

Jiawen (Erica) Li

778-869-0715 | kizuna021997@gmail.com | <https://www.linkedin.com/in/jiawen-li-898851118/> | <https://erica97.github.io/>

EDUCATION

Columbia University , <i>New York, United States</i>	2021 - 2022
Master in Operations Research, Machine Learning & Artificial Intelligence Concentration	
University of Toronto , <i>Toronto, Canada</i>	2016 - 2021
Bachelor of Applied Science, Major in Industrial Engineering	Major GPA: 3.61/4.00
Minor in Engineering Business & AI Engineering	

SKILLS

Data Analytics & Visualization	R (optimization packages, ggplot2), Python (Pandas, NumPy, SciPy, seaborn, matplotlib)
Programming & Database	HTML/CSS, JavaScript, C#, JAVA (JDBC), SQL (SQL Server, Oracle, SQLite), Spark, SAS 9
Machine Learning	Python (PyTorch, TensorFlow, Scikit-learn, nltk), Linear & Logistic Regressions, Tree-based Models, Classifiers, Support Vector Machine, Neural Networks (CNN, RNN, LSTM)
Tools	AWS (SageMaker, S3), MS Excel, Tableau, Power BI, MATLAB, Minitab, Git, Linux
Certificates & Awards	AWS Solutions Architect Associate (Candidate), SAS 9 Advanced Programmer, RBC AmpHacks 2019 Hackathon Top 4 Finalist

PROFESSIONAL EXPERIENCE

Technology Analyst, Summer Intern Citi Group , <i>Mississauga</i>	May - Aug 2020
<ul style="list-style-type: none">Brainstormed and prototyped 2 Fintech application demos, then pitched to the manager. <ol style="list-style-type: none">Citi-Fund: Utilized APIs to extract stock price data from <i>Yahoo Finance</i>. The Python app could take in user-defined stock portfolios and find the minimum variance portfolio using Pandas and NumPy, and plot the efficient frontier using matplotlib.Citi-Balance: A money management app which allows users to record the amount and attitude about every expense, with the front end implemented by HTML/CSS and the back end implemented by Node.js and MongoDB.	
Research Assistant, Part-time Intern Röst Lab (http://www.roestlab.org/), <i>Toronto</i>	Jun 2020 - May 2021
<ul style="list-style-type: none">Performed exploratory analysis and visualized mass spectrometry data (high dimensional multi-channel time series data) in Python using seaborn and matplotlib.Imported and cleaned new experiment data into SQLite database. Assisted in creating Python scripts which parsed mass spectrometry data and trained a convolutional neural network model on the parsed data.Including the model prediction scores in Röst Lab's mass spectrometry data analytic tool, Swath-MS, increased peptide detection accuracy by 15% based on 100+ samples.	
Corporate Real Estate Data Management, Summer Intern City of Toronto , <i>Toronto</i>	May - Aug 2019
<ul style="list-style-type: none">Collected and aggregated data from various sources (SAP/spreadsheets/etc.), then performed data cleaning and ETL process using Excel and SQL for data reconciliation. File search time was optimized by 30% as a result.Eliminated data redundancy and error by 60% with normalized data models and redesigned relationship in SQL Server.Created and maintained 5 insightful Tableau Dashboards to enable dynamic KPIs tracking for internal consulting.	

RELEVANT PROJECTS

Natural Language Processing Analysis on Twitter Analysis	Apr 2021
<ul style="list-style-type: none">Utilized POS-tagging, stop words and stemming for building the term matrix by nltk in Python.Tokenized 1000+ text files to extract terms from reviews and performed sentiment analysis to score terms to inspect whether the tweets were positive or negative.	
Evaluating Schedules with Discrete Event Simulation (Capstone Project) Ceridian	Sep 2020 - May 2021
<ul style="list-style-type: none">Implemented a proof-of-concept discrete event simulation JAVA application from scratch. The prototype could simulate any queuing system given customer arrival information and worker schedules, then project key performance metrics.The application could be used to maximize server utilization rates and reduce customer wait times and therefore was estimated to help the client in our case save more than \$500,000 per year.	
Chest X-ray Diagnosis on Pneumonia based on deep learning	Jul - Aug 2020
<ul style="list-style-type: none">Preprocessed 5000+ chest X-ray images of 3 classes: normal, viral, bacterial. Used Generative Adversarial Network to generate artificial training data for imbalanced dataset.Implemented and performed parameter-tuning for 3 candidate models: SVM, CNN, transfer learning model with <i>AlexNet</i>.Evaluated model performances against Confusion Matrix and F1 score. The best model CNN achieved a test accuracy of 93%.	