Rohan Shinde

(213) 234-8020 | rohanshi@usc.edu | linkedin.com/in/rohanshinde24 | Los Angeles, CA

Education

University of Southern California — Los Angeles, CA
Master of Science in Computer Science — *Jan.'25 – Present*
Coursework: Analysis of Algorithms, Information Retrieval and Web Search Engines

Maharashtra Institute of Technology — Pune, MH
Bachelor of Technology in Computer Science and Engineering | GPA - 3.74 — *2020 – 2024*
Relevant Coursework: Software Engineering, Database Management Systems, Machine Learning, Big Data Technologies

Experience

DMI Finance — Jul 2024 – Oct 2024
Data Science Intern — *Delhi, DL, In*

- **Fine-tuned Llama 3.0 and Llama 3.1 models** for synthetic financial data generation, accomplishing a **191.67%** accuracy improvement of data generated across **35 classes**, utilizing advanced token compression and Chain-of-Thought prompting techniques with a focus on reducing hallucination
- Independently optimized **BERT-based models**, attaining **98% accuracy** in classifying 35 types of financial data
- Collaboratively engineered **data generation and labeling pipelines** based on **Groq** and **vLLM**, facilitating efficient training and data labeling for large-scale financial applications
- Boosted inference throughput by **100%**, leveraging **LlamaCpp** for efficient model deployment and low-latency inference at scale

ResoluteAI Software — Dec 2023 – May 2024
Data Science Intern - GenerativeAI — *Bengaluru, KA, In*

- Developed an **advanced data visualization and analysis web application** leveraging **Streamlit, OpenAI, LangChain, and Pandas**, integrating audio recording and authentication functionalities
- Spearheaded migration from ChatGPT to CodeLlama, achieving **annual cost savings of $10,000** by optimizing codebase efficiency and cutting down resource usage
- Collaborated with senior data scientists and engineers to refine generative AI models, conducting thorough error analysis and debugging, **resolving 95% of bugs** and improving model robustness

Maharashtra Institute of Technology — July 2022 – Feb 2023
Machine Learning Research Assistant — *Pune, MH, In*

- Applied TF-IDF for feature extraction, boosting model precision and accuracy by 15%, and achieved **78.57% accuracy** in requirement type classification using Random Forest and K-Nearest Neighbor algorithms
- Automated software requirement classification with machine learning models, **reducing manual classification time by 50%** and increasing efficiency in the software development lifecycle

Projects

BrainInsight | *Javascript, Python, Flask, PyTorch, SQLAlchemy, Git* — Jul 2023 – Oct 2023

- Designed and developed a full-stack web application to serve a **REST API** for brain tumor classification based on **RESNET50**, with optimizations for real-time predictions and seamless user interaction.
- Engineered a K-Means algorithm with **99.32% test accuracy** for brain tumor segmentation and devised a real-time anomaly detection system, **reducing diagnosis time by 50%** and improving diagnostic accuracy

Movie Recommendation App | *Python, Streamlit, Natural Language Processing, NLTK* — Jan 2023 – Apr 2023

- Led development of a Movie Recommender System using Demographic/Content-Based Filtering, **increasing user interaction by 25%** and **recommendation accuracy by 20%** through **personalized strategies**
- Designed and deployed an intuitive Streamlit web interface, **promoting user adoption by 40%** and enhancing overall user experience and engagement with the recommendation system

Technical Skills

Languages: Python, SQL, JavaScript, TypeScript, HTML, CSS
Frameworks: Flask, FastAPI, PyTorch, TensorFlow, ReactJS, NodeJS, ExpressJS
Developer Tools: AWS, Git, Github Actions, Jenkins, Docker, PostgreSQL, MongoDB, Redis, Tableau, Excel
Libraries: scikit-learn, Pandas, NumPy, Matplotlib, LangChain, Keras, Streamlit