Aditi Tripathi

(+1) 412-708-7569 | adititripathi694@gmail.com | karasuno7.github.io | karasuno7 | tripathi-aditi/

Education

Carnegie Mellon University (CMU)
Pittsburgh, PA

Masters in Electrical and Computer Engineering — *2019 - 2020*

- 11785 - Introduction to Deep Learning • 10601 - Introduction to Machine Learning • 15319 - Cloud Computing • 15640 - Distributed Systems • 18793 - Image and Video Processing • 18889 - Internet Services • 17681 - Data Structures • 18613 - Introduction to Computer Systems

National Institute of Technology (NIT)
Bhopal, India

Bachelors in Electrical Engineering — *2012 - 2016*

Skill Set

- **Languages** : Python, Java, C, C++, SQL, Scala, Matlab • **Containerization :** Kubernetes, Docker
- **Tools** **:** NLTK, Scikit-Learn, Pytorch, Tensorflow, OpenCV, Apache Spark, OpenMP, AWS, GCP, Bash, Git, Vim etc.
- **Skills** **:** Multimodal Machine Learning, Natural Language Processing, Image Processing, Distributed Systems

Work Experience

Research Assistant - advised by Prof. Carolyn Rose, CMU SCS, Pittsburgh, PA
Feb 2021 - Present

- Developing intelligent conversational agents to support collaborative learning using machine learning and probabilistic models.
- Designed a performant event-based lambda architecture with ElastiCache and JSON-RPC API to generate automatic hints for the students' code in real time.

Data Scientist Intern, Boeing, Bellevue, WA
June 2020 - Aug 2020

- Developed a more memory efficient and a 4x faster image processing framework (Using Clustering algorithms for segmentation and using OCR based image stitching instead of overlap detection).
- Achieved a cross validation accuracy of 95% with an XG-Boost clustering model for airplane part price prediction.

Graduate Research Assistant - advised by Prof. Richard Stern, CMU, Pittsburgh, PA
Oct 2019 - Dec 2019

- Developed a java user-interface for a GIS based air traffic noise prediction model for overlaying helicopter trajectory on top of raster maps created using LIDAR data points to display noise levels around Allegheny General Hospital.

Research Consultant, Worldquant LLC, Virtual Research Center, India
Dec 2018 - Jul 2019

- Developed new ML models to seek out sources of inefficiencies, and build predictive profitable strategies.
- Developed new alpha strategies using online machine learning methods for RavenPack market sentiment data.

Engineer, L&T Mitsubishi Hitachi, Chennai, India
Aug 2016 - Nov 2018

- Reduced downtime by 20% by switching from scheduled maintenance to predictive maintenance, using machine learning anomaly detection models.
- Brought down triple-redundant equipment cost by 33%, by using an ensemble of regression models based on Principal Component Analysis, to enable early fault detection and time to failure prediction in sensor based systems.
- Improved PCA by modifying data pre-processing pipeline to remove random fluctuations with a sliding window average and a variable 3σ to include local changes.

Academic Projects

Deep Learning
CMU, Pittsburgh

Tag: PyTorch, AWS, Tensorboard — *Fall 2020*

- **Attention-based End-to-End Speech-to-Text Deep Neural Network**– Implemented the Listen, Attend and Spell paper to design a system for speech-to-text transcription using **Locked Dropout, Teacher Forcing and padded Cross Entropy Loss** to achieve a Levenshtein Distance of 17.7 on Librivox dataset and top 11% on Kaggle Leaderboard. (kaggle)
- **Multimodal Visual Question Answering System**– A question answering system using **BERT embeddings and multistage fusion** for user's photo albums and photo metadata to enable multiple-choice question-answering and providing grounding photos based on the user's question. Achieved 5% improvement for "how many" and "where" type questions over the SOTA implementation for MemexQA. (github)

Social Networking Website
CMU, Pittsburgh

Tag: MySQL, Neo4j, MongoDB, cache — *Fall 2020*

- Built a social network website that displays a user's timeline and top users activity by retrieving data from different back-end systems like MySQL, Neo4j and MongoDB with a light weight, in-memory cache for the frontend.
- **Deployed a video processing pipeline on CloudTube** as a serverless architecture to generate GIFs,captions and thumbnails using FFMPEG and Azure's Cloud Vision API with support for label search and indexing using Azure's cognitive search.
- **Driver Matching and Advertising service** Implemented a solution to join and process multiple input streams of GPS data, IoT device data and user data from Kafka producers to enable an Uber-like driver matching service by deploying a Kafka and Samza stream processing system on a YARN cluster.
- **Research papers to audio-books converter** Using Google Cloud's computer vision OCR to enable ML based filtering of most common fonts in a research paper to generate audio from text-to-speech cloud API.

Other Experience

Course Developer, CMU, Pittsburgh, PA
Jan 2021 - Present

- Teel Lab AI Practitioner Course - Developed course content for Computer Vision, Information Retrieval and Question Answering Systems module.
- Teaching Assistant for 18474 - Developed Reinforcement Learning Car Racing Environment with **DDPG and DQN agents in Matlab with Python bindings** (github).
- Undergraduate Research Assistant, NIT Bhopal - Developed optimal bidding strategy by comparing **Genetic Algorithm and Monte Carlo approach in solving bi-level stochastic optimization problem**. Paper presented in i-Fast Savishkar 2015 conference.