

YIFAN LI

267-781-5871 | liyifan.steven616@gmail.com

EDUCATION

University of Pennsylvania

Master of Science in Engineering (MSE) in Data Science

Sep 2021 - May 2023

Philadelphia, PA

- GPA: 4.0/4.0
- A+ courses: Algorithms & Computation, Machine Learning, Deep Learning, Artificial Intelligence, Computational Linguistics

Tsinghua University

Bachelor of Management in Information Management and Information Systems, Major

Sep 2017 - Jun 2021

Beijing, China

Bachelor of Science in Data Science and Technology, Minor

- GPA: 3.84/4.0
- Awards: 2019 National Scholarship (Top 0.01%), 2018 & 2020 Academic Excellence Scholarship

PROFESSIONAL EXPERIENCE

Amazon

Software Development Engineer Intern

May 2022 - Aug 2022

Seattle, WA

- Designed and implemented a state machine with AWS Step Function and Lambda, which orchestrates the backend data ETL pipeline of the Physical Store Analytics team, loading data from AWS S3 to Redshift and conducting further transformations
- Compared to the old system, the new architecture allows automatic retries on temporary errors, enables the team to launch manual re-drives easily, and provides more detailed monitoring and customized alerting for the team to manage the data pipeline

Microsoft

Research Intern, Microsoft Research

Aug 2020 - Mar 2021

Beijing, China

- Published research paper, *HALO: Hierarchy-aware Fault Localization for Cloud Systems*, on SIGKDD 2021 as the third author
- Designed and implemented “HALO”, a large-scale failure diagnosis algorithm for cloud services, with other team members. Reproduced two related works and conducted comparative experiments of accuracy and runtime efficiency with Python
- Developed Flask API services and Kusto plugins that allow other internal teams to use our novel algorithm. According to feedback from members of Office 365 and Azure, our algorithm significantly reduced engineers’ anomaly mitigation time
- Awarded the “Stars of Tomorrow” award of excellence at the end of the internship

ByteDance

Big Data R&D Intern / Data Science Intern

Jan 2020 - Aug 2020

Beijing, China

- Applied Hive SQL and Spark in corporate advertisement data warehouse construction and governance
- Maintained the schema and data pipeline of two data-mart-level tables related to “Dou+”, the advertising product of TikTok. Implemented feature requests such as adding new metrics related to coupons, shopping carts, and live stream promotions
- Optimized the A/B tests anomaly detection mechanism on the corporate online experiments platform. Conducted research and implemented a new algorithm with time series data mining techniques that increased the alerting accuracy by 30%

RESEARCH EXPERIENCE

Quantitative Marketing Research Assistant

The Wharton School, University of Pennsylvania

Sep 2021 - Present

Philadelphia, PA

- Built a preference measurement web application for data-driven marketing research. Debugged, refactored, and optimized the legacy Flask and React code extensively, leading to significant improvement of code robustness and readability
- Assisted the maintenance of data ETL pipeline of several other projects using Python, MongoDB, and AWS S3

LEADERSHIP AND EXTRACURRICULAR ACTIVITY

Ernst & Young NextWave Data Science Challenge

Team Leader

Apr 2019 - May 2019

Shanghai, China

- Led a team to win the China Regional Championship
- Contributed feature engineering ideas, did data wrangling, model selection, and hyper-parameter optimization with pandas and scikit-learn. Built a machine learning model that accurately predicts residents’ travel path
- Structured the winning presentation, which showed the exploring process, analysis, and potential applications

CNKI-2-BibTeX

Open Source Developer and Maintainer

Feb 2019 - Present

- Built an open-source library, CNKI-2-BibTeX, for using BibTeX with a Chinese academic database CNKI
- The project helped many other Chinese scholars conduct researches, and got more than 130 stars and 72k downloads

ADDITIONAL INFORMATION

- Skills: Python, SQL, Hive, Scala, Spark, TypeScript, Node.js, Data Mining, Machine Learning, AWS