Shaan Pakala

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

Education

University of California, Riverside, B.S. in Data Science & Engineering — 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
- Includes developing potentially novel tensor completion algorithms for these applications

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 guidance 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
- Applied Machine Learning algorithms to automate protein sequence analysis
- Fine-tuned large-language model for protein cleavage site prediction for drug discovery

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
- Reconstructed spatio-temporal map of hearts using a Convolutional Neural Network
- Mapped ECG data from 12 leads over 500ms into heart activity of 75 locations over 500ms

Paper

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

Simulated Online Store Database – UC Riverside [GitHub]

- Developed PostgreSQL relational database for online store sales data and records
- Implemented triggers and indexing for improved efficiency and reliability

Wikipedia Article Visualization & Classification [GitHub]

- Used Google's BERT on Wikipedia articles to generate hidden states
- Applied series of matrix operations and t-SNE to hidden states for visualization
- Visualization revealed close distance for Wikipedia articles with similar topics

Skills

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

Technologies: Tableau, PyTorch, TensorFlow, tensorly, OpenCV, scikit-learn