YIFAN LI

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

EDUCATION

University of Pennsylvania   Sep 2021 - May 2023
Master of Science in Engineering (MSE) in Data Science   Philadelphia, PA
- GPA: 4.0/4.0
- A+ courses: Algorithms & Computation, Machine Learning, Deep Learning, Artificial Intelligence, Computational Linguistics

Tsinghua University   Sep 2017 - Jun 2021
Bachelor of Management in Information Management and Information Systems, Major   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

University of Minnesota   Aug 2019 - Dec 2019
Exchange Student at Carlson School of Management   Minneapolis, MN
- GPA: 3.898/4.0
- Selected Courses: Big Data Engineering and Architecture (Graduate Level, won the first place in the coursework big data analytics project), Interactive Data Visualization for Business Analytics, Analysis and Modeling of Business Systems

PROFESSIONAL EXPERIENCE

Amazon   May 2022 - Aug 2022
Software Development Engineer Intern   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   Aug 2020 - Mar 2021
Research Intern, Microsoft Research   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   Jan 2020 - Aug 2020
Big Data R&D Intern / Data Science Intern   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%

Continental Automotive Systems (Shanghai) Co., Ltd.   Jul 2019 - Aug 2019
Intern, Industrial Engineering Department   Shanghai, China
- Collected data from Oracle database, and utilized MicroStrategy business intelligence tools and KNIME data analysis platform to conduct analysis on the data gathered
- Analyzed and predicted such key production indicators as primary pass rate and logistic shipment volume, and plotted interactive visualization charts for further review
- Took deep involvement in the PRA (Robotic Process Automation) project, studied and applied UiPath RPA to develop three projects
- Hosted three internal workshops and finished a 110-page tutorial documentation to promote the implementation and application of RPA project in the whole company

Summitview Capital Co., Ltd.   Jul 2018 - Aug 2018
Intern, Investment Department   Shanghai, China
- Independently accomplished the 38-page research report of autonomous car sensor industry, which well prepared for the whole team to investigate the investment target of the on-board MMW (Millimeter Waves) Radar Chip, and gave a presentation to help our team know more about the industry vision
- Participated in two interviews of two CEOs about their enterprises according to SWOT and Michael Porter's Five Forces Model, responsible for preparing questions, recording contents, and arranged them into an article

- Translated several Chinese reports into English version, designed two PPT documents on investment research projects, and took minutes of two conferences

RESEARCH EXPERIENCE

Quantitative Marketing Research Assistant   Sep 2021 - Present
The Wharton School, University of Pennsylvania   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

Carlson Leader Time Use Data Analysis   Sep 2019 - Dec 2019
Research Assistant to Prof. Le Zhou of University of Minnesota   Minneapolis, MN
- Wrote Python scripts to extract time use data from 500 Excel spreadsheets into one CSV. Conducted the data cleansing and help with preliminary analysis like reporting descriptive statistics, correlation and regression
- Built a website with Vue.js, Flask and MongoDB to replaced the old Excel data collection tool, providing both data collection and participant feedback functionality. The veracity of data and the user-friendliness are both improved.

Research on Price Index Construction and Monetary Mechanism in Fictitious Economy   Oct 2018 - Jun 2019
Research Assistant to Prof. Wenlan Luo and Prof. Qing Liu of Tsinghua University   Beijing, China
- Cooperated with Tencent Internet Innovation Technology Joint Lab to preprocess and mine the data of inter-player transactions and player-company transactions and eventually established a related econometrical model
- Analyzed the reasons for the serious inflation of game currency and severe loss of players from FIFA Online 3, and proposed suggestions for the transaction, currency, and reward mechanisms for FIFA Online 4

LEADERSHIP AND EXTRACURRICULAR ACTIVITY

Ernst & Young NextWave Data Science Challenge   Apr 2019 - May 2019
Team Leader   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   Feb 2019 - Present
Open Source Developer and Maintainer
- 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

Chinese Young Development Program   Jan 2019 - Feb 2019
Participant, Columbia University   New York City, NY
- Engaged in various lectures in Behavior Economics, Financial Risk Management, Value Investing, Design Thinking, Marketing, International Organizations taught by Columbia University faculty and industry practitioners
- Visited New York organizations, including IBM and Morgan Stanley
- Participated in a design and business plan competition in a team of 9; played the role of team leader and won second prize among seven teams; took charge of organizing teamwork, formatting the report using LaTeX and construct the advertising website

Taekwondo Delegation of Tsinghua University   May 2018 - Jun 2019
Team Captain   Beijing, China
- Planned and led daily training, member recruitment, and organized shows in the School Anniversary Event
- Ranked 5th, individually, in the Beijing University Taekwondo Competition, and led the team to win 2nd Place Prize

Academic Development and Scientific Innovation Association   May 2018 - Jun 2019
Principal of Match Department, Tsinghua U. SEM   Beijing, China
- Took charge of planning the highest-level contest on economic hot spots analysis at Tsinghua - Jin Jing Le Dao, attracted over 200 groups from nationwide universities in the main stadium, and organized the freshman competition whose scale was double that of last year
- Completed the most important section of writing the competition manual, and used Python script program to process the competition reports of over 200 groups

ADDITIONAL INFORMATION

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