HANAN ATHER

https://hananather.com https://www.linkedin.com/in/hanan-ather/ https://github.com/hananather

EDUCATION

University of Ottawa

Master of Science Mathematics and Statistics Concentration in Statistics

2021-2024

University of Ottawa

Honours Bachelor of Science - Mathematics & Statistics

2015-2020

PUBLICATIONS

- **Reinforcement Learning and Function Optimization**
- Large Deviations and the Law of the Iterated Logarithm
- Big Data and Sentiment Classification via Apache Spark

EXPERIENCE

Mathematical Statistician

Statistics Canada

July 2022- Present

- Developed de-duplication and entity resolution system by **fine-tuning encoder-only transformer models**. Project is scheduled to be presentated at the **2024 Statistics Canada Symposium**.
- Leading the initiative on **fine-tuning open source LLMs** such a **Llama 3.1** via **LoRa**.
- Translated and optimized legacy data processing queries from SAS to **Python** achieving a **50% reduction in run time**.
- Developed a **SQL-based data pipeline** for error detection for 32 million+ businesses in Canada.
- Designed and implemented a data integration system for large datasets. This pipeline processes between **10 million to over 100 million rows of data monthly**, used by both external and internal federal government clients.
- Conducted theoretical research on outlier detection methods, making mathematical adjustments to enhance state-of-the-art techniques, and implementing them in **Python** to assess survey quality (deployed the tool internally for statisticians).

Data Scientist

Treasury Board of Canada Secretariat

December 2020 - July 2022

- Automated routine data processing tasks using Python scripting and ETL tools, previously done with Excel macros, **saving 2 weeks of manual effort per fiscal year** and increasing overall productivity by streamlining workflows in the data ingestion process.
- Fine-tuned **semantic search** with **BERT** to audit data using **Python** enabling Treasury Board senior management to extract insights from thousands of comments through **topic clustering** and **sentiment analysis**. These insights were presented to over **35+ department heads** in Government of Canada at the **Departmental Audit Committees Symposium**.
- Built and maintained a **Flask** application in **Python** to facilitate data labeling and model training for machine learning.

EXPERIENCE

Data Science Writer

Medium & Analytics Vidhya

September 2020

- **Most Recent Article**: End-to-End Guide: Creating a Web Application using Dash

Machine learning Researcher

University of Ottawa, Department of Mathematics and Statistics

April 2020- September 2020

- Leveraged **Apache Spark** and AWS Elastic Map Reduce (EMR) to analyze massive text datasets.
- Deployed multiple word embedding algorithms and tested the performance of various statistical machine learning algorithms to classify the sentiments of documents.

Undergraduate Researcher

University of Ottawa, Department of Mathematics and Statistics

May 2019- August 2019

- Automated Data Collection for NLP; Built a pipeline for automating collection of unstructured data through web-scraping and other sources in R.

Research Assistant

University of Ottawa, Department of Cellular and Molecular Medicine

May 2018- August 2018

- Generated image data and automated image recognition using machine learning scikit-learn.
- Performed multivariate statistical data analysis at KIM lab.

AWARDS

- **University of Ottawa Dean's Honour List**
- CANDEV Data Challenge Top Qualifier (2020)
- Ontario Volunteer Service Award (2019)

SKILLS

Languages	Frameworks & Tools
Python	**PySpark**, Flask, TensorFlow
R	**Tidyverse**, dplyr
Java	Eclipse
SQL	PostgreSQL, MySQL
Scala	**Spark**