Pranitha Velusamy Sundararaj

(313) 414-4989 | pranithavs2310@gmail.com | linkedin.com/in/pranitha-v-s | github.com/pranithavs | pranithavs.github.io/Portfolio/

Summary

Data Scientist with expertise in predictive modeling, cloud-native solutions, and large-scale data processing. Proficient in **Python, SQL, AWS** (Lex, Lambda, Rekognition), and **ML frameworks** (TensorFlow, Scikit-Learn). Skilled in ETL workflows, Spark-driven analytics, and end-to-end deployment of AI tools. Passionate about transforming raw data into actionable insights for scalable, real-world impact.

Experience

Internship - SAMSUNG, India *Jul 2021 – Mar 2022*

- Developed Bi-Directional LSTM RNN with **Manhattan Distance**, achieving **81.62%** similarity accuracy on Quora dataset.
- Preprocessed **417K+** sentence pairs using **NLTK** for tokenization, cleaning, and lemmatization.
- Trained model with Keras embeddings, deployed using **TensorFlow** Lite for on-device efficiency.
- Integrated model into Android app using **Stanford CoreNLP** for dynamic MWUC sentence handling.
- Achieved **95.65%** accuracy in semantic similarity and improved voice assistant usability via MWUC reformulation.

Personal Project - Deep Audio Classifier *Jan 2025 – Apr 2025*

- Engineered a data pipeline for efficient loading and preprocessing of audio files, converting them into spectrograms for **CNN** input.
- Preprocessed rainforest audio using **Librosa** with spectrograms, noise reduction, and data augmentation.
- Applied early stopping, learning rate scheduling, **L2** regularization, and dropout to reduce overfitting, achieving **96%** accuracy and **75%** precision.

Personal Project - Student Performance Prediction *Aug 2024 – Dec 2024*

- Engineered a high-accuracy **Multiple Linear Regression (MLR)** model using Python/Scikit-Learn for performance prediction on a **10,000-record dataset**.
- Evaluated and compared MLR (**Test R^2: 0.9884**) against Simple Linear Regression (**Test R^2: 0.8351**) and Random Forest Regressor (**Test R^2: 0.8165**) to select the optimal predictive model.
- Identified key predictors via EDA/correlation analysis leveraging Matplotlib and Tableau visualizations.

Personal Project - Dining Concierge Chatbot *Aug 2024 – Dec 2024*

- Deployed serverless chatbot using **AWS Lex** (NLP), **Lambda** (Python), and **API Gateway** for restaurant suggestions.
- Integrated **Yelp API** and **DynamoDB** (NoSQL), enabling real-time data storage and dynamic retrieval.
- Developed Lambda functions orchestrating conversational AI flow and backend API resource interactions.

Personal Project - Intelligent Photo Album Search Application *Aug 2024 – Dec 2024*

- Developed a search Lambda function to process **natural language queries** and query **ElasticSearch** using the ElasticSearch SDK.
- Implemented serverless Lambda (Python) triggered by **S3** events for automated photo metadata indexing.
- Integrated Amazon Transcribe to enable speech-to-text functionality, allowing users to perform voice-based photo searches.

Personal Project - Smart Health Disease Prediction *Jan 2024 – Apr 2024*

- Implemented a probabilistic **Naive Bayes classifier**, leveraging its feature independence assumption for efficient symptom-based disease prediction.
- Performed comprehensive **data preprocessing** to handle missing values and **encode categorical symptom data** for model compatibility.
- Achieved **92.7%** prediction accuracy and performed comprehensive model evaluation to ensure the classifier's robustness.

Personal Project - Recommendation Systems *Aug 2023 – Dec 2023*

- Developed a robust product recommendation model utilizing Amazon large data sets (**8M rows**) in **Pyspark**.
- Optimized big data through **preprocessing** techniques, **strategic data mapping**, and implementation of **collaborative filtering** and **content-based filtering** methodologies.
- Achieved optimized **ALS model** performance (**RMSE 0.19**) through evaluation within the Spark ecosystem.

Education

University of Michigan *Aug 2023 – May 2025*
Master of Science in Data Science *GPA: 3.77*

PSG College of Technology- Coimbatore, India *Aug 2019 – July 2023*
Bachelor of Engineering in Computer Science *GPA: 3.33*

Skills & Interests

Languages: Python, SQL (MySQL, PostgreSQL), R Programming, PySpark
Frameworks & Tools: Pandas, Hadoop, Apache Spark, Docker, NLP
Cloud Platforms: AWS (S3, Lambda, Rekognition, API Gateway, DynamoDB, Lex, IAM, Kinesis, VPC), Snowflake
Data Visualization: Tableau, Power BI, Matplotlib, Seaborn
Machine Learning: Scikit-Learn, TensorFlow, Keras
Developer Tools: Jupyter Notebook, GitHub, RStudio, Google Colab, VS Code, MS Word, Excel, PowerPoint