SAMIKSHA KALE

✉ samikshakale@gmail.com ◇ ☎ +1 703 726 4328
linkedin.com/samikshakale ◇ github.com/samikshakale
samikshakale.github.io/personal-website/

Education

- **Carnegie Mellon University** Pittsburgh, PA
  BS in Computer Science; minor in Machine Learning *Aug 2019 – May 2023*

Work Experience

- **Data Engineer at Hildreth Institute** *Aug 2023 - Present*
  - Designed and deployed automatic data pipeline using AWS Lambda, S3, RDS, and CloudWatch for CFPB student loan complaint data ingestion, transformation, storage, and integration
  - Constructed highly interactive data dashboard on Tableau connecting to an AWS RDS MySQL database instance for visualizing significant KPIs in addition to geospatial and temporal trends
  - Engineered topic modeling and sentiment analysis solutions to extract key insights from the complaint data
- **Teaching Assistant for Intro to ML at CMU SCS** *Aug 2021 - May 2023*
  - Revamped course website built with Ruby and Jekyll using front-end frameworks including HTML5, CSS, and JavaScript and hosted on Github Pages
  - Supervised and collaborated with my team of 20 TAs in the development and routine testing of course material for 400+ students (homework, exams, recitation) using a Docker-based autograder system with custom compilers
  - Lectured on RNN/CNNs, Transformers, RL, HMMs, Regression, Feature Engineering, Ensemble Methods (Bagging, Boosting), and Recommender systems
- **Machine Learning Engineer Intern at Acuity Diagnostics** *Summer 2022*
  - Utilized an open-source computer vision pipeline, TIAToolbox, to perform instance segmentation on H&E stained whole slide image (WSI) data
  - Extracted tumor-specific features using pretrained HoVer-Net models with custom parameters for use in proprietary model predicting chemotherapy responsiveness
- **Data Science Intern at CMU Software Engineering Institute** *Oct 2021 - May 2022*
  - Built an end-to-end jet engine predictive maintenance tool for an air force base client, reducing costs associated with engine failure
  - Performed NLP text data pre-processing using Python's nltk package and high-dimensional language representation using Google's BERT model on maintenance engineer notes
  - Automated ETL process by integrating relational databases (MySQL) and employed distributed version control with BitBucket Cloud for the CI/CD pipeline
  - Programmed multiple clustering and advanced probabilistic models to facilitate further data mining
- **Teaching Assistant for Functional Programming at CMU SCS** *Jan 2021 - May 2021*
  - Led weekly recitations of 20+ students as well as office hours and weekend review sessions for 300+ students

Programming Skills

- **Languages :** Python, Java, C, R, SML, Javascript, Typescript, SQL, MATLAB, HTML/CSS
- **ML Packages :** Tensorflow, Pytorch, Pandas, SparkML, nltk, Sci-kit Learn
- **Full Stack:** AWS, GCP, Flask, Weights and Biases, MySQL, NodeJS, React, Docker

Project Experience

- **LLM-Powered Autofill Chrome Extension for AI Incidents Database (AIID)**
  - Developed a custom browser extension (using Flask-RESTful in Python, Javascript, HTML, CSS, and GatsbyJS in React) to expedite submissions of AI incidents to the AIID
  - Employed LLMs, including deepset's RoBERTa and Meta's BART, for summarization and question-answering functionality
  - Designed and developed RESTful API using Flask as the backend component
- **Kaggle Amex Fraud Detection**
  - Developed powerful models with >75% accuracy for predicting customer credit default on industrial scale time-series behavioral dataset
  - Benchmarked and evaluated Logistic Regression, SVM, XGBoost, LGBM, CatBoost, and Neural Models on the aggregated features of customer's monthly profiles using cross validation for hyperparamter optimization
  - Conducted numerous data compression techniques to optimize results for training datasets larger than 16GB
- **Speech-to-Text Model for WSJ Article Recordings**
  - Built an attention-based end-to-end sequence conversion deep learning model using encoder-decoder transformer architecture
  - Performed ablation studies on Weights and Biases (MLOps platform) using GCP for hyperparameter tuning and optimization on a 17+ million parameter model
- **Othello AI**
  - Implemented minimax algorithm with alpha beta pruning and iterative deepening that wins against human players
  - Developed unique weighting strategy combining more than 5 validated, evidence-driven heuristics for game play
- **Feature Derived Popularity Prediction of Music Tracks**
  - Utilized AWS EC2 instance with attached EBS volume to perform data conversion on 250 GB million song dataset with S3 for storage
  - Applied distributed computing principles running Spark on AWS EMR for data cleaning, EDA, feature engineering, and model creation

Extracurriculars and Leadership

- **Member of Alpha Chi Omega** *Sep 2019 - May 2023*
  - **Organizational chair**: Supervised 60+ members in designing, planning and building a themed one-story structure over the course of 1 year
  - Other positions: **panhellenic delegate**, **social chair**, and **nominating committee representative**
- **VP of Community Standards on CMU's Panhellenic Council** *Feb 2022 - Dec 2022*
  - Worked closely with university administration and the entire Greek community of 1,100+ members to ensure integrity of each chapter is upheld