Yiwen Li

Toronto, Canada | +1 437-989-7435 | yiwenli2001@gmail.com | linkedin.com/in/yiwen-li-8671b7209

EDUCATION

University of Toronto Sept 2019 - Jun 2023

Bachelor of Science in Statistics and Economics (double majors) | ***GPA: 3.99/4.00***

- **Relevant Courses:** Machine Learning (ML), Time Series Analysis, Python Programming, Advanced Data Analysis, Methods for Multivariate Data, Advanced Calculus, Advanced Linear Algebra, Probability, Financial Economics
- **Awards**: Science & Mathematics Scholarship (Top 1%), Reuben Wells Leonard Scholarship (Top 1%), C. L. Burton Open Scholarship, U of T Special Admission Scholarships, Dean's List Scholar (3 years)

INTERNSHIPS

Jd.Com, Inc Aug 2022 - Dec 2022

Data Analyst *Beijing, China*

- Implemented automatic workflows in **Power Query** to standardize large-scale data and generate weekly reports within 5 seconds, empowering faster data-driven marketing decisions through 20+ KPIs like conversion rate
- Optimized processing efficiency for 100K+ eCommerce data rows using advanced Excel functions (**VLOOPUP**, **Pivot Tables)** and Python **Pandas**, enabling accurate data extraction and reducing human error rates by 30%
- Created 10+ **Tableau** validation dashboards to ensure consistency between **SQL** queries and reporting logic
- Increased business sales by 25% through optimizing logistics resource allocation with an in-depth quantitative analysis of regional gross profit and customer engagement, supporting cross-functional teams in establishing a new warehouse

UnionPay May 2022 - Jul 2022

Data Analyst *Shenzhen, China*

- Designed 20+ **ETL** (Extract, Transform, Load) data pipelines in **SQL** to process 600K transactional data from financial institutions, increasing reporting efficiency by 40% and saving the team 8 hours per week
- Assessed annual company performance in **Python** by benchmarking 200+ business indicators, with Exploratory Data Analysis (**EDA**) to examine distributions and the Kruskal-Wallis test to identify statistically significant indicators
- Built a risk prediction model (**Logistic Regression**) to detect suspected cash-out activities with a 90% accuracy, informing the executive board of high-risk merchants and recommended solutions through clear **presentation**

RESEARCH EXPERIENCES

Surrogate Assisted Positive Unlabeled Learning on EHR data May 2023 – Present

Supervisor: Prof. Jessica Gronsbell *Toronto, Canada*

- Developed a **semi-supervised ML** algorithm for phenotype prediction with an **AUC** score exceeding 93 that outperformed all 7 baseline models, offering a substantial and accurate solution to reduce manual chart-review efforts for data labeling
- Achieved robust feature selection with adaptive **LASSO** and automated hyperparameter tuning with **R** (*glmpath*)
- Conducted **NLP** analysis on real doctor notes to extract disease-indicative terms using the Unified Medical Language System, improving the model's predictive accuracy by 10%
- Demonstrated model effectiveness in handling **high-dimensionality** through extensive model robustness testing on 42K Electronic Health Records (EHR) from the MIMIC database with a generation of 1100 covariates
- Presented research poster to 300+ professionals and published the model as an **R package** (*SAPUL*), contributing to the open-source development and allowing the research community to reproduce statistical findings

Monty Hall Meets AI: The Influence of AI on Decision-Making Apr 2023 – Present

Supervisor: Prof. Boris Babic *Toronto, Canada*

- Coded survey questions in Qualtrics to collect **multi-level data**, facilitating the evaluation of participants' trust in AI-generated suggestions for the Monty Hall problem via quantitative and qualitative metrics
- Established a connection between **Prolific** and **Qualtrics**, streamlining the recruitment process of 2K participants
- Enabled dynamic survey design by leveraging predefined hypotheses based on **Bayes theorem** to find optimal solutions and implementing logic structures to target different participant segments

Regulatory and Ethical Concerns on AI-based Medical Devices May 2023 – Present

FORCOLAB | Supervisor: Prof. Shurui Zhou, Prof. Boris Babic *Toronto, Canada*

- Synthesized a comprehensive literature review focusing on algorithms employed in AI-based medical devices and data challenges throughout development and deployment phases, providing valuable insights into state-of-the-art approaches
- Interviewed stakeholders from medical companies on the AI-based medical **software development lifecycle**, identifying the need for improving existing regulatory policies and refining best practices

Curriculum Development: Fundamentals of AI, Data, and Algorithms Aug 2023 – Present

Supervisor: Prof. Boris Babic *HongKong, China*

- Contributed to the development of a new Master's program at the HKU Data Science Institute, structured **LaTeX** course materials on Overleaf, covering topics on Probability, Bayesian Inference, R programming, and advanced ML algorithms
- Crafted 15 scaffolding **R scripts** and **Jupyter Notebooks** to illustrate data and model details, deepening students' understanding of the course content and enhancing their engagement with interactive visualizations

SKILLS

Programming & Software: R (4 years), Python (3 years), SQL (2 years), Microsoft Office Suite (Excel, PowerPoint, Word)

Languages: English **(**IELTS 7.5, GRE 333), Mandarin, Cantonese