CURTIS LI

4A 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, Web Development, OOP

EXPERIENCE

DATA SCIENCE CO-OP · ThinkData Works   Toronto, ON · 05/2021 - 08/2021
- Developed features for a NLP Entity Resolution application using **PySpark**, **Hadoop**, **PostgreSQL**, **GCP**, and **Docker**.
- Implemented new address-matching feature and documented tradeoffs in precision, recall, and overall performance
- Updated **Pytest** 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 and Memory usage under different configurations and scenarios

DATA SCIENTIST CO-OP · Canadian Imperial Bank of Commerce   Toronto, ON · 09/2020 - 12/2020
- Developed features for a document summarization Web application using **React**, **Flask**, and **PostgreSQL**
- Used Google Places API to extract reviews data and performed text clustering using **Unsupervised Learning**
- **Created visualizations** of sentiment, traffic, and topic trends using dashboards in Google Data Studio, and presented findings to prominent stakeholders
- Collaborated with teams across CIBC to write and post company-wide weekly newsletters on tech trends and innovations

DATA SCIENTIST CO-OP · Canadian Imperial Bank of Commerce   Toronto, ON · 01/2020 - 04/2020
- Scraped and cleaned news articles and text data from public financial websites using **Selenium** and **BeautifulSoup**
- 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 extracting using RAKE, EmbedRank, and spaCy dependency parsing to improve search relevance

TECHNICAL SECURITY ANALYST · DaVinci Assortment Solutions   Toronto, ON · 09/2018 - 12/2018
- Performed Web App penetration testing using OWASP ZAP and standards; identified and documented security risks
- Developed admin functionalities using **AngularJS**, **Spring**, and eliminated Broken Access Control vulnerabilities
- Conducted research on and experimented with the implementation of **token-based authentication** using JWT
- Upgraded existing user authentication process to incorporate LDAP and Microsoft Azure AD credentials; improving scalability of new AD user creation and reducing manual new user configuration steps from linear to constant

PROJECTS

NEURAL DEPENDENCY PARSER
- Implemented part of an academic paper (Manning et al., A Transition-Based Dependency Parser, 2014)
- A transition-based dependency parser using the shift-reduce framework, a neural network, and Word2Vec embeddings
- Created and trained the ML model with Keras on the Penn TreeBank dataset

EDUCATION

UNIVERSITY OF WATERLOO · Bachelor of Mathematics - Honours, Co-op
09/2017 - 04/2023
- **Term Distinction:** Spring 2020 - Present
- **Extra-curriculars:**
  - Volunteered on Faculty Orientation Committee of 4 and managed all logistics of Math OWeek 2021 for 1900 incoming students and led 70 upper-year Orientation Leaders
  - Planned OWeek events and mentored incoming students as an Orientation Leader in 2020/2019
  - Studied abroad in Italy and France during Summer 2018