

Sai Charan Somineni

Fairfax, VA | Open to Relocate

📞 +1-832-931-1949 | ✉️ ssominen@gmu.edu | 🔗 linkedin.com/in/sominenisaicharan | 🌐 github.com/saicharansom

EDUCATION

George Mason University

M.S. in Data Analytics Engineering, GPA: 3.8/4.0

Aug 2023 – May 2025

Fairfax, VA

SRM Institute of Science and Technology

B.Tech in Computer Science and Engineering (First Class with Distinction)

Jun 2018 – May 2022

Chennai, India

TECHNICAL SKILLS

Languages: Python, SQL, R, Scala, C++, Java, JavaScript

ML/AI: Supervised/Unsupervised Learning, GPT, BERT, RAG, LLM Finetuning, Transformers, NLP, Machine Learning Algorithms, Statistical Methods, A/B Testing, Data Mining, Advanced Analytics

Frameworks & Packages: Scikit-learn, TensorFlow, Keras, PyTorch, FAISS, spaCy, NumPy, Pandas, GeoPandas, Matplotlib, Seaborn, SciPy, Plotly

Cloud/Data Platforms: AWS (Redshift, S3, Neptune), GCP (BigQuery), Azure, PySpark, Databricks, Apache Spark, Apache Hadoop, Apache Kafka, Apache Beam, Data Lake, Delta Lake

Databases: MySQL, PostgreSQL, MongoDB, AWS Neptune, Neo4j, GraphQL, Snowflake

ETL & Data Tools: Airflow, Glue, Informatica, Talend, SSIS, SSRS, SSAS, ETL/ELT, Data Transformation, Data Warehousing, Data Quality and Governance

DevOps & CI/CD: Docker, Kubernetes, Jenkins, Terraform, Git, GitHub

Visualization & BI: Tableau, Power BI, Looker Studio, Advanced Excel (Pivot Tables, VLOOKUP)

IDEs & Tools: VS Code, PyCharm, Jupyter Notebook, IntelliJ, OpenRefine, Streamlit, Figma, Nimble

Methodologies: SDLC, Agile, Waterfall

Operating Systems: Windows, Linux, macOS

PROFESSIONAL EXPERIENCE

Data Engineer – Intern

Tanvi IT Solutions, Inc.

Sep 2025 – Present

Remote

- Assisted in designing and maintaining ETL pipelines, improving data processing efficiency by 25% across multiple internal projects.
- Cleaned, validated, and transformed 50K+ rows of raw data into analytics-ready datasets used for reporting and dashboarding.
- Monitored and optimized pipeline runs, reducing data delivery delays by 15% while ensuring accuracy and consistency.
- Collaborated with senior engineers to debug and resolve workflow issues, cutting average resolution time from 2 hours to 30 minutes.
- Documented ETL workflows and pipeline logic, creating a reusable knowledge base that decreased onboarding time for new interns by 40%.

Data Analysis & Visualization Intern

Pearl Matibe — International Security & Counterterrorism

Jun 2025 – August 2025

Washington, DC

- Cleaned, unified, and enriched multi-source datasets (OSINT, government releases, news crawls) covering 100+ security events across 50+ African states using Apache Spark jobs; automated ingestion pipelines with Airflow and Glue to deliver analysis-ready tables overnight.
- Built interactive geospatial dashboards and story maps that surface diplomatic travel patterns, aid flows, and crisis flashpoints; outputs feed directly into Matibe's columns and policy briefings.
- Co-authored data-driven investigative pieces, translating statistical findings into plain-language narratives for public audiences and aligning every visual with editorial goals.

Research Assistant

CAPS Lab, George Mason University

Jan 2025 – May 2025

Fairfax, VA

- Conducted NLP-driven literature synthesis using NLTK/spaCy over 250+ research papers on adaptive authentication.
- Modeled usability-privacy-security trade-offs with Scikit-learn, informing secure design prototypes.
- Researched trends in infrastructure security and adversarial AI, shaping strategic lab publications.

