

CURTIS LI

4B STATISTICS AND COMPUTATIONAL MATHEMATICS, CO-OP

✉ sz7li@uwaterloo.ca 🌐 sz7li.github.io ☎ 647-272-6690 in shao-curtis-li 🐙 github.com/sz7li

Languages and Technologies: Python, R, SQL, C, C++, PySpark, Pandas, Scikit-learn, Keras, Git, Docker, Bash

Skills: Natural Language Processing, Machine Learning, Data Analysis, Web Development

Relevant Courses: Data Types and Structures, Obj. Oriented Software Development (C++), Statistical Learning - Classification, Deploying Machine Learning Models in Production (Coursera)

EDUCATION

UNIVERSITY OF WATERLOO • Bachelor of Mathematics - Honours, Co-op

Sep 2017 - Dec 2022

- Achieved **Academic Term Distinction** from Spring 2020 - present
- **Leadership and Extracurriculars:**
 - Volunteered on Faculty Orientation Committee of 4 and managed all logistics of the Math Faculty's OWeek in 2021 for 1900 incoming students and led 70 upper-year Orientation Leaders

EXPERIENCE

ThinkData Works • Data Science Co-op

Toronto, ON • May - Aug 2021

- Developed features for a NLP Entity Resolution application using **PySpark, Hadoop, PostgreSQL, GCP**, and **Docker**.
- Updated **Pytest** Continuous Integration and Unit Testing modules, increasing test coverage for new **Spark** features
- Automated dependency tracking for **Spark jobs**, improving testability of legacy modules without cluster downtime
- Investigated **multi-threaded performance** of sparse matrix multiplication; monitored and gathered data on CPU/memory usage under different configurations - identifying **speed-up of 22%**

Canadian Imperial Bank of Commerce • Data Scientist Co-op

Toronto, ON • Jan - Apr / Sep - Dec 2020

- Developed features for a document summarization Web App for CIBC Capital Markets using **React, Flask**, and **PostgreSQL**
- Reproduced existing research on the YouTube 8m Dataset by experimenting with building and training **CNN** architectures
- Used Google Places API to extract reviews data from CIBC branches and performed unsupervised text clustering
- **Created visualizations** of topic, sentiment trends in Google Data Studio and presented findings to prominent stakeholders
- Built and trained sentiment classifiers using **Scikit-learn** and **Keras** on SemEval datasets to analyze stock trends
- Developed semantic search functionality using **Sentence-BERT** in Covid response Web App used by CIBC Contact Center
- Incorporated keyphrase extraction using RAKE, EmbedRank, and spaCy dependency parsing to improve search relevance

KPMG • Summer Student

Toronto, ON • May - Aug 2019

- Assisted in the **T661 Scientific Research & Experimental Development** claim process for Canadian technology firms
- Gathered and compiled technical documentation to justify SR&ED eligibility and drafted technical sections of T661 forms
- Performed data entry in Excel and developed automation VBA scripts for Microsoft Word

DaVinci Assortment Solutions • Technical Security Analyst

Toronto, ON • Sep - Dec 2018

- Developed features for inventory Web App with **AngularJS, Spring, PostgreSQL** and built **REST APIs** for user management
- Experimented with the implementation of **token-based authentication** using **JSON Web Tokens**
- Upgraded existing user authentication process to incorporate LDAP and Microsoft Azure AD credentials; improving scalability of new user configuration in steps **from linear to constant**

PROJECTS

KAGGLE COMPETITION - PAWPULARITY

- Used a **Transformer-based** image classification model with **Keras** to predict "Pawpularity" scores from pet photos
- Trained ensemble model using **TPUs**, achieving a 17.847 RMSE score in the **93rd percentile** out of ~3100 teams

COVID-19 CLASSIFIER

- Created a **CNN model** using **Keras** to detect cases of COVID-19 from chest X-Ray scans
- Trained model using the Mirrored-Distributed Strategy on custom **GCP** cluster with GPUs, achieving a **.95.7%** accuracy