VENKAT SRINIVASA RAGHAVAN

srinivasaraghavan.v@northeastern.edu | (206) 518-8678 | LinkedIn | GitHub | Portfolio

SUMMARY

Master's in Data Science with 3+ years of experience designing end-to-end ML pipelines, deploying models at scale, and building cloud-native, data-driven applications for real-time, impactful business solutions across diverse domains.

EDUCATION

Master of Science in Data Science - Northeastern University | 4.0/4.0 **Jan 2023 - Apr 2025**

Bachelors in Electronics & Instrumentation - R.V College of Engineering | 8.69/10 **Aug 2016–Aug 2020**

Relevant Coursework: Cloud Computing, Algorithms, Machine Learning, Large Language Models, Data Management.

PROFESSIONAL EXPERIENCE

Data Scientist, Fidelity Investments | *Boston, MA* **Jan 2024 - Jul 2024**

Skills—System Design, ETL Pipelines, LLMs (RAG), Microservices, MLOps, NER, AWS (SageMaker, Lambda, EC2), Kubernetes, CI/CD.

- Reduced data processing time from 190 to 18 hours by designing a **Kafka**-based ETL system and deploying a stock trend forecasting pipeline using Transformer models, **MLflow**, **S3**, and **Snowflake**.
- Built a sentiment analytics engine using **TF-IDF**, **LDA**, and **BERT** with **CI/CD** and **containerized** workflows, cutting processing time by **90%**. Created **S3** to **Snowflake** data pipelines using **AWS Lambda**, boosting investment decision speed by 60%.
- Built a GPT-4-powered RAG system on EC2 using **ChromaDB**, **Elasticsearch**, **GraphQL**, and **SageMaker** for **NER** and deployed a conversational AI chatbot on top, achieving **99.99%** uptime and significantly boosting accuracy.

Data Science Engineer, Airbus | *Bengaluru, India* **Jan 2021–Dec 2022**

Skills—AWS (Redshift, SageMaker, MWAA), Databricks, Airflow, TensorFlow, ETL Pipelines, MLOps, SQL Optimization

- Built a pipeline for validating time series data using **Prophet** and **z-scores**, orchestrated via Amazon **MWAA** (Managed **Airflow**) with integrated drift detection and auto-retraining on **SageMaker**, reducing production error rates by **95%**.
- Engineered multivariate anomaly detection pipelines using **TensorFlow** (CNNs & LSTMs), developed in **Databricks** and deployed as scalable distributed workflows on **EKS**, achieving **92%** accuracy and enhancing system robustness and reliability.
- Developed modular **ETL** pipelines using **Airflow** and **Databricks** to ingest and transform high-volume IoT data into **Redshift**, reducing latency by **80%** and enabling near **real-time analytics** across business units.
- Re-architected SQL Server models and **ER diagrams**, then migrated and optimized them on **Redshift** with partitioning and distribution strategies, halving query times and improving pipeline scalability by **70%**.

Associate Data Engineer, Airbus | *Bengaluru, India* **Jun 2020–Dec 2020**

Skills—SQL, Node.js, Angular, AWS (RDS, Athena, Redshift), data processing, and CI/CD.

- Optimized SQL procedures and indexing for large datasets, reducing runtime from **12 hours to 4 hours** and enhancing data retrieval. Developed a forecasting tool using **AWS RDS** (**SQL Server**), **Athena**, and **Redshift**, boosting efficiency by **65%** and **doubling** users.
- Designed an interactive forecasting dashboard using **Power BI** (with data from Redshift and RDS), enabling trend analysis and KPI tracking; improved decision-making efficiency by 65% and doubled stakeholder engagement.

Machine Learning Researcher, R.V College of Engineering | *Bengaluru, India* **Aug 2019 - Aug 2020**

Skills—Python, Javascript, Deep Learning, Machine Learning, IOT, and Github.

- Developed an IoT-based spine diagnostic system at BGS Global Hospital to detect lumbar lordosis using sensor instrumentation, capturing posture data in real-time via a **JavaScript–Python** stack. Patented our study: **No. 202241054225 A.**
- Built and trained logistic regression and ANN models for spine condition classification, achieving an F1 score of **0.98** and reducing clinical diagnosis time from **days to minutes**.

PROJECTS

Nutribuddy : *Transformers, FastAPI, PostgreSQL, PyTorch, React, Docker.* **Apr 2025**

Developed a personalized nutrition and exercise recommendation system using **LLMs** to parse user history and biometric data; deployed to 22 beta users, delivering highly tailored outputs and receiving positive early feedback.

Codesage : *Mistral-7B, FastAPI, React, GitHub Actions, CI/CD, PyTorch.* **Jan 2025**

Built an AI-powered code review system using **Mistral-7B,** optimized fine-tuning on **multi-GPU infrastructure** for pull request analysis, and integrated it into the **CI pipeline with GitHub Actions**, reducing post-merge code errors by **80%** and improving review turnaround time.

TECHNICAL SKILLS

Programming & Development: Python, SQL, Java, AWS (S3, EC2, Lambda), GCP (BigQuery, Vertex AI), R, Apache Spark, Apache Hadoop.

MLOps & Deployment: MLflow, Kubeflow, AWS SageMaker, Airflow, TorchServe, FastAPI, Flask, Docker, Kubernetes, TFX.

Machine Learning: PyTorch, TensorFlow, Scikit-learn, Numpy, OpenAI APIs, LangChain, HuggingFace, DDP, Keras, Bayesian Inference.

Competitive Programming: Kaggle (3rd/50+ teams, 2023 Datathon, ML credit line model); Leetcode Knight, AWS Graduate Certifications.