Vamshi Saggurthi

P: +1 2016871827 | vamshisaggurthi@outlook.com | [LinkedIn](#) | [Github](#) | [vamshisaggurthi.com](#)

EDUCATION

Rutgers University, New Brunswick | Master's in Computer Science **New Jersey, NJ**
Coursework: Machine Learning,Design and Analysis of Algorithms, Distributed Systems,Data Visualization **Aug 2021-May 2023**

Osmania University | Bachelor's **Hyderabad, India**
Coursework: Data Structures and Algorithms, Operating Systems,Computer Networks,Database Systems, **Oct 2013 - Jun 2017**

Publications

TR-2-PATH and MYC as a marker of Enza resistance Nature Communications impact score **17.69** **2022**

Skills

Languages: Python, Java, R, JavaScript
Libraries/Frameworks: ReactJS, Redux, Firebase, Pandas, Dash, Airflow, Kafka, pySpark, TensorFlow, Sci-kit Learn, PyTorch,
Databases: MySQL, Postgresql, MongoDB
MISC: Git, Docker, Kubernetes, RESTful, Jenkins CI/CD, AWS

WORK EXPERIENCE

Striim Inc | Senior Software Engineer **July 2023 – Present**
- *Real Time PII Detection*: Researched ,Designed, implemented ,real time PII detection while streaming data at scale 3 million events/per hr , on a single GPU
- Lead introduction of Micro Services with gRPC in monolith applications of on-prem deployment and also scaled it cloud services
- *Striim Copilot* :Built Striim Copilot Agent with fine tuned OpenAI GPT solving customer issues , and automating the pipeline creations , for real time streaming.
- *Real time Encryption*: Designed ,Developed cloud kms based envelope encryption in real time streaming systems , enhanced the open source sdk's to match the high throughput demands (100k events /sec)
- *Observability*: Built POC to move from custom telemetry system to OTEL backed for better observability , moving from legacy lo4j log debugging .

Research Assistant | Big Data **Jan 2021 –May 2023**
- *Data Engineering*: Designed and built a Docker-based Apache Airflow data pipelines over HPC slurm with custom deferrable and trigger operator for ingesting datasets and performing mining
- *Network Mining* : Built an algorithm that integrates biological mechanisms to reconstruct a network and then mines to interrogate the network , improving their computational performance by 400%

Striim Inc | SDE Intern **Jun 2022 – Aug 2022**
- *Real Time streaming Analytics:* Reduced **37%** bugs , by building end to end analytics application which includes data scraping, **ETL**, and web interface for visualization, which formed as the baseline data for prioritizing development.
- *Library Diagnosis*: Developed library diagnosis module for analyzing jar dependencies installed on customer deployment.
- Built multiple ETL pipelines tailored for customers to tackle specific tasks demonstrating the real time analytics of the data

Dassault Systemes Inc | Site Reliability Engineer **Aug 2020 – Jul 2021**
- Developed & maintained cloud microservices for COVID research and clinical trials supporting Moderna and other institutes
- *Custom Observability Tooling*: Designed and Developed custom tooling for observability, monitoring and metrics reporting for advanced analytics.
- Built Predictive alert systems which reduced scaling failures by **15%**
- *MicroService Orchestration*: Transformed kubernetes based custom tooling for local microservices orchestration boosting developer productivity by 40%.
- Resolved scalability bottlenecks for backend services in AWS which improved service performance by **10%**

Hexagon Inc | Senior Software Development Engineer **Aug 2017 – Aug 2020**
- *Smart Form Generator AI*: Developed complex cloud based web applications for mechanical engineers' design workflows.
- Optimized data flow in the application by building custom middleware using React-Redux ,boosting load time by 30%.
- Lead "edge computing feature" improving performance by 40% in mobile browsers and devices with limited compute power.
- Engineered and implemented proof of concept for Computer Vision based text extraction from instrument specification images.

Academic Projects

[Criminalytics](#)   **Jan-May-2021**

- Formulated an ETL pipeline that periodically collects crime data for the last 15 years for the city of Chicago.
- Designed and built geographical dashboard for chicago crime data analysis (2012-present) consisting over 7 million data points.
- Implemented high performant api queries that allows users to bring out statistical outcomes on the fly for over 7 million records in under 1.5-sec using out of core dataframes.

[Maze Solver](#)   **Sep-Dec 2020**

- Designed multiple AI agents using Repeated A*, Inference, and Bayesian Networks for optimally searching a hidden target within a maze using Python and NumPy
- Optimized these agents to find the targets in (101) *(101) dimension mazes under 20 milliseconds
- Built a CNN with Dense layers using PyTorch to imitate these agents obtaining accuracy of 92% in solving the mazes.

[Image -to -Image Translation](#) **(GAN's)**   **Jan-May 2021**

- Explored Supervised I2I using Pix2Pix GAN to translate Street View Images to Aerial View Images and vice-versa
- Implemented CycleGAN framework for the task of translating Real Pizza to Synthetic and Live Pizza Image Domains
- Researched the drawbacks of CycleGAN framework and proposed an enhanced CycleGAN by 10% with reduced artifacts