Roochi Shah

+1 (815)-630-1201 | rtshah@andrew.cmu.edu | linkedin.com/in/roochi-shah-3087a5217/

Education

Carnegie Mellon University   **May 2024**
B.S. Statistics and Machine Learning | Dean's List with High Honors
Minors in Artificial Intelligence, Machine Learning

Graduate Coursework: Advanced Deep Learning, Multimodal Machine Learning, Convex Optimization, ML with Large Datasets, Federated Machine Learning, On-Device Machine Learning, Advanced Probability and Statistics

Teaching Assistantships (Graduate): Large Language Models, Convex Optimization, Statistical Learning
Teaching Assistantships (Undergrad): Machine Learning, Advanced Probability and Stat Inference, Stat Computing

Skills

Languages: Python, R, SQL, Julia, C, Java, HTML/CSS, JavaScript, SM, Scala
Programs: Linux (ubuntu), Git, Spark, Beam, GCP, AWS (EC2/S3), Jupyter Notebooks, Kubernetes, Tableau
Python Libraries: PyTorch, TensorFlow, Sklearn, NumPy, Matplotlib, Seaborn, Pandas, OpenCV, NLTK, Selenium

Work Experience

Machine Learning Engineer R&D | KBR, U.S. Space Force   **Fall 2023**
- Novel machine learning methods to automate star detection and astrometry from high-speed satellite imagery

Machine Learning Engineering Intern | Spotify (Advertising R&D)   **May 2023- Aug 2023**
- Developed first ad relevance heuristic allowing for user generalization through personalization embeddings
- Architected, analyzed, and automated ad audio features with Apache beam
- Formulated framework to incorporate generative background music for advertisers in Spotify's ad studio

Data Science Intern | John Deere (Intelligent Solutions Group)   **May 2022- Aug 2022**
- Architected satellite imagery semantic-segmentation pipeline for John Deere's autonomous tractors
- Performed distributed data preprocessing, and feature engineering on 4 petabyte data set using Spark

AI Instructor | InspiritAI   **May 2022- Aug 2022**
- Presented lectures on artificial intelligence fundamentals (supervised learning, CV, NLP, RL) to 200+ students
- Directed machine learning projects on large language models, object detection, and self-driving cars

Data Science Co-op | Wabtec Corp. (Advanced Technologies Team)   **Oct 2021- Feb 2022**
- Modeled railcar delay from time-series data of railcar movement for Wabtec's 2030 de-carbonization initiative
- Architected railcar movement database with automated updating through web-scraping

Research Projects

SightSense (On-Device ML Course Project) | CMU Dept. of ML   **Present**
- Building ML app to help the visually impaired that summarizes surroundings through real-time object tracking

Generative Models for Adversarial Robustness | CMU Dept. of ML   **Jan 2023- Present**
- Developing and testing novel training procedure for increased adversarial robustness in vision models

BasicML Library | CMU Dept. of Data Science   **Jan 2022- Present**
- Authoring the first AutoML package in R to automate creation of data science reports and visualizations
- Includes: pre-processing, visualization, modeling, hyperparameter tuning, and inference

Speaker-Follower for ALFRED Subgoals | CMU Dept. of ML   **Jan 2023- May 2023**
- Designed architecture combining vision and language transformers for multimodal "butler" agent
- Implemented "subgoal-encoder" which creates an intermediate representation of language instruction input

Auto-Regressive Models for Stock Market Prediction | CMU Dept. of Data Science   **Jan 2023- May 2023**
- Coordinated team of four in data science capstone project predicting stock prices for individual firms
- Designed auto-regressive linear models to predict stock prices from financial data reported in W2 statements