

Bharani Ujjaini Kempaiah

ukbharani@gmail.com ♦ (412) 805-9987 ♦ bharaniuk.github.io

EDUCATION

Carnegie Mellon University – School of Computer Science

Master of Computational Data Science | QPA: 4.14/4.33

Selected Coursework: Intro to NLP, Multilingual NLP, Intro to Machine Learning, ML with Large Datasets

Intro to Deep Learning, Cloud Computing, Interactive Data Science

Pittsburgh, PA

December 2022

PES University

Bachelor of Technology in Computer Science & Engineering | GPA: 9.83/10

Selected Coursework: Data Structures & Algorithms, Operating Systems, Linear Algebra

Object Oriented Modelling & Design, Databases, Data Analytics, Big Data, Machine Learning

Bangalore, India

May 2021

SKILLS

Programming Languages/ Version Control: Python, Java, C, C++, R, SQL, Scala, Javascript, PHP, HTML, CSS, Bash, Git

Data Science Tools: Numpy, Pandas, PyTorch, Keras, WandB, scikit-learn, NLTK, ESPNet, Fairseq, Huggingface

Cloud/Big Data tools: Docker, Kubernetes, Kafka, Spark, Terraform, MySQL, HBase, Neo4j, DynamoDB, AWS, GCP, Azure

EXPERIENCE

Cognistx

Data Scientist Intern

Pittsburgh, PA

May 2022 - August 2022

- Lead development of production-grade AWS-based (SQS, Lambda, DynamoDB, S3) Web Crawler to enhance domain generalization and performance of downstream QA model, resulted in acquisition of three new customers
- Designed de-coupled architecture of fetcher and downloader for link deduplication and storing crawled data into DynamoDB database
- Reduced crawling time by at least 35% compared to standalone EC2-based deployment

AppDynamics, Cisco Systems (India) Private Limited

Technical Undergraduate Intern

Bangalore, India | Remote

January 2021 - May 2021

- Created library to dynamically alter agent configurations. Written in Java, it uses SnakeYAML to construct custom Java objects and is currently being employed by the OpenTelemetry team to test future builds of agent
- Facilitated automated testing of multiple agent configurations, leading to 38% reduction in testing time

Centre for Cloud Computing and Big Data, PES University

Research Intern

Bangalore, India

June 2019 - December 2020

- Examined microservice benchmarks to identify performance bottlenecks using tools such as Docker Stats, perf, mpStat and TShark
- Formulated placement algorithm to map containers of microservice to cores of server hardware and devised mechanism to coalesce services on commonly called paths into single container, achieving up to 26% improvement in latency and 29.5% in throughput
- Published and presented a paper in Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP-2021)

PROJECTS

Denoised smoothing as an adversarial defense for ASR

Carnegie Mellon University | August 2022

- Evaluated effectiveness of Denoising Diffusion Probabilistic Models (DDPMs) as defense against adversarial attacks on ASR systems
- Achieved 40% improvement in WER and better robustness compared to sequential randomised smoothing on LibriSpeech dataset

ALFRED-Speech: An Embodied Vision-Audio-Navigation Task

Carnegie Mellon University | August 2022

- Developed flask based web-app to crowdsource audio annotations for ALFRED benchmark on Amazon Mechanical Turk to create a first of a kind dataset to enable currently “deaf” embodied agents to learn to navigate by hearing and seeing
- Awarded Best Solution Capstone project and is currently enabling the collection of 25K+ audio annotations

Face Classification & Verification using Convolutional Neural Networks.

Carnegie Mellon University | August 2022

- Implemented and trained ResNet-34 and ConvNeXt-T models for face classification on subset of VGGFace2 dataset using image augmentation techniques and Stochastic Depth, achieving an accuracy of 94%
- Finetuned ConvNeXt-T model with center loss for face verification and achieved accuracy of 66%

Iterative Back-Translation-Style Data Augmentation for ASR and TTS

Carnegie Mellon University | January 2022

- Adapted back-translation style data augmentation using ASR and TTS outputs to improve performance for low resource languages
- Authored ESPNet recipes to train conformer-based ASR model employing linear fusion of HuBERT and spectrum-based features
- Achieved 6.91% reduction in word error rate (WER) for ASR and 2.91% improvement in Mel cepstral distortion (MCD) for TTS model

Neural Machine Translation for Low Resource Languages

Carnegie Mellon University | January 2022

- Trained baseline bilingual transformer-based models to translate between Belarusian/Azerbaijani and English using fairseq
- Improved BLEU score by 3 points through back translation and cross-lingual transfer by augmenting dataset with bitext from similar high resource languages chosen by LangRank

Cloud Native Micro-service Based Twitter Analytics

Carnegie Mellon University | August 2021

- Worked in team of 3 to design and develop scalable, fault-tolerant web-service with 3 microservices to recommend similar Twitter users with ~55K RPS and budget of \$1.12/hr using AWS EKS, RDS, Kubernetes and Docker
- Performed ETL on 1.2TB Twitter dataset using PySpark on Azure Databricks

PUBLICATIONS

Characterization and Detection of Parkinson's Disease, A Data Driven approach, 2020 IEEE International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)