Naga Harshita Marupaka

213-272-3555 marupaka@usc.edu linkedin.com/in/nagaharshitamarupaka github.com/NagaHarshita

Education

University of Southern California **Aug 2022–May 2024**
Master of Science in Computer Science *GPA: 3.73/4.0*

Indian Institute of Information Technology, SriCity **Aug 2018–May 2022**
Bachelor of Technology in Computer Science and Engineering *GPA: 9.12/10.0*

Technical Skills

Programming Languages: Python, Java, SQL, C/C++, JavaScript, TypeScript, Go, Swift
Web Frameworks: Django, React.js, Next.js, Node.js, Angular, Ajax, jQuery, HTML5/CSS
Data Science and ML: Pandas, Scikit-Learn, XGBoost, Spacy, Seaborn, PyTorch, TensorFlow, Spark, Hadoop
Tools and Platforms: GCP, AWS, Docker, Kubernetes, CUDA, Git, Firebase, Jira, Apache Airflow
Techniques: Exploratory Analysis, Predictive analytics, Deep Learning, Natural Language Processing (NLP), Recommender Systems, Statistical Analysis, Hypothesis Testing, Data Visualization

Work Experience

Machine Learning Intern, Informatics Team **June 2023–Present**
USC Alzheimer's Therapeutic Research Institute *San Diego, CA*

- Crafted a comparison report detailing 15+ open-source libraries for **speech recognition** with a specific focus on resource allocation, performance metrics, and benchmark analysis
- Engineered an efficient pipeline to provide video transcription with speaker distinction for cognitive sciences
- Integrated the video transcription outcomes into a React application, offering an intuitive visual representation
- Spearheaded the development of an interactive dashboard for a Laboratory Information Management System (LIMS) project, adhering to the principles of Test-Driven Development (TDD)

Software Developer Intern, Backend Team **Dec 2021–Jun 2022**
Josh Talks *Bengaluru, India*

- Designed and implemented a **group voice-calling feature**, enabling customizable speaker and listener settings
- Conducted a successful pilot for multiple audio chat rooms, each accommodating over 100 participants
- Contributed to configuring virtual machine instances, setting up CI/CD pipelines with Cloud Build trigger, and databases with cloud SQL auth proxy in Google Cloud Platform
- Programmed SQL views and queries to enhance the user engagement by scheduling engagement notifications based on user inactivity and analyze the million-user database for daily and monthly active users

Projects

Integrated Data Analytics Platform, Develop for Good - *Python, ETL, SQL, Airflow* **June 2023–August 2023**

- Led a team of 7 engineers in developing a streamlined, cost-efficient airflow pipeline to transfer Google Analytics (GA4) data into SQL warehouse and create materialized views for further analysis

Toxicity Detection & Mitigation in Reddit Comments, NLP - *Python, POS* **March 2023–May 2023**

- Trained a bi-directional GRU model to predict the toxicity score of scraped 100k reddit comments
- Implemented a novel algorithm that leverages POS tags to generate candidate sentences for offensive text and fine-tuned DistilRoBERTa model to fill the masked words of the candidate sentences

Event Finder, Web Technologies - *Node.js, React.js, GCP, Swift* **January 2023–May 2023**

- Orchestrated end-to-end development and deployment of a robust full-stack web application, integrating Ticketmaster API to empower users with advanced event search capabilities
- Developed a responsive iOS mobile app using Swift with similar functionalities, widening the application's accessibility

KPT - Open Source Project, Google Container Tools - *Go, GCP, Kubernetes* **May 2021–June 2021**

- Implemented an end-end feature request create-setters functionality for kpt tool to parameterize field values by adding setter comments in Kubernetes configuration files
- Assessed the output of the feature to handle different inputs through the unit tests and manual testing

Image Inpainting using Deep Learning, IIITS - *Computer Vision, GAN* **Jan 2021–Nov 2021**

- Reconstructed the lost pixels using Spectral Normalised GAN with Gated convolution and Self-Attention.
- Modified the TRANSGAN to produce results for the Image Inpainting subtask that uses only vision transformer blocks to learn the features of the image without convolution.