

YASHASHVINI RACHAMALLU

rachama2@msu.edu | [linkedin.com/in/yashashvinir/](https://www.linkedin.com/in/yashashvinir/) | github.com/yashu-dot | yashu-dot.github.io/ | +1-(517) 282-3989

PROFESSIONAL EXPERIENCE

Software Engineer - LLM Modeling | Rice F.W. Technologies | East Lansing, MI

May 2024 – Present

- Developed and optimized an in-house training codebase for large-scale multimodal transformer models using Whisper for speech-to-text and LoRA-fine-tuned LLaMA2-70B, reducing document processing time by 60% and costs by 50%.
- Produced a RAG-based AI chatbot by integrating PyTorch, Azure OpenAI, Search AI Services, and Flask APIs within an Agile SDLC to support query resolution and streamline onboarding, achieving a 50% reduction in time and costs.
- Leveraged advanced time-series forecasting to develop a customer-centric inventory prediction solution proactively aligning stock with demand, preventing stockouts, and driving a 20% increase in annual sales.
- Migrated over 200K+ legacy MySQL records with custom ETL pipeline, including end-to-end data preprocessing, schema optimization, and attribute mapping, enabling seamless deployments using Azure ARM templates to automate server provisioning and infrastructure setup for enhanced **E-commerce Platform** performance.

Software Engineer - Machine Learning | Michigan State University | East Lansing, MI

Jan 2023 – Dec 2024

- Accelerated geospatial data analysis as measured by processing 1M+ points per second by developing a classification system using Open3D, cKDTree, and multithreading optimization.
- Enhanced Mask R-CNN with transformer-based architectures, boosting crack detection accuracy by 40% and accelerating response times by 30% for real-time infrastructure monitoring.

ML Engineer - ASIC & ML Design | Intel Corporation | Bengaluru, KA

Jun 2021 – Jul 2022

- Reduced project costs by 70% and improved chip-placement accuracy by integrating GraphSAGE neural networks and K-means clustering, leveraging graph-based ML techniques for enhanced ASIC workflows.
- Increased throughput by 25% by developing advanced ETL pipelines with TCL scripting, enabling seamless extraction and preprocessing of complex design data for efficient integration into EDA tools.
- Optimized ASIC design metrics, achieving a 30% improvement in wire length, power, and performance through ML-driven automated placement guidance and enhanced design workflows.

EDUCATION

Michigan State University, East Lansing, MI

Jan 2023 - Dec 2024

Master of Science, Computer Science

GPA - 4.0/4.0

PES University, Bengaluru, India

Aug 2018 - Jun 2022

Bachelor of Technology, CSE with Machine Intelligence & Data science specialization

GPA - 3.8/4.0

TECHNICAL SKILLS

Programming Languages: Python, SQL, C, C++, MATLAB, JavaScript, Java.

Machine Learning & AI • Frameworks: TensorFlow, PyTorch, Scikit-learn, PySpark • GenAI: LangChain, HuggingFace, OpenAI, Gemini • Libraries: Pandas, NumPy, OpenCV, NLTK, XGBoost • Vector Stores: FAISS, ChromaDB

Cloud & DevOps • AWS: SageMaker, Bedrock, AMI • Azure: AI Services, VMs, KeyVaults, SQL Server, IAM • Tools: Docker, Git, JIRA, Postman, Linux/Unix

Data & Databases • SQL: PostgreSQL, MySQL • NoSQL: Neo4.js • Visualization: Tableau, Matplotlib, Seaborn, D3.js

Web Development • Backend: Flask, RESTful APIs • Frontend: HTML/CSS, Bootstrap, Streamlit

PUBLICATIONS

- Rachamallu, Y., et al. (2022). Image caption generation for low light images. *ICICCT 2022*, Springer. https://doi.org/10.1007/978-3-031-20977-2_5
- Rachamallu, Y., et al. (2022). Fake news detection on Indian sources. *ICIPCN 2022, Lecture Notes in Networks and Systems*, 514. Springer. https://doi.org/10.1007/978-3-031-12413-6_3
- Rachamallu, Y., et al. (2021). Similar face detection for Indian faces using Siamese Neural Networks. *2021 2nd INCET*, 1–5. <https://doi.org/10.1109/INCET51464.2021.9456199>

PROJECTS

Restaurant Recommendation System | Built a scalable, real-time restaurant recommendation and ranking engine achieving 87% precision by processing a large scale dataset of 1M+ Yelp reviews using Hadoop, Spark for distributed data analysis

COVID-19 Classification Using Computer Vision | Boosted COVID-19 diagnostic accuracy to 97% by training Xception and MobileNet-v2 CNNs on GAN-enhanced lung MRI data, deploying the model via AWS SageMaker for scalable solution.