SAMIKSHA KALE

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

Education

- **Carnegie Mellon University (School of Computer Science)** Pittsburgh, PA
  BS in Computational Biology; concentration in Machine Learning *Aug 2019 – May 2023*

Programming Skills

- **Languages** : Python, Java, C, R, SML, Javascript, SQL, MATLAB **Cloud** : AWS, GCP, Databricks
- **ML Packages** : Tensorflow, Pytorch, SparkML, nltk **Full Stack** : Flask, MySQL, NodeJS, React, Docker

Work Experience

- **Teaching Assistant for Intro to ML at CMU SCS** *Aug 2021 - May 2023*
  - Revamped course website built in Jekyll using front-end frameworks including HTML5, CSS, and JavaScript
  - 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
- **Machine Learning Engineer Intern at Acuity Diagnostics** *Summer 2022*
  - Utilized an open-source computer vision pipeline, TIAToolbox, to perform instance segmentation on image data
  - Extracted tumor-specific features using pretrained HoVer-Net models to predict chemotherapy responsiveness
- **Data Science Intern at CMU Software Engineering Institute** *Oct 2021 - May 2022*
  - Automated ETL process by integrating relational databases and employed distributed version control with BitBucket Cloud for the CI/CD pipeline
  - Performed text data pre-processing using Python's nltk package and language representation using Google's BERT model
  - Programmed multiple clustering tools to model predictive maintenance and 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

Project Experience

- **LLM-Powered Autofill Chrome Extension for AI Incidents Database (AIID)**
  - Developed a custom autofill extension (using Flask-RESTful in Python, Javascript, GatsbyJS in React) to expedite submissions of AI incidents to the AIID
  - Employed Large Language Models, including deepset's RoBERTa and Meta's BART, for summarization and question-answering functionality
- **Kaggle Amex Fraud Detection**
  - Trained XGBoost to develop powerful predictive models for credit card defaulting with over 75% accuracy
  - 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 architecture
  - Performed ablation studies on Weights and Biases (MLOps platform) using GCP for hyperparameter tuning and optimization on a 17+ million parameter model
- **Feature Derived Popularity Prediction of Music Tracks**
  - Used AWS (EMR, EC2, S3) to apply distributed computing principles in the creation of predictive models
  - Performed data compression, data cleaning, and EDA on the million song dataset (280 GB)

Extracurriculars and Leadership

- **Member of Alpha Chi Omega** *Sep 2019 - May 2023*
  - **Organizational chair**: Supervised 60+ members in building a themed one-story structure over 6 months
  - 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