Disha Thotappala Jayaprakash

(408) 614-1508 | thotappa@usc.edu | www.linkedin.com/in/dishajayaprakash/

EDUCATION

University of Southern California Aug 2023 – May 2025
Master of Science focused on AI - Machine Learning and Data Science GPA: 3.93/4.0
Coursework: Natural Language Processing, Data Management, Digital Image Processing, Computational Introduction to Deep Learning, Machine Learning I: Supervised Methods, Linear Algebra, Probability

Dayananda Sagar College of Engineering Jul 2019 – Jun 2023
Bachelor of Engineering (BE) in Electronics and Telecommunication GPA: 3.96/4.0
Coursework: Artificial Intelligence & Machine Learning, Data Science, Python, Java, Complex variables and distribution, IoT and cloud computing, Digital Signal Processing

SKILLS

Programming Languages: Python, R, Java, C/C++, MATLAB, Linux, Bash/Shell Scripting, SQL, MySQL, PostgreSQL
Machine Learning & Modeling: Scikit-learn, XGBoost, SMOTE, SVM, Random Forest, RNN, MLP, Hyperparameter Tuning, Cross-Validation, Model Evaluation, Classification, Regression, Recommendation Systems
Languages & Programming: Python, Java, C++, JavaScript, SQL, Bash, TypeScript, HTML, CSS
Backend & Systems: Linux, Unix, Git, Docker, Kubernetes, REST APIs, FastAPI, Flask, Node.js, MongoDB, PostgreSQL, MySQL, Redis, Nginx, Kafka, gRPC, CI/CD, GitHub Actions
Software Engineering & Cloud: Object-Oriented Programming, Data Structures, Algorithms, Design Patterns, AWS, GCP, Azure, Microservices, Load Balancing, API Design, Unit Testing, Integration Testing, Agile, Scrum

RESEARCH EXPERIENCE

NLP Graduate Research Assistant | USC Keck Sep 2024 – May 2025

- Engineered real-time audio processing pipelines for MERLS using Python, enabling scalable emotion recognition and behavioral analysis in bilingual parent-child speech interactions
- Integrated multilingual speech models (e.g., Wav2Vec, SeamlessM4T) into production workflows; improved transcription accuracy by 30% in noisy environments
- Built an end-to-end system combining speech-to-text, LLMs, and custom logic for learning disorder detection; reduced annotation effort by 60% and optimized performance for low-resource deployments

EXPERIENCE

Software Engineering Intern (ML) | WorkUp June 2024 – July 2024

- Prototyped and shipped end-to-end NLP pipelines using spaCy, NLTK, and OpenAI APIs to parse job descriptions and match resumes; integrated transformer models into backend logic for real-time matching on a job-search platform
- Built a low-latency speech processing module with Librosa and Praat to extract prosodic features (pitch, tempo, intonation); accelerated response time by 30% using lightweight models in TensorFlow, PyTorch, and scikit-learn
- Containerized a real-time recommender system using Nvidia Merlin on SageMaker; integrated MLflow for lifecycle tracking and tuned performance with CUDA and Docker for scalable deployment

Software Engineering Intern (ML) | DigiAdd Technologies Sep 2022 – Nov 2022

- Designed and implemented a time-series processing module to track and forecast flight delays, improving system responsiveness and reducing reliance on third-party data services by 80%
- Developed scalable data ingestion and transformation pipelines to handle 1.2M+ flight records using SQL and Power BI; enabled automated reporting and regional analytics across operational teams
- Built a text similarity service using TF-IDF and cosine distance to enhance search and recommendation functionality, integrated preprocessing and retrieval logic into a modular backend pipeline

ACADEMIC PROJECTS

Design Pattern Playground (Link) Apr 2025 – Present

- Developed a real-time design pattern sandbox using Flask and WebSockets; enabled interactive backend logic demos with <200ms UI update latency under concurrent sessions
- Refactored pattern modules into plug-and-play services using Python's AST and strategy pattern; reduced code duplication by 60% and cut onboarding time for new patterns by 50%

Trojan Map (Link) Feb 2025 – May 2025

- Implemented a modular C++ system supporting location search, regex matching, topological sorting, and pathfinding
- Integrated graph and string algorithms to power shortest path routing, fuzzy search, and delivery planning; emphasized performance trade-offs, runtime analysis, and maintainable code architecture for large-scale geographic data

Stutter Classification (Link) Nov 2023 – Dec 2023

- Processed and analyzed speech recordings from SEP-28k and FluencyBank to extract disfluency patterns for classification
- Extracted acoustic features (pitch, pause duration, speech rate) using Librosa, Praat, and Pandas to support stutter pattern analysis