Bharani Ujjaini Kempaiah

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

EDUCATION

Carnegie Mellon University – School of Computer Science — Pittsburgh, PA
Master of Computational Data Science | QPA: 4.12/4.33 — December 2022
Selected Coursework: Multilingual NLP, Cloud Computing, ML with Large Datasets

PES University — Bangalore, India
Bachelor of Technology in Computer Science & Engineering | GPA: 9.83/10 — May 2021

SKILLS

Programming Languages: Python, Java, C, R, SQL, Scala
Frameworks and Tools: Pytorch, Tensorflow, PySpark, ESPnet, Fairseq, Docker, Kubernetes, Kafka, Terraform
Databases: MySQL, HBase, MongoDB, Neo4j

EXPERIENCE

Cognistx — Pittsburgh, PA
Data Scientist Intern — May 2022 - Present

- Leading design and development of scalable web crawler, capable of crawling web pages with up to ~100K links and supports deduplication of embedded documents
- Implemented crawl component with AWS SQS URL frontier triggering AWS Lambda based recursive crawls while tracking crawl history in DynamoDB
- Developed decoupled AWS Lambda based downloader component to concurrently download crawled URLs into S3
- Projected to enhance domain generalization and increase downstream QA model performance across various domains

AppDynamics, Cisco Systems (India) Private Limited — Bangalore, India | Remote
Technical Undergraduate Intern — January 2021 - May 2021

- Developed Component Tests for OpenTelemetry supported agent using JUnit
- Created library to dynamically alter agent configurations. Written in Java, it uses SnakeYAML to construct custom Java objects and is currently being utilized 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 — Bangalore, India
Research Intern — June 2019 - December 2020

- Examined DeathStarBench and TeaStore microservices to identify performance bottlenecks responsible for reduced efficiency of microservices based applications using tools such as Docker Stats, perf, mpStat and TShark
- Identified and optimized kernel parameters for NUMA machines such as disk write frequency, kernel thread affinity
- 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

PROJECTS

Iterative Back-Translation-Style Data Augmentation for Low Resource ASR and TTS — Carnegie Mellon University | January 2022

- Adapted back-translation style data augmentation to speech processing by leveraging ASR and TTS outputs to improve each other's performance iteratively for low resource languages
- Implemented a conformer-based ASR model employing linear fusion of HuBERT and spectrum-based features. Utilized a combination of Glow-TTS and Hifi-GAN for TTS model
- Achieved up to 6.91% and 10.87% reduction in word error rate (WER) and character error rate (CER) respectively for ASR and 2.91% improvement in Mel cepstral distortion (MCD) for TTS model

Twitter User Recommendation — Carnegie Mellon University | August 2021

- Worked in a team of 3 to build an application for recommending similar Twitter users
- Designed an efficient and fault-tolerant web tier consisting of 3 microservices using Amazon EKS with managed node groups to handle high loads (~tens of thousands of RPS) under a constrained budget
- Performed ETL on a large Twitter data set (~1 TB) using Apache Spark on the Azure Databricks platform and deployed storage tier on an AWS RDS MySQL instance

PUBLICATIONS

Rao, Vishal, et al. "Scheduling Microservice Containers on Large Core Machines Through Placement and Coalescing." *Workshop on Job Scheduling Strategies for Parallel Processing*. Springer, Cham, 2021.

Mampilli, Ruben John, et al. "Characterization and detection of Parkinson's Disease, A data driven approach" International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE 2020)