Venkat Srinivasa Raghavan

Boston, MA 206-518-8678 srinivasaraghavan.v@northeastern.edu Linkedin GitHub

EDUCATION

Northeastern University **Jan 2023 - April 2025**

Master of Science in Data Science, GPA: 4.0/4.0 *Boston, MA*

Coursework: Algorithms, Machine Learning, Data Mining, Deep Learning, LLMs, Cloud Computing.

R.V College of Engineering **Aug 2016 – Aug 2020**

Bachelor of Engineering in Electronics and Instrumentation *Bengaluru, India*

Coursework: Coursework: OOPS, Graph theory, Java, Algorithms, Discrete Mathematics.

SKILLS

Languages & Frameworks: Python, R, SQL, PyTorch, TensorFlow, Scikit-learn, Keras, Pandas, NumPy, SciPy.
Systems & APIs: Airflow, Snowflake, MongoDB, PySpark, Kafka, AWS (SQS, SageMaker), MLFlow, Docker, Git.
Cloud & Databases: AWS (S3, EC2, SQS), GCP (BigQuery, Vertex AI), Kubernetes, CI/CD, Time Series Analysis.

EXPERIENCE

Fidelity Investments **Jan 2024 – Jul 2024**

Software Data Engineer *Boston, MA*

- **Tactical Data Capturing Team**: Led efforts in web scraping, data transformation, data loading, and managing ETL workflows using **Airflow and AutoSys** while enhancing data quality using machine learning models.
- Optimized investment decision-making by reducing data processing time from 190 to 18 hours through an advanced data transformation algorithm and a **multithreaded** information retrieval system, using Python, **FastAPI**, and **AWS**.
- Developed a sentiment analysis pipeline using **FinBERT**, **PyTorch**, and **MLflow** to predict stock trends, enabling informed investment decisions by structuring data into a tabular format with sentiment scores for efficient analysis.
- Built a GPT-4 RAG-based system delivering **48** half-hourly notifications to investment officials, enabling real-time data tracking with millisecond precision. Deployed on AWS with **EKS**, **Snowflake**, and **ChromaDB**.
- Improved data quality by **75%** and operational efficiency by **80%** by developing a time series data validation pipeline using z-scores and Exponential Moving Average (EMA).

Airbus **June 2020 – Dec 2022**

Full Stack Software Engineer *Bengaluru, India*

- **Methods and Tools:** Managed aircraft system software and built tools to improve operability and efficiency. Developed analytics software for hardware data and enhanced flight testing through Agile application development.
- Reduced processing time by **55%** by creating **MS SQL stored procedures**, optimizing runtime through data indexing, and building views and functions to enhance the aircraft hardware forecasting system.
- Improved business efficiency by **70%** by using **SSIS** and **C#** to automate data integration and streamline operations for managing the aircraft hardware system.
- Reduced system latency by **80%** by designing data models using **ER diagrams** to optimize large-scale data handling and support simultaneous usage with JWT user authentication. Deployed the solution on OpenShift using CI/CD pipelines.

PROJECTS

Multi-Modal Content Moderation System | *NLP, CV, MLFlow, PyTorch, AWS* **May 2024 – Oct 2024**

- Built a scalable content moderation system using with computer vision and NLP to analyze text and video content.
- Leveraged PyTorch and MLflow for model tuning and deployment, ensuring efficiency and scalability.

CodeSage | *LLM's, Microservices, Python, FastAPI, GPU, Quantization, NLP* **Sept 2024 – Dec 2024**

- Developed an automated code review system using fine-tuned LLMs, achieving a ROUGE score of 0.68.
- Integrated the system into CI pipelines via GitHub Actions, enhancing performance with Quantization and LoRA.

Big Data Recommendation System | *Hadoop, Pig, SQOOP, Hive, React* **Aug 2024 – Nov 2024**

- Designed and implemented recommendation models using user-user and user-item collaborative filtering for insights.
- Enhanced user engagement through hybrid recommendation systems powered by Hadoop.

LEADERSHIP & AWARDS

- Secured **3rd place** out of 50+ teams at Kaggle's 2023 Datathon by developing an ML model to predict credit lines.
- Ranked in the **top 400** of 25000 participants in the November LeetCode Competitive programming championships.
- Three time **champion** with Northeastern University in the National Badminton Championships.
- Certifications: Hackerrank – SWE, Problem Solving (Basic to Advanced), SQL (Basic to Advanced).