XIYANG "TIM" MO

|| Tucson, AZ || (612) 385-5952 || timxymo@gmail.com || github.com/timxymo ||

EDUCATION

University of Arizona, Tucson, AZ — August 2021 – Present
PhD of Statistics || Minor in Information Science || GPA 3.86/4.0

University of Virginia, Charlottesville, VA — December 2018
Master of Science in Statistics || GPA 3.34/4.0

University of Minnesota-Twin Cities, Minneapolis, MN — December 2016
Bachelor of Science in Mathematics || Minor in Statistics || GPA 3.2/4.0

SKILLS

- **Programming Tools**: **Python** | **R** (dplyr) | SQL | SAS (Certified Advanced Programmer) | Linux |
- **DS Specialization**: **LLM** (IR, Transformers, BERT, Hugging Face, LangChain) | **Machine Learning** (classification, clustering) |
- **Intelligence Tools**: **AWS (Certified Cloud Practitioner)** | Gith | Azure | Databricks | Salesforce | Shiny | Tableau | Power BI | Spark |

EXPERIENCE

Liberty Mutual Insurance, Boston, Massachusetts
Applied Scientist Intern, Solaria Lab — June 2024 – August 2024

- Trained and implemented a SPLADE model, a novel embedding and information retrieval method, for predicting business class codes, **improved MRR@10 by 5%** and recall@10 by 7% on current production model
- Developed parsing scripts to transform data into SPLADE-compatible formats, and architected the migration of training processes to EC2 instances, leveraging GPU capabilities to handle large-scale datasets efficiently
- Explored the integration of Retrieval-Augmented Generation (RAG) with the a off-the-shelve SPLADE model to enhance the output

Data Scientist Intern — June 2023 – August 2023

- Evaluated an internally developed matrix completion package, implemented enhancements, and introduced new functionalities including imputing categorical variables, optimizing hyperparameters and computing confidence intervals for each imputation
- Optimized and implemented the package in real-time applications, resulting in more than 400 hours saved for processing 1.5M records
- Built, trained, and scored new models to compare imputed data with original data, showing improvements across various metrics

Capital One Financial Corporation, Richmond, Virginia
Data Analyst — February 2020 – August 2021

- Utilized **AWS services tools** (S3, EC2, lambda, etc.) to design easy, efficient, and encrypted ways to build and streamline email campaigns, generating 10M+ monthly marketing emails and expanding new lines of business every two months
- Built Fractal campaigns and developed **automation scripts** to generate SQL queries which saved 4 hours per week on average
- Supported card acquisition teams and served as a contact with half dozen line of business partners to accommodate ad hoc business needs and guarantee successful campaign deliveries, resulting a satisfying 80% increment in capacity since joined

University of Virginia, Charlottesville, Virginia
Data Analyst — February 2019 – May 2019, September 2019– February 2020

- Built a **recommender system** based on collaborative filtering and cosine similarity, helping faculties and research admins to discover comparable researchers, match faculties with potential funding opportunities, and vice versa
- Developed an internal research intelligence database on local host of MS SQL server which stored data in a securely, systematic, and cost-efficient fashion and saved hours for researchers and admins from recreating and accessing the related research information
- Designed and deployed an interactive interface with Shiny to visualize the collaboration network between principal investigators within UVA; widely used by research admins across the university to identify research collaboration opportunities among schools and departments (https://timxymo.shinyapps.io/Collaboration_Visualization/)
- Utilized various research intelligence tools to answer specific questions and to support senior management making strategic decisions with data to further develop organizational and strategic improvement efforts for UVA research

Crutchfield Corporation, Charlottesville, Virginia
Data Scientist Intern — June 2019 – August 2019

- Extracted, transformed data in SSMS; cleaned, imputed, and synthesized data in R and applied **classification algorithms** (rf, xgboost, LDA, logistic, SVM), then used the results to build an ensemble model and further improved the accuracy rate to 99.5%-and-above
- Stored data in Azure blob storage and set up Azure HDInsight to move local computations to the cloud; used Spark (PySpark) for parallel computing and got up to 50 times faster performance compared to R CRAN
- Serialized ensemble model as a R object, stored in the server and ran as SQL objects to save hours of training time; created an internal webpage for order processing agents to check fraud probability which minimized manual work and potentially saved ¢40 per order