

Kaibo Zhang

+1 (438)8650693 | kb.zh@outlook.com | 45 Hayward Street, MA, US | [LinkedIn](#) | [Personal Website](#) | [GitHub](#)

EDUCATION

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Master of **Business Analytics**, Sloan School of Management

MCGILL UNIVERSITY

B.Com. in **Business Analytics**, Desautels Faculty of Management

- Minor in **Statistics**, Department of Mathematics and Statistics

- Scholarships and Awards: Dean's Honour List – McGill (top 10%), Accenture Prize in Management (\$1,590)

Cambridge, United States

Aug. 2025 – Aug. 2026 (Expected)

Montreal, Canada

Aug. 2022 – May 2025

GPA: **3.95/4**

WROK EXPERIENCE

Data Modelling and Machine Learning Engineer

Alimentation Couche-Tard Inc. & McGill University (Capstone)

Montreal, Canada

Jan. 2025 – May 2025

- Built an **Integer Optimization model** in Gurobi to design combo offers that maximize expected profit and customer-product affinity, **driving a 15% increase** in Net Promoter Score compared to the control group.
- Trained an ensemble forecasting model using **Regression** and **pre-trained NeuralProphet** to predict 6-month sales for the identified products, achieving an RMSE of 247.27 and outperforming the baseline model by 23%.
- Applied **Association Rule Mining** on previously untapped POS data to uncover consumer preferences in high-frequency market baskets, identifying 39 top-selling candidates from 5,000+ products.

Machine Learning and AI Engineer Intern

Bosera Asset Management Co. Limited

Shenzhen, China

May 2024 – Aug. 2024

- Integrated **K-Means** algorithms into a **Multi-Agent** workflow using Python to extract and cluster performance metrics from research reports, **reducing operation time by 10%** by enabling efficient comparison of companies.
- Designed and implemented **relational database** and APIs with SQLAlchemy and FastAPI to facilitate efficient **LLM** data retrieval, achieving 85% accuracy in critical information retrieval and outperforming that of sample responses by 12%.
- Researched and enhanced **Retrieval Augmented Generation** performance with Landmark Embedding and tree-organized Natural Language Retrieval workflow, improving semantic and answering accuracy by 45%.

Data Scientist Intern

VariFlight Technology Ltd

Remote

Sept. 2023 – Dec. 2023

- Utilized **Mixed-Integer Optimization** and sensitivity analysis on resource constraints to discover revenue management insights for airlines and airports, projecting a **4% increase in profit**.
- Designed 3 **MySQL** relational databases and built 4 **ETL** pipelines with Python to load API data for model training and weekly report publications, **automating 71% +** of manual, repetitive tasks in data extraction and cleaning.
- Implemented interactive features for 2 existing visualization dashboards with **Power BI** and **Python**, simplifying graphs and enhancing readability.

Data Scientist Intern at Online Ride-hailing Area (Urban Central Platform)

Beijing DiDi Infinity Technology and Development Co.

Tianjin, China

May. 2023 – Aug. 2023

- Cleaned 5 datasets with **SQL** and built data visualization and **Regression** models with Python for online ride-hailing operational needs, unifying data formats and helping to **put 13% of stale data** into use.
- Built data automatic processing models that execute **SQL** queries to update datasets from source data at preset times, resulting in a **30% increase in efficiency**.
- Led and delegated tasks for colleagues to collect first-hand online ride-hailing data, and used **hypothesis testing** for measuring price-related market standings with competing brands.

RESEARCH EXPERIENCE

Research Assistant at Professor Setareh Farajollahzadeh's Lab

Timing the Release: A Study on Market Crowdedness and Book Launch Outcomes

Montreal, Canada

Oct. 2024 – Now

- Employed **BERT**, **UMAP**, and **HDBSCAN** to generate multilingual embeddings and cluster novels by textual features, downscaling 4,889 novels into 140 genres and labeling them with meaningful keywords using **c-TF-IDF**.
- Constructed a **Cosine Similarity Matrix** from multilingual embedding vectors to retrieve the top 6 semantically distinct topics, enabling downstream feature engineering for modeling pre-release sales trends of comparable genres.
- Applied offline **Bayesian Changepoint Detection** using a customized **Poisson-Gamma Prior** to identify exogenous shocks in user order time series, supporting analysis of lagged effects on other novels' sales trajectories.

SKILLS&ACTIVITIES

- Technical Skills: Python (Pandas, Sklearn, Torch, Matplotlib, etc.), MySQL, R, Neo4j, Git, Slurm, JavaScript, Java, HTML, CSS, Canva, Figma, Microsoft Office Suites
- Interests: Photography, Guitar, Alto-saxophone, and Piano