chongwengliang@gmail.com
peter-chong.github.io

PETER CHONG

Skills Summary

- *Languages*: Python, SQL, R, C++, TypeScript, HTML/CSS, MATLAB, Swift
- *Technologies*: AWS, GraphQL, PySpark, Node.js, Teradata, SQLite, Scikit-learn, Git, Bash, Jest

Work Experience

Amazon Web Services

Software Developer Engineer Intern — May 2022 – Present

- Built an internal auditor tool that detects data discrepancies between various databases such as Amazon DynamoDB & AWS OpenSearch
- Devised the infrastructure of the tool by utilizing Amazon API Gateway & AWS Lambda
- Wrote various TypeScript scripts to perform CRUD operations on databases

Intact Financial Corporation

Data Science Co-op — January – April 2022

- Improved pre-trained BERT model's accuracy by 16% and increased annotation speed by 2.7x using Snorkel weak supervision technique
- Designed and evaluated statistical tests to make informed recommendations using non-parametric bootstrap and Wilcoxon rank-sum test
- Held leadership roles such as daily scrums facilitator, sprint planning and sprint review host

XE.com

Data Science Associate — May – August 2021

- Performed entity resolution and fuzzy string matching to de-duplicate customer records
- Reduced fuzzy search runtime by 82% using NLP methods such as N-gram and TF-IDF
- Implemented and designed an end-to-end object-oriented microservice to validate customers' addresses by leveraging Google Geocode API
- Constructed SQL queries to retrieve data from AWS S3 and project them in QuickSight dashboards

Loblaw Companies Limited

Data Science Analyst — January – April 2021

- Saved 4 hours of weekly manual work by automating standard report ETL across different teams
- Wrote Python scripts in Databricks that discover fraudulent email addresses using regular expressions
- Conducted statistical analysis on store transaction history using Apache Spark on Databricks to better detect anomalies and discover data patterns

Marlena Books

iOS Developer — September – December 2019

- Developed and integrated Apple Push Notification using Google's Firebase Cloud Messaging service for backend support

Projects

- Developed a customer segmentation model by clustering customers into 3 different priority groups using K-means Clustering based on RFM score [Link]
- Built a content-based podcast recommendation engine that recommends most similar podcasts to user [Link]
- Explained and compared the strengths and weaknesses of different dimension reduction techniques [Link]
- Created a marathon finishing time predictor that helps runners obtain their estimated finishing time [Link]

Education

University of Waterloo

Bachelor of Mathematics: Statistics and Computational Math — December 2022 (Expected)

- *Relevant coursework*: Mathematical Statistics, Regression, Classification, Forecasting, Data Visualization, Experimental Design, GLMs, Data Structure & Algorithm, Databases, OOP