LINGESH S

Email: lingeshsrinivasan1@gmail.com
LinkedIn: www.linkedin.com/in/lingesh-s29
Phone No: 7299307308
GitHub: github.com/Lingesh-S

ABOUT ME

Data Scientist with hands-on experience building production-grade ML and GenAI systems at a product-based company. Strong background in XBRL auto-tagging using LSTM & NER, taxonomy-specific ML pipelines, and RAG-based systems for real-time applications. Experienced in deploying scalable ML services, working across multiple regulatory domains, and building LLM-powered features for enterprise products.

WORK EXPERIENCE

Data Scientist

DataTracks | Feb 2025 – Present

- Built and maintained LSTM + NER–based XBRL auto-tagging models for multiple region-specific taxonomies across core compliance products.
- Designed independent ML pipelines per taxonomy, and trained & deployed two new taxonomy pipelines end-to-end (data prep → training → production integration).
- Developed a RAG-based next-word prediction system for a Word-like editor in a flagship product using Haystack, FastAPI and RabbitMQ.
- Implemented production monitoring and observability for ML services using Grafana and Prometheus.

LLM Intern

Dtyle.AI (IIT Madras Incubated) | Nov 2024 – Jan 2025

- Appointed as a virtual intern to work on Large Language Models and Generative AI.
- Collaborating with the tech team under the CTO's mentorship on NLP-centric tasks.
- Contributing to prompt engineering, fine-tuning, and evaluation of LLM-based systems.
- Engaging in research, experimentation, and practical application of transformer models.
- Ensuring data privacy and following responsible AI practices in project workflows.
- Developing RAG-based architectures using Flan-T5 and LangChain for modular, context-aware Q&A applications.

SKILLS SUMMARY

- **Languages:** Python, SQL
- **Databases:** MySQL, Vector Databases (ChromaDB / FAISS)
- **ML & NLP:** LSTM, NER, Transformers, Fine-tuning (LoRA, PEFT)
- **GenAI / RAG:** LangChain, Haystack, Retrieval-Augmented Generation
- **ML Stack:** Scikit-learn, TensorFlow, PyTorch
- **Backend, Infra & Monitoring:** FastAPI, RabbitMQ, Grafana, Prometheus
- **Data & Tools:** Pandas, NumPy, Git, Streamlit, Flask

EDUCATION

B.E. Computer Science Engineering — **2020 - 2024**
Sathyabama University, Chennai — **CGPA:** 8.41

PROJECTS

1. Codebase RAG Assistant

- Designed a Retrieval-Augmented Generation (RAG) system that answers questions about code using Mistral-7B, combining transformer-based generation with semantic search from a vector store.
- Built a pipeline to load and chunk .py, .md, and .json files, then embed them with SentenceTransformers and store in a Chroma/FAISS vector database for high-precision code retrieval.
- Integrated LangChain, Mistral-7B, ChromaDB, and Streamlit to deliver real-time code understanding, explanation, and developer support via a modular assistant.
- GitHub Link: (bit.ly/4kxpzFu)

2. LDA-Powered Resume Screening Dashboard

Presented at ICCCAI-2024 (Taylor's University, Malaysia)

- Developed an interactive Streamlit dashboard using LDA (Latent Dirichlet Allocation) for extracting dominant topics and keywords from resumes.
- Designed a custom NLP-based job matching algorithm with token indexing, achieving >80% accuracy in simulated resume screenings.
- Integrated OpenAI API to auto-generate personalized interview questions, enhancing recruiter productivity through AI-powered automation.
- GitHub Link: (bit.ly/438nhXw)

CERTIFICATES

- Machine Learning Internship – Cognibot Labs
- Cloud Data Management Associate - Oracle
- Data Analytics Virtual Experience - Accenture

ACHIEVEMENTS

- Published and presented a research paper titled "LDA-Powered Resume Screening Dashboard with Token Indexing & Streamlit" at ICCCAI-2024, in association with Taylor's University, Malaysia.
- Developed and deployed a personalized learning chatbot for students using LangChain and Flan-T5, enabling real-time, context-aware Q&A from academic notes.
- Engineered a RAG-powered Codebase Assistant using LLMs, LangChain, and vector databases to help programmers understand, query, and explore codebases through natural language.