Zach Zhu

(773)739-4294 | zachzhu@u.northwestern.edu | https://www.linkedin.com/in/zazhu/ | https://zekaizhu.github.io/

EDUCATION

Northwestern University, Evanston, IL, United States — *September 2019 - December 2020*
Master of Science in Analytics — GPA: 3.9/4.0

Soochow University, Taipei, Taiwan — *September 2015 - June 2019*
B.B.A in Financial Engineering and Actuarial Mathematics — GPA: 4.0/4.0

TECHNICAL SKILLS

- **Programming languages:** Python, SQL, R, Scala, Shell scripting, D3.js, HTML
- **Framework & tools**: Git, Bash, Hive, Presto, Excel, AWS, Airflow, Tableau, Spark, Azure, Docker, Jira, Confluence, Looker

PROFESSIONAL EXPERIENCE

Pinterest — San Francisco, California
Data Scientist, Browse & Search — *February 2021 – Present*

- Owned multiple search product analytics workstreams: search landing page, search recommendations, people search, and search refinements; collaborated with product managers and software engineers to design success and guardrail metrics
- Conducted search use case expansion analysis using causal inference (propensity matching, IPW, and Doubly robust estimation) to draw causal relationship between user interest expansion and long-term retention on Pinterest; presented to the head or product and head of product marketing and helped shape search discovery product roadmap
- Performed search landing page user segmentation analysis and dynamic ranking analysis to understand how different user segments engage with different content types (shopping, curated, and personalized) and how engagement differ among slots; helped achieve team key OKRs and provided directional guidance on the whole page ranking strategy
- Worked with data scientists to design search intent sessions and search fulfillment metric by leveraging text analytics techniques; search fulfillment rate was adopted by the search teams as north star metrics to measure experiment success
- Created multiple automated experiment reports in Python with significance tests to better interpret results of A/B experiments and inform launch review decisions; reports have been adopted by three search teams (40+) and included in 50+ experiments
- Designed workflow and data lineage to generate downstream tables using Airflow; designed dashboard to report key metrics and conduct metrics drop investigation; performed frontend logging audit and PR review to ensure data quality
- Organized Data Education Program (>500 participation) to help employees learn how to work with data and provided mentorship to a 15-week project: help center tickets classification and sentiment analysis using NLP

LinkedIn — Sunnyvale, California
Data Scientist Intern, Growth — *June 2020 – September 2020*

- Used Presto and Spark to analyze massive amounts of data of app activation, notifications, and sessions, etc.; migrated metric datasets by running flows on Azkaban to cope with changing data schemas and designed metrics reporting dashboards
- Employed machine learning classification model (Random Forest, XGBoost, Logistic Regression) to understand the relationship between 110 million infrequent members' email open and member life cycle upshift; proposed most valuable emails to be prioritized for the next two quarters of email edge building strategy
- Conducted deep dive analysis of onboarding flow to improve email confirmation (2% lift in weekly active users); identified emails with deep link issues that mostly impact members' mobile experience and persuaded engineers to fix the problem
- Performed funnel analysis of phone address book imports to find massive opportunities to optimize recommendation for connection and help infrequent members connect with frequent members; worked with product and engineering teams to drive corresponding experiments

NASA Jet Propulsion Laboratory (Learn More) — Pasadena, California
Data Science Researcher — *October 2019 - June 2020*

- Conducted exploratory data analysis and 3D visualization in Python on massive electron number density data collected from THEMIS probes; detected repetitive cycles and anomalies by applying data mining, machine learning and clustering techniques (t-SNE, PCA, K-means) to help increase accuracy for future disturbance prediction in the physical system
- Built an interactive web dashboard to visualize the clusters and anomaly detections by leveraging D3.js and html (Learn More)

TransUnion — Chicago, Illinois
Analytics Consultant - Data Science & Analytics team — *September 2019 - June 2020*

- Employed XGBoost to develop credit risk models using 1500 attributes of credit risk reports to measure potential lift in the prediction of account delinquency; elevated the AUC score of the benchmark model by 2% with the addition of macroeconomic predictors
- Leveraged multivariate time series model VAR and LSTM to predict macroeconomic trends using key consumer credit attributes; employed forward chaining to validate the models and achieved less than 0.5 RMSE in the prediction of fraud-related macroeconomic indices