Varun Woodi Raghavendra

860-994-9535 | woodirvarun@gmail.com | linkedin.com/in/varun-w-r/ | wrvarun-96.github.io/varun_wr_portfolio/ | github.com/wrvarun-96

Professional Summary

Data scientist with 3+ years of experience in predictive modeling, analytics, and cross-functional collaboration. Proficient in Python, SQL, and R for data analysis and validation, with expertise in developing advanced models and implementing LLM techniques. Delivered impactful results in medical image analysis and network security through scalable pipelines and data-driven strategies. Skilled at identifying user patterns, refining metrics, and driving business-aligned solutions

Skills

Languages:	Python, SQL, R.
Framework/Libraries:	TensorFlow, PyTorch, LangChain, Spark,Seaborn, Streamlit, OpenCV, Pandas, NumPy, Scikit-learn.
Machine Learning:	Linear & Logistic Regression, Boosting, Bagging, SVM, Naive Bayes, PCA, Forecasting.
Deep Learning :	ANN, CNN, RNN, LSTM, Encoder-Decoder, Transformers, Computer Vision, Natural Language Processing.
LLMs:	Fine-tuning, Prompt Engineering, RAG.
MLOps:	CI/CD, Docker, MLflow, Kubeflow, Model Monitoring, Version Control (Git).
Cloud:	AWS (SageMaker, S3, EC2, ECR).
Databases:	PostgreSQL, MongoDB.
Mathematics:	Statistics & Probability, Calculus, Linear Algebra, A/B testing, Hypothesis Testing.
Visualization:	Tableau, Power BI.

Work Experience

HEALTHI AI (Volunteer) • Massachusetts **Dec 2024 – Present**

DATA SCIENTIST

- Enhancing Caddy Chatbot with amazon.titan-embed-text-v2:0 in AWS us-east-2, improving performance and refining logic using FastAPI. Contributing to caddy-scraper integration with DynamoDB and PostgreSQL for optimized data integration and retrieval processes.
- Developing scalable deployment pipelines using Docker and AWS technologies, while documenting APIs with Swagger to enhance usability and debugging. Assisting in streamlining data storage and retrieval processes through integration of various tools and technologies.

INFOSYS • Bangalore, India **May 2021 - Aug 2022**

SYSTEMS ENGINEER

- Architected an LSTM-based time series forecasting model for demand planning. Implemented feature engineering and hyperparameter tuning, enhancing forecast model accuracy by 15%. This optimization led to on-time release of new medical device versions by accurately predicting OS release dates.
- Established quality assurance machine learning practices to address inconsistent data analysis approaches and lack of AI adoption within the team. Introduced best practices for feature engineering, model validation, and structured/unstructured data analysis. This initiative resulted in a 25% increase in AI-driven process improvements.
- Tested medical devices for script accuracy and functionality. This rigorous approach resulted in a significant reduction in analysis time, enhancing overall testing efficiency and device reliability while ensuring compliance with safety standards.

VARUN ENTERPRISES • Bangalore, India **Sep 2020 - Apr 2021**

DATA SCIENTIST

- Spearheaded CNN-based models for manufacturing anomaly detection, addressing high false-positive rates in defect identification. Enhanced defect detection accuracy, leading to 12% revenue increase through improved product quality and reduced waste.
- Deployed deep learning models using Flask on AWS EC2. This resulted in a 50% reduction in false positives, ensuring scalability and reliability in the manufacturing process.

Projects

Multimodal LLM-powered Medical Image Analysis [GitHub] **Jan 2024 - Sep 2024**

- Engineered a question-answering diagnostic tool for brain MRI analysis, collaborated with medical professional to address time-consuming manual image interpretation challenges. This innovation enabled automated insights from uploaded brain MRI images, potentially reducing analysis time by 40%.
- Optimized medical image analysis by fine-tuning three multimodal LLMs (IDEFICS2, LlaVA 1.5 7B, and LlaVA-Med). LLaVA-

Med achieved the highest performance, enhancing diagnostic accuracy by implementing telemetry to monitor usage patterns.

- Developed an interactive Gradio-based interface with model selection, image upload, and query input via text or speech. Implemented adjustable parameters for tailored responses, improving user experience and potentially reducing diagnostic decision-making time by 30%.

End-to-End MLOps Project with ETL Pipeline- Network Security System [GitHub] **Jun 2024 - Nov 2024**

- Architected ETL pipelines using MongoDB Atlas to address data quality issues in network security systems. Implemented robust error handling and validation, reducing data processing errors by an estimated 20% and enhancing threat detection reliability.
- Optimized network threat detection models through hyperparameter tuning and MLFlow experiment tracking. Leveraged Daghub for version control, improving model performance and potentially increasing threat detection accuracy by 35%.
- Streamlined deployment processes using Docker and GitHub Actions, managing artifacts with AWS S3 and ECR. Deployed the solution on AWS EC2, reducing deployment time and enhancing real-time threat detection capabilities.

Education

UNIVERSITY OF MASSACHUSETTS, DARTMOUTH **Aug 2022 - Sep 2024**
Master's in Data Science GPA 3.97 / 4

NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY, Bangalore, India **Aug 2014 - Jul 2018**
Bachelor of Engineering, Electronics and Communication GPA 3.1 / 4