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: 3.88/4.0

- Relevant Coursework: Analytics Value Chain, Big Data Analytics (Hadoop & Spark), Databases & Information Retrieval, Data Mining, Data Management for Business Intelligence, Deep Learning, Data Visualization, Predictive Analytics, Text Analytics

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

- Dean's List for all semesters attended and graduated with honor
- **Society of Actuaries (SOA):** Exam**:** Financial Mathematics, Probability; VEE: Corporate Finance, Applied Statistics

TECHNICAL SKILLS

- **Programming languages:** Python (Pandas, NumPy, scikit-learn, Keras), SQL, R, Scala, Shell scripting, D3.js, JavaScript, HTML
- **Framework & tools**: Git, Bash, Hive, Presto, Excel, AWS (S3, EC2, RDS), Tableau, Spark (PySpark, SparkML), Hadoop, MapReduce, SPSS, Azure, Docker, Jira, Confluence, Looker, A/B Testing, Airflow, Predictive Analytics, Funnel Analysis, Agile Project Management

PROFESSIONAL EXPERIENCE

Pinterest — San Francisco, California
Product Analyst – Inspire & Search — *February 2021 – Present*

- Perform deep dive analysis using SQL and Python to understand and optimize the key levers of the search and shopping products; search query clustering analysis to detect user intent of shopping and people search; user segmentation and module analysis to understand shopping and curated content impact to different user segment; construct regression models to understand relationship between users engagement with search products and long term retention; Help achieve team key OKRs by the end of Q3 and proposed new target goal
- Design success and guardrail metrics that serve as the North Stars for the team efforts and model trade-off decisions across product areas
- Create compelling experiment jupyter reports to better interpret results of A/B experiments and inform launch review decisions
- Design workflow and data lineage to generate downstream tables which can be used for reporting metrics on dashboard and data analysis
- Perform frontend logging audit, metrics drop investigation, and user case study to ensure data quality and identify product opportunities
- Work with product managers and engineers to design data products, prove their value by running experiments and release into production

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

- Leveraged Presto, Hive, Python, and Spark to analyze massive amounts of data of app activation, notifications, sessions, and members, etc.
- Employed machine learning classification model to understand the relationship between 110 million infrequent members' email click 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 that will largely lift long-term 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
- Migrated metric datasets by running flows on Azkaban to cope with changing data schemas and designed metrics reporting dashboards

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

- Conducted exploratory data analysis and 3D visualization using Python (Plotly, Seaborn) on massive electron number density data collected from THEMIS probes; detected repetitive cycles and anomalies by applying data mining, machine learning and clustering techniques (Anomaly detection algorithms, 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
Student Analytics Consultant — *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
- Performed mapping visualization of quarterly data in R (ggplot): number of new accounts opened and bad rate for each county and MSA
- 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

PROJECT EXPERIENCE

House Sale Price Prediction Application (Learn More) — *April 2020 – June 2020*

- Leveraged AWS (S3 and RDS), docker and Python to build an automated regression model pipeline to predict New York house sale price
- Implemented offline model pipeline in an interactive web-based application where users get forecasted sale price after providing information

Deep Learning Face Mask Detection (Learn More) — *April 2020 – June 2020*

- Implemented VGG16 Neural Network to detect human faces on public images (WIDER Face, MAFA) and classified images into without wearing masks, wearing normal masks, and wearing N-95; achieved 82% validation accuracy

Society of Actuaries Student Research Case Study Challenge — *April 2018*

- Led a team of four to analyze household historical data with R and constructed reduction factors model and Markov multiple-state model to predict the future balance of a public long-term care insurance program; conducted sustainability assessment through sensitivity analysis
- Optimized the public long-term care program to yield a 138% increase of its balance in 2028 by implementing solutions such as introducing progressive tax rate with surcharges for childless families, allocating budget for healthy habits advocacy, etc.