Sree Charan Reddy Kailasam

| +1 443-285-9885| [Portfolio](#) | [sreechk1@umbc.edu](#) | [LinkedIn](#) | Open to Relocate |

Professional Summary

Machine Learning Engineer specializing in NLP and LLM-based systems, with experience building scalable data pipelines and retrieval systems for financial and nonprofit datasets. Proven experience developing production-ready ML applications using AWS, transformer models, and vector search to solve real-world data challenges.

Education

Master of Science in Data Science | University of Maryland, Baltimore County   -   GPA : 3.84 / 4.0

Bachelor of Technology in Computer Science and Engineering | Vellore Institute Of Technology, Vellore, India   -   GPA : 8.45 / 10

Skills

Languages: Python, SQL, Java, C++, Bash

ML & AI: Machine Learning, Deep Learning, NLP, LLMs, RAG, Information Retrieval, Time-Series

Frameworks: PyTorch, TensorFlow, Scikit-learn, XGBoost, Hugging Face

Cloud & MLOps: AWS (SageMaker, S3, Lambda, EC2), Docker, CI/CD

Data Systems: Spark, FAISS, Airflow, Pandas, NumPy

Data Visualization & BI: Tableau, Power BI, Matplotlib, Seaborn

Backend & Databases: FastAPI, Flask, PostgreSQL, MySQL, Redis

Work Experience

WikiCharities, Bountiful, UT (Remote)   **Aug 2025 – Present**

Machine Learning Engineer

- Architected LLM-driven data pipelines using transformer models and RAG to process nonprofit datasets across **190+ countries**, improving metadata accuracy by **42%** and reducing manual curation effort by **35%**
- Built microservices-based backend systems (FastAPI) integrating GPT-based models for automated financial insights and reporting, improving system scalability and reducing analysis time
- Designed and deployed scalable ML infrastructure on AWS (SageMaker, Lambda, S3), enabling automated training and batch inference, improving pipeline efficiency by **30%**
- Implemented semantic retrieval using FAISS and dense embeddings, improving document discovery and search relevance across large unstructured datasets
- Built evaluation and re-ranking pipelines using embedding similarity and prompt optimization, improving LLM response consistency and reducing hallucinated outputs

Progment Software Technologies, Hyderabad, India   **Apr 2021 – Nov 2023**

Data Scientist

- Designed end-to-end ML pipelines combining time-series forecasting (LSTM, XGBoost) with automated feature engineering, improving prediction accuracy by **24%** on financial datasets
- Built large-scale NLP pipelines using BERT and Hugging Face Transformers to process **12K+ financial news records/day**, improving sentiment classification performance and downstream analytics
- Applied unsupervised learning techniques (LDA, K-Means) to segment financial datasets, improving pattern discovery and enabling more targeted analytics
- Developed anomaly detection systems using Isolation Forest and Autoencoders, identifying **250+ anomalous transactions/month** and improving fraud detection coverage
- Engineered data pipelines using Python, SQL, and Linux workflows, improving data availability and reducing processing time for ML systems
- Built and deployed ML pipelines on AWS SageMaker, reducing deployment latency by **45%** and improving scalability
- Conducted model validation using cross-validation and A/B testing, improving model reliability and decision accuracy

Data Analyst   **Feb 2020 – Mar 2021**

- Performed exploratory data analysis on financial datasets, identifying trends and anomalies that improved reporting accuracy
- Built SQL-based data pipelines and dashboards, improving reporting efficiency and reducing manual data handling by **20%**
- Assisted in implementing clustering (K-Means) and anomaly detection models, improving early-stage fraud detection capabilities

Projects

- **Distributed Semantic Search Engine (AI Search System) :** Built a semantic search system using transformer embeddings and FAISS to enable efficient similarity search across large unstructured datasets. Designed ANN-based indexing and embedding-driven re-ranking pipelines to improve top-k retrieval accuracy and reduce query latency. Integrated REST APIs for real-time search and ranking, improving system responsiveness and throughput.
- **Machine Learning Feature Store for Scalable Model Training :** Developed a centralized feature store to standardize and reuse features across ML pipelines, incorporating feature versioning and lineage tracking. Built automated feature pipelines using Spark and SQL to reduce redundant feature engineering and improve training efficiency. Integrated the feature store with ML workflows to improve consistency and model reproducibility.