CURTIS LI

4B STATISTICS AND COMPUTATIONAL MATHEMATICS, CO-OP

sz7li@uwaterloo.ca sz7li.github.io 647-272-6690 shao-curtis-li github.com/sz7li

Languages and Technologies: Python, R, SQL, C, C++, Pandas, Numpy, Matplotlib, PySpark, Keras, Git, Docker, Bash
Skills: Natural Language Processing, Machine Learning, Data Analysis, Software Development
Relevant Courses: Data Structures, Object Oriented Software Development, Big Data, Statistical Machine Learning, Forecasting

EDUCATION

UNIVERSITY OF WATERLOO • Bachelor of Mathematics - Honours, Co-op
Expected Dec 2022

- Achieved **Academic Distinction** from Spring 2020 - present
- **Leadership and Extracurriculars:**
  - Orientation Committee of 4 - managed logistics of Math O-Week 2021 for 1900 students and 70 Orientation Leaders

EXPERIENCE

Headspin • Data Science Team
Palo Alto, CA (Remote)
Engineering Intern
May - Aug 2022

- Researched, developed, and deployed 2 new image matching features using **OpenCV, keypoint,** and **template matching**
- Collaborated with the team to identify customer use cases and curated labelled dataset to evaluate against success criterias
- Conducted independent research and applied existing literature on object detection and image retrieval algorithms
- Documented experiments and iterations by presenting visualizations and animated plots of video data
- Created Objected-Oriented modules and Unit Tests in the HeadSpin platform using **Python** under the **MVC** framework
- Shipped features to production, improving image matching by adding invariance to rotation, scale, or background changes

ThinkData Works • Data Labs Team
Toronto, ON
Data Science Co-op
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 - **improving computation speed by 22%**

Canadian Imperial Bank of Commerce • Data Studio / Enterprise Innovation Team
Toronto, ON
Data Scientist Co-op
Jan - Apr / Sep - Dec 2020

- Developed features for a document summarization Web App for CIBC Capital Markets using **React**, **Flask**, and **PostgreSQL**
- Applied 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 • Tax Incentives Practice Team
Toronto, ON
Summer Student
May - Aug 2019

- Assisted in the **T661 Scientific Research & Experimental Development** claim process for major Canadian technology firms
- Gathered and compiled project documentation to articulate 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 • Engineering Team
Toronto, ON
Technical Security Analyst
Sep - Dec 2018

- Developed features for inventory Web App with **AngularJS**, **Spring, PostgreSQL** and built **REST APIs** for user management
- Implemented **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**