Shaan Pakala

San Jose, CA; Riverside, CA | shaan.pakala@gmail.com | Website | LinkedIn | GitHub

Education

Data Science & Engineering, B.S. – University of California, Riverside Sept 2021 – June 2025

- Upper Division GPA: 3.9/4.0

Professional Experience

Data Science Research Assistant – UC Riverside Sept. 2024 – Present

- Extending work with Professor Papalexakis on tensor completion for automating data science pipelines
- Developing new tensor completion algorithms for these applications
- Experimenting with sampling techniques to improve efficiency

NSF REU Data Science Research Intern – UC Riverside [GitHub] June 2024 – Sept. 2024

- First-author paper accepted to IEEE International Conference on Big Data 2024
- Under the supervision of Professors Vassilis Tsotras, Vagelis Papalexakis, and Jia Chen
- Leveraged tensor completion algorithms to accelerate hyperparameter tuning by inferring the results of all predefined hyperparameter combinations (5x – 25x faster than gridsearch)
- Includes deep learning architecture tuning in addition to non-deep learning models
- Additionally, applied this to estimate SQL query cardinalities for the purpose of query optimization

Bioinformatics Research Assistant – UC Riverside March 2024 – Present

- Bioinformatics Research Assistant under Professor Stefano Lonardi
- Using machine learning to automate protein sequence analysis
- Feature engineering protein sequences to improve performance
- Refined noisy dataset to enhance machine learning classification

Data Science Fellow – Lawrence Livermore National Laboratory [GitHub] July 2023

- Designed Neural Network for automated heart disease diagnosis using ECG data with PyTorch
  - Improved minority classes' accuracy 10% with custom loss function
- Mapped ECG data from 12 leads over 500ms into heart activity of 75 locations over 500ms using a CNN

Publication

Shaan Pakala, B. Graw, D. Ahn, T. Dinh, M. T. Mahin, V. Tsotras, J. Chen, and E. E. Papalexakis, "Automating Data Science Pipelines with Tensor Completion," *IEEE International Conference on Big Data, 2024*.

Projects (full list)

Simulated Online Store Database – Database Management Systems Final Project [GitHub]

- Developed PostgreSQL relational database for online store sales data, using Java for user interface
- Implemented triggers and indexing for improved efficiency and reliability

Wikipedia Article Visualization [GitHub]

- Used Google's BERT to generate hidden states, and applied matrix operations and t-SNE for 2D visualization
- Scatter plot revealed close distance for Wikipedia articles with similar topics, making classification trivial

Skills

Coding Languages: Python, SQL, R, C++, Java

Technologies: Tensorly, PyTorch, TensorFlow, Scikit-Learn, Tableau