Sai Siva Reddy Maddula

maddulasaisivareddy@gmail.com | 3614881962 | Tampa, FL | https://o1mssr.online
https://www.linkedin.com/in/sai-siva-reddy-m-964504224/ | https://www.github.com/Mpr9640

Profile:

Machine Learning Engineer with strong software engineering skills and hands-on experience in building ML-powered applications, including NLP, OCR, and model deployment pipelines. Passionate about developing intelligent systems that are efficient, ethical, and production-ready. Proven ability to integrate ML models with scalable APIs and real-time user interfaces.

Skills:

- **Programming & Core Concepts**: Python, C++, PostgreSQL, Data Structures & Algorithms.
- **ML & DL Concepts**: Supervised & Unsupervised Learning, Model Training, Evaluation, Tuning, Linear & Logistic Regression, Decision Trees, KNN, SVM, Neural Networks (MLP), Fairness Indicator.
- **Libraries & Tools**: Scikit-learn, Pandas, TensorFlow, PyTorch, NumPy, Keras, Matplotlib, Seaborn.
- **APIs & Deployment**: Fast API, Google Colab, VS Code, GitHub, GCP(cloud), Docker.
- **Mathematics**: Linear Algebra, Probability & Statistics, Calculus, Optimization

Experience:

ML Software Engineer, CORtracker Inc., 03/2025 - Present

Project: Real-time Extraction-based Q&A ML System (SQUAD Model)

- Collaborating on the development of an ML-driven Q&A system by **integrating pre-trained transformer models for document-based question answering**, focusing on real-time inference and latency optimization.
- Designing backend logic to **handle document ingestion, text extraction, and query preprocessing**, ensuring accurate context generation before model inference.
- Implementing **answer ranking pipelines** that evaluate multiple model outputs based on confidence scores and contextual relevance, displaying only the most reliable response to end users.
- Continuously fine-tuning deployment pipelines by **monitoring model performance in production**, adjusting parameters for faster response times while maintaining output quality.
- Supporting a minimalistic, task-focused frontend that allows users to upload files, submit queries, and view staged answer states (thinking, blurred preview, final answer) — optimizing for ML-specific use cases.

ML Software Engineer, **Datics Inc (**Intern -> Contract) 01/2024 - 03/2025

Project: Job Application Automatic Filler (Web App + Chrome Extension)

- Developed ML models for form input field classification using surrounding labels and DOM context, **helping users autofill forms accurately without manual corrections.**
- Built resume parsing pipelines using NER models to extract skills, education, and experience, **enabling automatic form pre-filling and reducing user effort during job applications.**
- Implemented skill-matching algorithms comparing resume skills with job descriptions, **providing users with a match percentage to help them decide whether to apply.**
- Implementing MLOps workflows including CI/CD pipelines, automated retraining triggers, and experiment tracking to **maintain model performance** post-deployment
- Deploying ML models in cloud environments (GCP), ensuring scalability and secure API accessibility for both internal tools and customer-facing applications.
- Applied backend caching strategies for parsed resumes and model predictions, **reducing repeated server calls and cutting autofill response time from seconds to milliseconds.**

Projects:

- Fairness Evaluation using TensorFlow Fairness Indicators – ACS Dataset
- Transfer Learning (Fashion MNIST)
- Data Augmentation (MNIST)
- Linear Regression (Chicago Taxi Dataset)

Education:

- **Master** of Science in **Computer Science**, Auburn University at Montgomery, Graduated: May 2024, **CGPA: 3.83**
- **Bachelor** of Technology **in Electrical and Electronics Engineering**, India, Graduated: June 2022, **CGPA: 3.41**

Certificates:

- AWS Academy Graduate – **AWS Academy Cloud Foundations**.
- **Ucertify** course **for Data Mining and Data Visualization**.