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

Work Experience

Research Assistant — Prof. Carolyn Rose, CMU SCS — Pittsburgh, PA — *Mar 2021 - Present*

- Developing intelligent collaborative learning environment to assist students with real-time hints for programming assignments.
- Removed ambiguous hints and improved overall hint quality by improving the reward function for the Markov Decision Process.
- Designed clustering based recommender system that uses ASTs of previous students' data as a basis for automatic hint generation.
- Deployed a performant event-based lambda architecture with ElastiCache and JSON-RPC API to deliver next-step hints to students in real time.
- Evaluated students' learning experience with A/B testing.

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

- Developed a memory efficient and a 4x faster image processing framework using image-segmentation and OCR based image stitching instead of overlap detection.
- Identified and visualized patterns in airplane part prices with respect to 3D part dimensions, using pandas and python.
- Improved cross-validation accuracy by 8% for the XGBoost model using SHAP value analysis.

Graduate Research Assistant — Prof. Richard Stern, CMU — Pittsburgh, PA — *Oct 2019 - Dec 2019*

- Developed a javascript user-interface with React and Redux for a GIS based air traffic noise prediction model to display noise levels around Allegheny General Hospital.
- Efficiently stored multi-dimensional LIDAR into pointcloud using geohash spatial data compression in PostgreSQL.

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.

C&I Engineer — L&T Mitsubishi Hitachi — Chennai, India — *Aug 2016 - Nov 2018*

- Reduced downtime by 20% by switching from scheduled maintenance to predictive maintenance, using anomaly detection models.
- Enabled early fault detection and time to failure prediction in sensor based systems.
- Developed and deployed custom SQL functions to streamline common investigatory processes.
- Improved quality of the sensor database (100+ GB data) by recommending improvements to the database structure, integrating third-party data, and documenting patterns and causes of missing/duplicate/inaccurate/outlier data.

Academic Projects

Deep Learning — *CMU, Pittsburgh*

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

- **Attention-based End-to-End Speech-to-Text Deep Neural Network**–
- Implemented a speech-to-text transcription system 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**–
- Implemented a neural network to retrieve user-photos based on questions asked by the user - using BERT embeddings and multistage fusion.
- Achieved 5% improvement for "how many" and "where" type questions over the SOTA implementation for MemexQA. (github)

Cloud Computing and Distributed Systems — *CMU, Pittsburgh*

Tag: MapReduce, Scala, Java, AWS, Kafka, Samza — *Fall/Spring 2020*

- **Fault-tolerant 2PC Application** — Implemented 2PC protocol and a fault tolerant server with write-ahead-logging which publishes photo-collage after consensus from globally distributed contributing users.
- **Big Data Analytics with MapReduce** — Processed 120GB dataset of Wikipedia traffic log to analyze current popular events using robust OS independent parallel algorithms running as MapReduce jobs on AWS EMR cluster.
- **Multithreading and Distributed Key-Value Store**— Built a globally distributed in-memory key-value store with strong consistency, using java multithreading API , to store sales records for a company.
- **Pagerank in Scala** — Analyzed the Twitter social graph by implementing the PageRank algorithm in Spark (Scala) using RDDs and DataFrame by performing iterative processing.
- **AWS Automatic Scaling Group** — Analyzed workload pattern and developed elasticity policies for AWS Automatic Scaling Group, to maintain the Quality of Service (QoS) of a web service.
- **Driver Matching and Advertising service** — Implemented a cloud service to join and process multiple input streams of GPS data, IoT device data and static 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.