Yongya Li

1101 Bay St, Toronto, ON M5S 2W8
E-mail: liyongya6@gmail.com **Cell phone:** (905) 324-8689

Highlights of Qualifications

- Data scientist with a strong analytics background and 3 years of research experience in Biostatistics and Cheminformatics
- Financial data science skills developed as a Data Scientist at Vhinny
- Proficient in building statistical models with complex and large datasets, including machine learning, clustering, NLP, deep learning, recommender system, time-series analysis, A/B testing design

Technical Skills

- **Programming Languages**: Python, SQL, R, Java, MATLAB, HTML, JavaScript
- **Analytical Tools**: Pandas, NumPy, Tableau, Looker, Power BI, MySQL, PostgreSQL, DynamoDB, BigQuery
- **ML Cloud Platforms**: Amazon SageMaker, Google Colab, Microsoft Azure, IBM Watson Studio

Education

Master of Science, Chemistry, Brock University (CGPA: **4.0/4.0**)   May 2018 – Aug 2020
- Research-based, emphasis in Genomic Data Science, Bio-analytics and Cheminformatics

Honors Bachelor of Science, Chemistry, Brock University (CGPA: **3.73/4.0**)   Sep 2015 – Jun 2018
- Distinguished Graduating Student Award, First-Class Standing, Dean's Honors List – 2017 & 2018

Professional Experience

Data Scientist (Contract), Vhinny, Toronto (vhinny.com)   Jun 2020 - Current
- Collect, study, and interpret financial information; perform accurate financial data management
- Develop and implement models for financial analysis, stock price prediction and portfolio optimization

Data Analyst Intern, Anogen-Yes Biotech Laboratories LTD., Mississauga (anogen.net)   Jan 2020 – Jun 2020
- Supported sales analysis and customer behaviour analysis in the Business Development department
- Extracted and analyze bio-pharmaceutical product information from a variety of databases using SQL and Excel
- Built and maintained an interactive Tableau dashboard for E-commerce sale analysis and website traffic analysis

Research Assistant, Feng Li Research Group, St. Catharines (ligroupbrocku.com)   May 2017 – Aug 2020
- Designed and implemented experiments, extracted useful data from raw datasets, removed outlier noise, normalized data into a standardized format that allows further processing and modelling
- Contributed to 5 bio-analytical research project developments as the project leader and first author

Projects

Sentiment Analysis for Hotel Reviews
- Developed a sentiment analysis model to understand customer experience from review using Doc2Vec and Tf-idf
- Detected negative reviews using logistic regression classifier for hotel service improvements

Recommender System for E-commerce Products
- Built a recommender model to attract customers, boost sales, improve customer experience and retention
- Combined demographic, content-based and collaborative filtering techniques with NLP and neural network embeddings to construct a hybrid system for product recommendations

Drug-resistant Tuberculosis Diagnosis
- Built a simulation-guided model using MATLAB to detect Tuberculosis drug resistance caused by a single nucleotide mutation (published in *Chemical Science*, impact factor 9.5, DOI:10.1039/C8SC02761G)

Grapevine Virus Detection
- Built an automated diagnosis based on the image classification model to detect grapevine virus infection
- Utilized CNN models to classify clinical sample pictures and deployed model in mobile setting for portable testing