

Girish Manjunatha Saraswathipura

gsaraswathipura@gmail.com | +1 (703)-400-9756 | [LinkedIn](#) | [GitHub](#) | Bloomington IN

Education

Master of Science in Data Science - ML, Data Mining, Advance Database, AI Research, Applied Algorithm

08/2024 - 04/2026

Indiana University Bloomington, USA

Bachelor of Engineering in Computer Science - Mathematics, DevOps, Software Engineering, MongoDB

08/2018 - 06/2022

Visvesvaraya Technological University, India

Work Experience

IU Innovates

06/2025 - 08/2025

Machine Learning Research Intern - Deep Learning, HPC, Fine Tune LLM, Statistics, Data Profiling, CNN, ANN, GPU, CUDA

- Researched how LLMs set task boundaries using system prompts, improving controlled response accuracy by 25%.
- Explored LLM meta-learning to enable quick adaptation across tasks, cutting model retraining needs by 30%.
- Built an AutoML tool that shrinks model search space, achieving 40% lower cost and 60% faster convergence during benchmarking.
- Applied advanced statistical and approximation methods to extract metadata from 50M+ records, cutting profiling time by 45%.
- Automated preprocessing and EDA workflows, reducing manual data exploration effort by 30% and accelerating model readiness.
- Developed a deep learning architecture generator in TensorFlow & PyTorch for image and tabular data, improving convergence with 35% fewer iterations.
- Leveraged NVIDIA A100 GPUs with CUDA and multithreaded processing to optimize training speed and computational efficiency by 2x.

Informatica Inc.

01/2024 - 07/2024

Modernization Associate Consultant - ETL optimization, Data Governance, Data Quality, Synapse SQL, AWS EC2, Azure, Snowflake

- Built a Python automation tool on AWS EC2 that reduced certificate generation time from 10 hrs to 3 hrs, saving \$4K+ monthly.
- Consulted Fortune 500 companies to migrate 370+ ETL workflows (8% of enterprise pipelines) to Informatica Cloud, accelerating adoption in 6 months and earning an MVP award for impact on Volvo and Charles Schwab projects.

Informatica Inc.

08/2022 - 12/2023

Migration Factory Associate Specialist - Data Engineering, Data Warehouse, Data Wrangling, Oracle, PySpark, Hadoop

- Migrated 1,100+ pipelines from PowerCenter to IICS for Volvo's master database, boosting workflow efficiency by 21% and ensuring 98%+ reliability on 20TB monthly data.
 - Automated lineage checks for 300+ dimensions and profiled streaming data across Snowflake, Hadoop, Spark, Oracle, AWS Redshift, and Azure Synapse, completing a major migration 7 weeks ahead of schedule.
-

Projects

Agentic AI Stock Portfolio Assistant - API Integration, Parallel Processing, Agentic AI, Prompt Engineering, GenAI

- Engineered an LLM-powered portfolio agent by integrating Perplexity API with smart routing and Alpaca API to automate stock actions, and implemented multithreaded monitoring to track portfolios and send alerts on major news with less than 5 seconds latency.

Traffic Sign Classification on AWS SageMaker - MLOps, Model Deployment, Neural Networks, AWS

- Built and deployed a CNN (LeNet) on AWS SageMaker with automated preprocessing, achieving 95% accuracy and processing 100K+ predictions/day at 60% faster runtime and 2/3 cloud cost.

Audio Signal Classification - TensorFlow, Class Imbalance Handling, Data Augmentation, CNN, Signal Processing, Feature Extraction

- Built an audio signal classification model on 800+ files (217 positive, 593 negative) using spectrogram features and CNNs in PyTorch, achieving 90%+ accuracy with class imbalance handling.

Fine-Tuning LLM on Environmental Protection Data - Hugging Face Transformers, Tokenization & Embeddings, Transfer Learning

- Collected and transformed Indiana environmental blogs and government records into Q&A datasets, enabling Hugging Face fine-tuning of LLaMA-8B with 15% faster convergence and improved model adaptability.
-

Skills

Programming & Data: Python, C++, C, R, SQL, PostgreSQL, Synapse SQL, MongoDB, FAISS, NeoJS, Hadoop, PySpark, AWS S3, Snowflake

ML Engineering & MLOps: Feature Engineering, Feature Selection, Hyperparameter Tuning, Cross-Validation, Regularization, Ensemble Learning, Grid Search, Model Monitoring, Drift Detection, CI/CD for ML, Reproducibility, Versioning, MLflow, Experiment Tracking, NLP, PyTorch, Scikit-learn, TensorFlow, Transformers, RAG, LangChain, Foundation Models, Agentic AI, Claude API, CNN, Backpropagation

Data & Cloud Platforms: Databricks, Informatica, AWS (Bedrock, Glue, SageMaker), Docker, Kubernetes, ETL Pipelines, Data Governance

Mathematics & Statistics: Linear Algebra, Probability & Statistics, Calculus, gradients, Numerical Optimization, Information Theory

Certifications and Credentials: IBM Data Science Professional, AWS AI Practitioner, Informatica Certified Data Integration Professional, DeepLearning Certified TensorFlow Developer, Intel AI Specialization, MongoDB University M001, Informatica Cloud Application Integration