(651) 795-8093
kzhang.matt@gmail.com
4715 25th Ave, Seattle, WA, 98105

Kaichong (Matt) Zhang

Portfolio:
github.com/mattkczhang
linkedin.com/in/mattkczhang

Education

Master of Science in Applied Data Science, *University of Chicago* **September 2021 - December 2022**
Coursework: Data Engineering, Data Visualization, Data Mining, Advanced ML & AI, Big Data Platform GPA: 3.89
Bachelor of Arts in Economics, Political Science; Minor in Statistics, *Macalester College* **September 2017 – May 2021**
Honors: Phi Beta Kappa (Top 10%), Economics Honors Program, 3M Scholar Awards GPA: 3.89

Technical Experience

Data Analyst Intern **April 2023 — August 2023**
MacHarry Education Consulting *Plymouth, MN*

- Developed an OLTP through data modeling and normalizing data in third normal form to improve DML performance and data integrity. Designed and implemented ETL, denormalized data, and loaded it into a star schema to optimize the DQL performance. Created index and conducted index tuning to speed up data fetching. Created views to simplify code complexity and improve data security.
- Built the MySQL Server-Tableau connection and designed 11 visualizations and 2 interactive dashboards monitoring customer purchasing behaviors and product popularity, helping the marketing team to identify and predict the trends and the seasonality.
- Conducted EDA to uncover the relationships among customers' demographic information, products' characteristics, and transaction status, cleaned the data by converting data types and handling outliers and missing values, and developed a Deep Factorization Machine model based on implicit feedback, achieving 95.5% F1 Score, 16.9% recall@10 and aiming to increase the revenue by 30%.
- Developed an automatic data flow from data loading from OLAP to data cleaning and transformation and to modeling via Python, deployed the final model on a web app demo via Streamlit where employees can log in, check the customers' profile and recommended products by entering their ID, and rerun the model on the up-to-date data, aiming to reduce the data analysis workload for 25%.

Data Scientist **March 2022 — December 2022**
Abzooba Inc. *Milpitas, CA*

- Developed an 8 times faster image segmentation tool for cellular research by building deep learning U-Net and FCN-8 models from scratch on cloud via Pytorch and deployed the best-performing U-Net on Android with an inference time 1 second per image.
- Achieved cross-validated 99.8% accuracy and 99.5% IoU on the final model by increasing the number of epochs, setting up early stop, and imposing image augmentation to overcome the overfitting problem and improve model generalization.
- Collaborated with cross-functional teams to build the end-to-end project pipeline from data extraction to model deployment on Abzooba's AI/ML operation platform Xpresso.ai, leading to a 30% increase in productivity.

Data Science Intern **July 2022 — October 2022**
Inference Analytics Inc. *Chicago, IL*

- Developed an AI-based text-mining system that extracts the keys of radiology reports and outputs structural data for UChicago Medicine.
- Tokenized the text data, analyzed the part-of-speech, dependency, and entity annotation of sampling sentences, identified 18 sentence patterns for keywords extraction using SpaCy and Stanza, and successfully influenced stakeholders to integrate the new patterns into the NLP module, leading to a 25% increase in accuracy.

Economics Data Analyst **June 2020 – May 2021**
Macalester College Economics Department *St. Paul, MN*

- Co-authored 2 papers with Professor Felix Friedt empirically investigating the mechanism of and trade effects of Covid-19 and published in Covid Economics (Issue 53) and Macalester Digital Commons, gaining 1200+ downloads and citations.
- Collected 7.4 GB data, transformed and merged it to panel data using Stata, invented 26+ fixed-effect regressions with hypothesis testing and robustness check to demonstrate the significance of the Covid shock, and presented the results to the department.

Technical Projects

Flight Cancellation Prediction and Features Analysis **September 2022 – December 2022**
University of Chicago Big Data Platform *Chicago, IL*

- Extracted, transformed, aggregated, and merged four datasets in total 116 GB from GCS data lake, dropped and imputed missing values, and loaded the data for EDA and feature engineering using PySpark SQL on GCP dataproc cluster.
- Developed PySpark data preprocessing and random forest and gradient boosted tree machine learning pipeline to predict flight cancelation and optimized the models using grid search, achieving a 30% increase in accuracy and F1 score.

Skills

Data Science Skills	Python, SQL, Tableau, Spark, Hive, GCP, Hadoop, R-Studio, MongoDB, MS Excel, Stata, D3, Streamlit
Developer Skills	JavaScript, CSS, HTML, Balsamiq Wireframes, Node, Express, Mongoose, Linux
Languages	English, Mandarin, Cantonese