

Disha Thotappala Jayaprakash

(408) 614-1508 | thotappala@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++, Linux, Bash/Shell Scripting, SQL, MySQL, PostgreSQL

Software Development & MLOps: Git, Docker, Kubernetes, Flask, FastAPI, REST APIs, GitHub Actions, MLflow (MLflow, Kubeflow, CI/CD, drift detection, versioning), Kubeflow, AWS SageMaker, GCP, Snowflake, Apache Spark, Linux, Data Structures and Algorithms

Machine Learning & Modeling: Scikit-learn, XGBoost, SMOTE, SVM, Random Forest, RNN, MLP, Hyperparameter Tuning, Cross-Validation, Model Evaluation, Classification, Regression, Recommendation Systems, Deep Learning, Computer Vision, Artificial Intelligence, Predictive Modeling, Statistical Analysis

RESEARCH EXPERIENCE

NLP Graduate Research Assistant, USC Keck

Sep 2024 – May 2025

- Developing **real-time ASR** pipelines and voice-driven models for MERLS to support emotion recognition and behavioral analysis in bilingual interactions
- Fine-tuning multilingual speech models (e.g., Wav2Vec, SeamlessM4T) with PyTorch and JAX for noisy Mandarin-English speech; **reduced WER by 30%**
- Leveraging speech-to-text, LLMs, and transfer learning for learning disorder detection; **cut annotation time by 60%** and optimized for low-latency, resource-constrained deployment

EXPERIENCE

Machine Learning Intern, WorkUp

June 2024 – July 2024

- Developed autonomous **AI agents** using spaCy and NLTK, semantic embeddings and entity recognition to parse job descriptions, and enhance resume-job alignment via transformers, OpenAI APIs, and semantic similarity models
- Extracted prosodic features (pitch, tempo, intonation) using Librosa and Praat; applied feature engineering and modeling with scikit-learn, PyTorch, and TensorFlow, **reducing speech processing latency by 30%**
- Built a personalized recommendation system on **AWS SageMaker** using Nvidia Merlin; deployed with Docker and MLflow, monitored via Kubeflow, and optimized inference performance with CUDA

Machine Learning Intern, DigiAdd Technologies

Sep 2022 – Nov 2022

- Devised a recurrent neural network (RNN) in TensorFlow to forecast flight delays with **95% accuracy**, **reducing radar dependency by 80%** through temporal modeling
- Investigated **1.2M+ flight records** using SQL and Power BI to uncover delay patterns across regions, enabling predictive insights and operational refinement
- Leveraged TF-IDF and cosine similarity on **50K+ song lyrics** to expose thematic clusters, informing personalized recommendation strategies and user segmentation

ACADEMIC PROJECTS

KPI Explorer

Jan 2025 – Present

- Built interactive dashboards in **Streamlit** and **Plotly** to track KPIs; applied business analytics methods to assess trends
- Created **ETL pipelines** from SQL/CSV; performed data cleaning, forecasting, regression analysis, and statistics with Pandas; embedded Tableau for operations management insights

Autocorrect system ([Link](#))

Oct 2024 – Dec 2024

- Designed a **context-aware autocorrect system** in Python using bidirectional bigram language models, Levenshtein distance and implemented efficient tokenization, candidate pruning, and correction heuristics for **corrupted text recovery**
- Outperformed baseline Ostrich algorithm with a **75% higher exact match rate** and **~40% lower WER** by balancing contextual accuracy and spelling similarity

Stutter Classification ([Link](#))

Nov 2023 – Dec 2023

- Built an ASR-based disfluency detection system using SEP-28k and FluencyBank; extracted features via wav2vec and MFCC for real-time speech analysis with **F1 score of 0.71**
- Adapted ResNet and ConvLSTM using **PyTorch, Librosa, and Praat**; optimized for latency and lightweight inference