

# Hai Bao

647-831-8623 | hbao12@gmail.com | hbao.ca

## SUMMARY STATEMENT

Senior Data Scientist with a background in Statistics and expertise in LLM integration, Python, and data visualization. Proven track record of transforming complex datasets into actionable insights that improved operational efficiency by 50%. Adept at communicating technical findings to diverse stakeholders to drive strategic decision-making and program optimization. Committed to leveraging data-driven approaches to drive business value and efficiency.

## EXPERIENCE

### Senior Data Scientist

**Jan 2023 - Mar 2026**

*Morningstar Sustainalytics, Toronto, ON*

- Responsible for maintaining the News Harvester System, which is a system that imports 400,000+ articles daily and filters them out for ESG relevance for ESG Analyst review.
- Integrated LLM-based solution (GPT) into the News Harvester System to help filter the news articles more efficiently, which resulted in the reduction of articles daily by 25% (approximately 1000 less articles per day).
- Created a Plotly dashboard in Python to monitor daily metrics and assist with anomaly detection. The dashboard is consumed by stakeholders through a daily automated email.
- Evaluated model performance using scikit-learn.

### Data Scientist

**Nov 2021 - Jan 2023**

*Morningstar Sustainalytics, Toronto, ON*

- Responsible for maintaining the Thematic Crawler (web/internal database crawler).
- Improved text mining performance of the crawler through use of multilayered keywords. Was able to reduce the number of hits by approximately 50%, effectively reducing workload for ESG Analysts in half.

### Network Capacity Planner

**Dec 2014 - Apr 2020**

*Zayo, Toronto, ON*

- Responsible for identifying cost optimization opportunities through SQL queries of the network equipment inventory database.

- Transitioned opportunities into realizing cost savings through implementation of network engineering projects (node installations/decommissions, customer circuit migrations).
- Worked on a project to vacate 2 floors of a data center in order to realize \$2 million dollars worth of cost savings.

## EDUCATION & CERTIFICATIONS

**Bachelor of Science**, University of Toronto

**Sep 2005 - Aug 2010**

*Statistics Specialist*

**Machine Learning**, IBM

**Sep 2021**

*Coursera*

**Data Analytics**, Google

**Jul 2021**

*Coursera*

**Natural Language Processing**, DeepLearning.AI

**Dec 2021**

*Coursera*

**Tableau Desktop Specialist**

**Nov 2020**

## SKILLS

Python Programming

Data Analytics

Data Visualizations

Machine Learning Modelling

(XGBoost, Random Forest)

Gen AI, LLM Integrated Solutions

Github