Siyi Wei

Master Student at University of Toronto

📞 (647) 870-8349 | ✉ siyia.wei@mail.utoronto.ca | 🏠 www.wei-siyi.com | ○ superp0tat0 | in siyi-wei

Education

University of Toronto   *Toronto, ON*

Master of Science in Department of Statistical Science   *August 2020 - Present*

University of Toronto Scarborough   *Scarborough, ON*

Honours Bachelor of Science (Graduate with Distinction)   *August 2015 - April 2020*

- Specialist in Statistics, Statistical Machine Learning and Data Science Stream
- Minor in Computer Science, Data Science Concentration
- Extracurricular: RBC AmpHack 2019. UofT DataFest 2019, 2020. Certificate for Algorithms Specialization. Certificate for Natural Language Processing Specialization.

Skills

Languages	Python, R, MATLAB, C#, SQL, Linux, Git
Technologies	TensorFlow, Trax, Scikit-Learn, Numpy, Jenkins, Docker, Selenium
Technical Skills	Statistical Analysis, Machine Learning, Agile Methodologies, Time Series Analysis, Data Mining, Feature Engineering

Experience

Rakuten Kobo   *Toronto, ON*

QA Analyst and Web development   *January 2019 - April 2019*

- Worked on Quality Assurance for new Kobo subscription services using Jenkins, Ruby and selenium. Wrote SQL and Ruby scripts for daily regression testing on production stage.
- Monitor the QA automation status and bug tracking system for 4 months, prevent service failure for multiple times and received praise from team lead and promoted to work on web development and ETL pipeline for new subscription.
- Developed web page and ETL pipeline using C# and MongoDB via CI/CD pipeline, automated the data collection process and Real-time display of subscription data. Cooperate with release manager to deliver stories during daily release meetings.

University of Toronto Scarborough   *Scarborough, ON*

Teaching Assistant   *August 2018 - Present*

- Hold weekly Tutorials and Office Hours to teach critical principles and answer students' questions. Built programming assignments for Machine Learning Courses and auto-marker for assignments. Wrote blogs on interesting topics to deepen understands. Received praise from multiple professors and thus involved in some future research.
- Taught multiple times for the following courses; Introduction to Machine Learning and Data Mining. Advanced Machine Learning and Data Mining. Stochastic Process. Introduction to Programming. Linear Algebra I&II.

Projects

Improved Monte Carlo Tree Search (Self-directed Project for Monte Carlo)

- Implement the naive Monte Carlo Tree Search (MCTS) algorithm in Python. Test it with multiple board games include 3x3, 4x4 Tic-Tak-Toe and modified Gomoku. Achieve a high winning rate in those games with a decent time complexity.
- Identify the disadvantage of naive MCTS in larger boards like 15x15 Gomoku/GO game board. Inspired by importance sampling Monte Carlo method. Then improve the simulation process of naive MCTS and make it workable on larger game board. Which balance the trade off of winning rate and time complexity.
- Inspired by Alpha Go. Currently working on replace Monte Carlo simulation by neural network. This could significant improve the efficiency on estimating the winning rate.

Data Science Project (Course Project for Big Data Analysis)

- Researched on ridge and lasso regression and implement them for gravity wave regression and feature selection using R. Implement feature engineering on predicting house price. Used logistic regression on diabetes classification and confusion matrix for further analysis. Received the highest grade among the class.
- Investigated current controversy on the ideology of Batch Normalization. Concluded the latest findings on the relation between Internal Co-variate Shift and Batch Normalization. Analyzed for different regularization method like Layer Normalization, further researched their ideologies and corresponding pros and cons. More details are in my blog.

Automated external feature selection for ARIMA (Methodology Project for Time Series)

- Investigated default auto.arima function in forecast package, and found the defects for further improvement. Improved the automation process by preventing overfit on external resources and saved tremendous time for data selection.
- Developed the automation program for the external resource selection in GDP deflator prediction and verified by time series Cross Validation. The improved model decreases the mean absolute percentage error rate from 11.77% to 9.37% for Oil industry. The automation process successfully generalized the improvement for other industries