Shakil Rafi, Ph.D.

479-301-1866 | sarafi@uark.edu | linkedin.com/in/shakilrafi | github.com/2shakilrafi | 2shakilrafi.github.io

Education

University of Arkansas — Fayetteville, AR
Ph.D. in Applied Mathematics, cGPA: 3.65 — *Jan. 2020 – May 2024*

Experience

Post-doctoral Fellow — Aug. 2024 – Present
University of Arkansas Division of Agriculture Research Extension — *Little Rock, AR*

- Wrangle terabyte scale genomic datasets, including denoising and demultiplexing.
- Use advanced machine learning tools like decision tree regression, heirarchical clustering, and deep learning to gain predictive insights from genomic data
- Collaborate and serve as external consultant to various parties across academia and industry. Some of the clients include: Cobb Genetics, Inc., UAB Heersink School of Medicine, and USDA Food Safety testing team.

Lecturer — Jan. 2023 – May 2024
Department of Data Science, Sam M. Walton College of Business, University of Arkansas — *Fayetteville, AR*

- Taught R, Python, Git, and bash to 100+ incoming first-year students. 80% received an A/B.
- Mentored 30+ students in an advanced course on optimization methods using Pyomo and Gurobi. 50% of students showed an improved grade compared to previous year.

Health Economics Intern — May 2023 – Aug. 2023
Arkansas Blue Cross and Blue Shield — *Springdale, AR*

- ETL'd data from Snowflake using SQL
- Developed LSTM models using Keras and TensorFlow for patient claims volumes from a dataset of 40000+ claims, going back to 2018. Model showed 85% accuracy.
- Explored the relationship between the social determinants of health and birth outcomes for our clients using sgradient boosted decision trees, XGBoost, LASSO and ridge regression models with 90% accuracy.
- Built a Markov model of childbirth using Pandas and scikit-learn.

Senior Graduate Assistant — Jan. 2020 – Dec. 2022
Department of Mathematics, University of Arkansas — *Fayetteville, AR*

- Taught pre-calculus, Calculus, Linear Algebra, Differential Equations and intro Statistics.
- Planned lessons and liaison-ed with the course coordinator in a team of five other instructors.

Projects

Gut feelings (tentative) | *Python, R, CUDA, qiime2, Bowtie, plink2, TensorFlow, HPC* — Aug 2024 – Present

- Currently working on using variational autoencoders to perform dimensionality reduction of gut bacterial genomic data to better understand biomarkers for psychiatric disorders
- Working on a project to see the effective ness of Transformer models to predict diseases gased on gut bacterial gene markers.

Who rides Uber anyway? | *Python, Pandas, GeoPandas, scikit-learn, Plotly, SQL* — Jan 2022 – Oct 2023

- Developed a K-means clustering alrgorithm to analyze how richer census tracts in Chicago used Uber compared to poorer census tracts.
- Created an analytical pipeline for large-scale (30 million rows+) transportation data from the City of Chicago data portal.

Pandemic loan inequality | *R, ggplot2, tidyverse, Shiny* — May 2022 – Aug. 2022

- Developed a regression model to see whether black and women owned business in Arkansas received lower initial federal loans.
- Work presented at the *Arkansas Summer Research Institute*, 2022.

Technical Skills

Languages: Python, C/C++, SQL, Julia, R, MATLAB, LaTeX
Tools:Microsoft Office, Bash, Git, Azure, pySpark, Tableau, XGBoost, LangChain, LLMs, MLOps
Libraries: pandas, polars, NumPy, Matplotlib, Keras, TensorFlow, tidyverse, GeoPandas, Pyomo, scikit-learn
Interests: Reading, hiking, and cooking.