies comments received, an indication of whether each comment was the root of a thread and an indicator of controversiality

- Built linear regression model via both Gradient Decent Algorithm and Closed Form Approach to predict the popularity of Reddit comments
- Submitted final model using Closed Form Approach with fine tuning hyper-parameters, achieved MSE of 1.25357 on the test set (very good result). Attained final grade 100% in this project

***Individual Researcher*, Artificial Intelligence Final Project (directed by Prof. Jakie Cheung)** **01/2018-05/2018**

- Employed Java to implement AI search algorithm to resolve Tablut Chess--an ancient chess game. Implemented Minimax Algorithm, Minimax with Alpha-beta Prunning Algorithm and Monte-Carlo Algorithm
- Designed both algorithms for two adversarial chess players, namely Muscovite (Black Chess) and Swedes (White Chess). Users could choose AI of either Muscovite or Swedes to play against AI from other students, or random AI, or Greedy Search AI
- Utilized AI search algorithm to beat both Random AI and Greedy AI with win rate of 100%. Provided A final report consists of Technical Approach, Pros/Cons of the Chosen Approach and Future Improvement