Jing (Lydia) Li

lydia001024@gmail.com (217) 328-6042
202 E Daniel St, Champaign, United States

EDUCATION

University of Illinois at Urbana-Champaign *Sep 2023 - May 2024 (Expected)*
Master of Science in Business Analytics Champaign, US

- **Relevant Courses**: Machine Learning in Finance, Big data Analytics, Enterprise Database Management, Data Story Telling.

University of Toronto, Scarborough *Sep 2018 - May 2022*
Bachelor of Science (Major Program in Statistics) Scarborough, Canada

- **Relevant Courses**: Measure-based Probability, Advanced Statistical Inference, Regression Analysis, Survey Sampling and Observe Data, Time Series Analysis.

Bachelor of Arts (Major Program in Economics for Management Studies) Scarborough, Canada

- **Relevant Courses**: Microeconomics, Macroeconomics, Price Theory, Quantitative Methods in Economics.

SUPPORTING ABILITIES

- **Languages**: Mandarin *(Native)*, English *(Proficient)*.
- **Technical Skills**: Python, R, SQL, Microsoft Office Suite *(Word, Excel, PowerPoint)*.
- **Statistical Skills**: Regression Analysis, Time Series Analysis, Machine Learning, Hypothesis Testing, and Correlation Test.

PROFESSIONAL EXPERIENCE

PingAn Bank *Feb 2023 - Aug 2023*
Data Analyst, Operation Big Data Department Shenzhen, China

- Built a tight connection between business requirements and data engineering.
- Translated business requirements into actionable data tasks, including specific fields extraction and generation of target tables using SQL.
- Developed an advanced optimization model, resulting in a 30% enhancement in the efficiency of the online credit card business channel.

WeBank (Tencent) *Aug 2022 - Dec 2022*
Data Scientist, AI Department Shenzhen, China

- Proposed an evaluation framework formulating the trade-off between privacy leakage and utility loss of a Vertical Federated Learning *(VFL)* system, treating the framework as a basis for standardizing the FL evaluation process.
- Applied the proposed evaluation framework to guide the comprehensive evaluations on various protection mechanisms against most of the state-of-the-art privacy attacks in VFL.
- Built an extensible code-base of the evaluation framework for evaluating current and future privacy protections and attacks.

SMY Digital and Technology Corporation *May 2022 - Aug 2022*
Data Scientist, AI Department Shanghai, China

- Generated a model to predict if a user's car insurance is fraudulent, based on the car insurance data of customer claims.
- Pre-processed data by label encoding the learning target"fraud", then built and trained the model using CatBoost classifier, Stratified K-fold, and calculated the prediction value.
- Improved and evaluated the model by AUC value and reached approximately 0.95, indicating the high efficiency of the model.

Talking Data *May 2021 - Jun 2021*
Financial Consultant, Financial Department Beijing, China

- Spearheaded the benchmarking process for the user rating system of Bank of China's Credit Card application through an in-depth competitive analysis of 12 market-leading rating systems.
- Formulated a comprehensive user data tracking strategy for Hebei Bank, ensuring data integrity and relevance.
- Implemented an employee training plan for Hebei Bank, leading to a projected reduction of 30% in junk data utilization.

ACADEMIC PROJECTS

Machine Learning Classification Analysis *August 2023 - Present*
Graduate Course Project: BADM550 Business Practicum Champaign, United States

- Accessing to 186,000 historical new stock request *(NSR)* items labeled with their outcome corresponding to the success of the sales at Graybar Electric Company, Inc.
- Developing a predictive model, including Catboost Classifier model, to accurately forecast the sales success of items.
- Integrating the predictive model into the decision-making process, optimizing approval strategies for new stock requests.

Statistical Analysis of Health Commodity Supply Chain & Pricing *May 2022 - Aug 2022*
Personal Research Project Scarborough, Canada

- Analyzed the supply chain of health commodity shipment, specifically Antiretroviral *(ARV)* and HIV, and eliminated time trend by calculating the difference between import and export time.
- Fit four different statistical models including smoothing splines, random forest, boosting and neural network.
- Acknowledged Xgboost is a proper method for reasonable training time and prediction error, and Neural Network is suitable for large training data.

Case Study on Time Series Analysis *Jan 2022 - Apr 2022*
Undergraduate Course Project: STA457 Time Series Analysis Scarborough, Canada

- Collected weekly mortality statistics published by the Canadian federal government from 2019 to the first 8 weeks of 2020.
- Performed seasonal difference on original data and proposed 4 SARIMA models by observing the ACF and PACF plot.
- Selected the model with the smallest AIC value to predict the following 12 weeks, compared the predicted mortality with the true mortality, concluded the excess mortality, and generated a report.

Case Study on Regression Analysis *Jan 2022 - Apr 2022*
Undergraduate Course Project: STAC67 Regression Analysis Scarborough, Canada

- Implemented a linear regression analysis on the 'Brown Fat' dataset, filtering out non-essential data for a refined model.
- Eliminated uncorrelated variables by observing Adjusted R^2, p-value, scatter plot and correlation tests.
- Obtained the optimal model through Backward Elimination and spotted the covariates influencing the volume of Brown Fat from the last model.

EXTRACURRICULAR EXPERIENCE

UTSC Concert Band *Jan 2019 - Apr 2022*
Flute player and Flute Section Leader Scarborough, Canada

- Organized weekly flute sectional rehearsal and provided constructive feedback for members.
- Performed multiple concerts and chamber music recitals, assumed flute solo part in several Repertoires.