Srikanth Malyala
Applied Data Scientist II

(+91) 9092929824
Bengaluru, Karnataka
malyala2srikanth@gmail.com

Leetcode: Malyala Srikanth
Github Profiles: Personal | Professional
Linkedin: srikanth-malyala

Experienced Data Science Professional with 4+ years in Natural Language Processing (NLP), Machine Learning, and Deep Learning. Proven expertise in building, training, testing, and deploying models into production. Strong proficiency in developing APIs using Flask, Django, FastAPI, and Celery. Adept at working in agile startup environments, collaborating with cross-functional teams to solve complex problems. Passionate about Generative AI, Large Language Models (LLMs), and Recommender Systems.

SKILLS

Tools and Languages	Python, C#, MySQL, OpenSearch, NEO4J, AWS (S3, SageMaker, Copilot, EKS, ECS, CodePipeline, OpenSearch), Linux Environments, Git, DevSpace, Kubernetes, Docker
Frameworks / Libraries	PyTorch, scikit-learn, SpaCy, NLTK, Flask, FastAPI, Celery, MlFlow, DVC, LangChain, CrewAI
Knowledge Areas	Machine Learning, Deep Learning, NLP (LLM, BERT, MPNet), Generative AI, RecSys, MCP

TECHNICAL EXPERIENCE

Applied Data Scientist II   **Nov 2024 — Present**
Future AGI (A Platform to build, evaluate, observe and optimize LLM applications)   *Bengaluru, Karnataka*

- Led the research and development of the futureagi SDK enabling seamless developer communication with platform features, and contributed to comprehensive SDK documentation.
- Developed an MCP server from scratch named futureagi-mcp-server to interact with futureagi functionalities using natural language.
- Built Django APIs with designers and frontend teams for seamless integration. and meeting product requirements.
- Researched on LLM as a Judge and setting up a benchmarking pipeline to evaluate the LLM as a Judge based evaluation agents using ground truth data.

Data Scientist II   **April 2023 — Oct 2024**
VMock India Pvt Ltd   *Gurgaon, Haryana*
*Smart Career Analytics platform intended to offer CV feedback services enabling users across universities like **CMU, Stanford, Chicago Booth, University of Oxford, Columbia, Kellogg** to help write more effective resumes.*

- Finetuned **LayoutLMForTokenClassification** language model using transformers library to extract named entities in experience section which uses both text and spatial information from resume obtaining overall f1 score of **0.844**.
- Researched on **efficient single GPU training** methodologies using **PyTorch** and presented it to the entire team of 12 data scientists, useful for faster and cost-effective finetuning of **LLM's**.
- Extracted bullets with weak action verbs from resume and leveraged **semantic search** using **multi-qa-mpnet-base-dot-v1** contextual embeddings to suggest relevant strong action verbs with the help of **OpenSearch (search engine)**.
- Built a scalable event-driven service using **AWS SNS** and **GPT-4o (LLM)** to extract and structure information from image PDFs, providing feedback using prompt engineering.

Data Scientist I   **Dec 2020 — Mar 2023**
VMock India Pvt Ltd   *Gurgaon, Haryana*

- Researched and implemented **ASPOSE**, a third-party .NET-based tool, to provide support for **DOCX Format** for parsing resumes uploaded by students across **260+ universities and 10000+ benchmarks**.
- Retrained **Partially Labelled Latent Dirichlet Allocation (PLDA)** to predict resume super-sections based on n-grams, improving top-1 accuracy from **81.9% to 89.6%** and top-2 accuracy from **92.2% to 96.7%** through k-fold cross-validation.
- Finetuned **bert-base-cased** large language model for merging bullets written in multiple lines (NSP), utilizing **DVC** for efficient data tracking and **MlFlow** for metrics tracking and obtained an accuracy of around **93%**.
- Implemented **recommendation algorithm** that suggests similar jobs and candidates based on users past interactions, enhancing the overall user experience using **node2vec** in **NEO4J** and obtained **NDCG of 0.76**.

PROJECTS

- Cleaned data, applied SMOTE, and trained a **Gradient Boosting model** to predict fraudulent vehicle claims (**94%** accuracy).
- Trained a CNN model using Keras for crack detection, achieving 87.4% validation accuracy based on **All-CNN paper**.
- Built a RAG framework with LangChain, Elasticsearch, OpenAI, FastAPI and deepeval implementing different **retrieval methods**

EDUCATION

Bachelor of Technology in Civil Engineering, *Indian Institute of Technology Madras*   2016 - 2020
Machine Learning in Civil Engineering   Data Structures and Algorithms in Python
Probability, Statistics and Stochastic Processes   Introduction to Programming

ACTIVITIES

Figured in top 1.2 percentile in JEE Advanced (AIR 2440) and top 0.5 percentile in JEE Mains (AIR 5582)
Awarded as Best Team Player for consistently assisting fellow teammates in the workplace
Won gold in Interdepartmental Chess Competition