Venkat Srinivasa Raghavan

Email: srinivasaraghavan.v@northeastern.edu | Phone no: +1 206-518-8678 | LinkedIn: Venkat Srinivasa Raghavan | GitHub Repository

EDUCATION

Northeastern University, Boston, MA — January 2023 - Present
Khoury College of Computer Sciences — Expected Graduation: December 2024
Master of Science in Data Science — GPA: 4.0/4.0
Coursework: Algorithms, Supervised Machine Learning, Data Mining techniques, Deep Learning.
Lead Player of the NEU badminton team. (**Silver Medalist** – Inter-zonal)

R.V College of Engineering, Bengaluru, India — September 2020
Bachelor of Engineering in Electronics and Instrumentation — GPA: 8.69/10
Coursework: OOPS with C++, Graph theory, Java programming, Algorithms, Discrete Mathematics.

SKILLS

Python, R, Java, GO, C#, SQL, Pytorch, Tensorflow, Scrapy, BeautifulSoup | Numpy, Pandas, Web scraping, ETL pipelines, DBMS, VectorDB, Snowflake, MongoDB, Kafka | Supervised ML and Data Mining, Deep Learning, NLP, GAN, Computer Vision, Large Language Models | Microservices, Distributed Systems, Angular, AWS, .NET core, NoSQL, System Design.| Jenkins Core, CI/CD, Git, Docker, Kubernetes, Airflow, Ansible, MLOps | Apache Spark, PySpark, MapReduce, Pandas, Big data Analysis.

PROFESSIONAL EXPERIENCE

Fidelity Investments, *Software Data Engineer - Adv. Strategies and Research Technologies* — January 2024 – Present | Boston, USA

- **Web Capturing Team**: Performed tactical data collection, processing, research, and validation using cloud environments.
  - Designed and developed a **microservices** based end-to-end framework using **Azure API's, Python, BeautifulSoup** with a **multithreaded** generic algorithm for data **extraction** and **analysis**, boosting the processing speed by **6x**.
  - Implemented a **distributed** data workflow management system using **AWS SQS** to manage the data flow between **AWS S3** and a **Snowflake database** enhancing the efficiency of data management and analytics.
  - Created an **Angular**-based UI for the framework facilitating easy automation and data pipeline management.
  - Led a 4-member team to **research** and apply **innovative data retrieval and processing techniques**, **optimizing** system architecture for advanced analysis, resulting in a 72% improvement in business efficiency and 40% reduction in cost.
- **Text Analysis Team**: Developed an Asset management-based Question Answering System.
  - Developed a scalable question answering system using **Vector databases**, **AWS SQS**, **Large Language Models**, and **NLP techniques** achieving data handling capabilities of up to 20 billion records.

Airbus, *Software Engineer L2 - Methods and Tools* — June 2020 - December 2022 | Bengaluru, India

- **Interface Point (IP) Data Center**: Worked in a team focused on aircraft configuration data engineering and analysis.
  - Designed **SQL procedures** and **SSIS** pipelines for aircraft data retrieval and KPI analysis, boosting the performance by 80%, earning a "**Spot Award**". Mentored new employees towards the governance of the IP Data Center.
- **Aircraft Test Manual Automation Team**: Automated and optimized test manual workflows ensuring quality standards.
  - Developed a document editor app using **C#.NET** following **microservices** architecture, using Entity Framework Core and **SQL Server**. Deployed on a **Kubernetes** platform named **OpenShift** using **Docker** and **CI/CD pipelines.**
  - Achieved a 95% improvement in handling large datasets, enhancing application scalability and robustness.

PROJECTS

AI Powered News Research Tool using LangChain. — January 2023 - February 2023
Developed an end-to-end news research tool with **LangChain**, **LLMs**, **OpenAI API**, and **Streamlit** for advanced financial Analysis.

Waste Object Detection using YOLOv5. — November 2023 – Februray 2023
Developed a YOLOv5 waste detection model with **PyTorch**. Deployed with **DVC**, **Docker** and **CI/CD** pipelines on **AWS** and **Azure**.

Text Summarization using Transformers. — December 2023 – February 2024
Developed a Text Summarization tool using **HuggingFace** API's, **Python**. Created **ML Pipelines** and deployed them on **AWS**.

Reddit Data Pipeline via Airflow. — January 2024 - February 2024
Developed a Reddit data Pipeline using **Airflow**, **Postgres**, **AWS Athena**, **Redshift** for ETL process and visualization in **Tableau**.

ACCOMPLISHMENTS

- Secured **the 3rd position in Kaggle's Summer 2023 DataThon**, a global data science hackathon.
- Developed a Data science-based injury detection method, published at **IEEE EMBC**, with a patent filed **(202241054225).**