Business Analyst

The Detailing Mafia

Aug 2022 – Jun 2023

Velachery, Chennai, India

- Developed full-stack Power BI dashboards with SQL and Apache Spark pipelines, delivering real-time insights across 25+ franchise locations and reducing manual reporting time by 75%.
- Designed optimized database schemas and API-driven integrations for POS, inventory, and labor systems, enabling automated hourly refreshes and improving stock accuracy by 15%.
- Applied machine learning (Python, scikit-learn, Spark SQL) to uncover product-affinity clusters, driving targeted upsell campaigns that boosted accessory conversions by 12%.

Implementation Consultant

Digite Infotech Pvt Ltd

Jan 2022 – Jul 2022

Remote

- Migrated four enterprise portfolios from spreadsheet-based waterfall tracking to Nimble's AI-powered Kanban/Gantt suite, raising on-time delivery to 93% and eliminating 400+ spreadsheets.
- Built secure Excel and REST connectors that stream live work-item data into Nimble dashboards, giving 250+ users real-time burndown and capacity views.
- Guided cross-functional teams through agile adoption, resource planning, and AI-driven backlog prioritization, boosting throughput by 40% while keeping oversight in a single hub.

Research Intern	Oct 2020 – Dec 2020
<i>SPIHER Lab, SRM University</i>	<i>Remote</i>
<ul style="list-style-type: none"> – Built a personalized cross-modal recommendation system with DLRM, improving engagement by 25%. – Integrated MongoDB with a Keras-trained backend for Bookworm-Innovations and conducted A/B testing for UX enhancements. 	

PROJECTS	
Student Survey Application <i>Spring Boot, Docker, Kubernetes, Jenkins, AWS EC2, AWS RDS, Rancher</i>	Mar 2025 – May 2025
<ul style="list-style-type: none"> – Built a production-ready Spring Boot REST API with Spring Data JPA and strict field validation (@NotNull, @NotBlank, @Email) to ensure strong data integrity during the pilot phase. – Provisioned Amazon RDS MySQL 8 and tuned HikariCP (maxPool=20, minIdle=5) for low-latency, high-throughput reads/writes under concurrent load. – Containerized the service into a slim OpenJDK 23 image, deployed a 3-replica Kubernetes workload on AWS EC2 via Rancher (NodePort), achieving 99.9% uptime. – Implemented a Git-triggered Jenkins pipeline (Maven build → Docker image → K8s rollout with health checks) for zero-downtime deployments in under 5 minutes. – Created Postman test suites and validated live production rollouts, successfully processing 10+ submissions without validation or runtime errors. 	
Taskify – Task Management Web App <i>Vite + React.js, Express.js, Node.js, MongoDB, Docker</i>	Jan 2025 – Feb 2025
<ul style="list-style-type: none"> – Developed a secure, role-based task management system with guarded routes and full CRUD workflows, enabling end-to-end team task tracking. – Implemented JWT authentication with bcrypt hashing to strengthen access control and reduce login issues by 25%. – Added real-time filtering, search, and sorting (20+ combinations) with optimized MongoDB indexes, reducing task discovery time by 15%. – Containerized the entire stack with Docker and Compose for predictable, cross-platform development and deployment. – Designed clear REST endpoints to handle concurrent edits smoothly, ensuring a consistent and responsive UI. 	
Augment Linkage: Evidence & Exploration <i>AWS Neptune, Neo4j, Gremlin, Python, SageMaker, Airflow, Glue</i>	Jan 2025 – May 2025
<ul style="list-style-type: none"> – Designed and deployed a scalable graph database for TraCCC, mapping 10k+ nodes and 15k+ relationships to uncover illicit fentanyl supply networks using AWS Neptune and Neo4j. – Automated evidence-weighted linkage algorithms and PII tokenization workflows, boosting network detection by 42% and reducing false positives across law-enforcement datasets. – Migrated Neo4j models to Neptune; built Gremlin queries, SageMaker inference hooks, and Airflow-Glue pipelines for five agencies cutting manual investigation time by 60%. 	
PrivAware Voice Assistant <i>GPT-3.5 Turbo, Streamlit, Python</i>	Sep 2024 – Dec 2024
<ul style="list-style-type: none"> – Built and fine-tuned a privacy-preserving voice assistant with real-time semantic filtering and role-based admin control. – Integrated RAG to raise response accuracy by 35% while ensuring ethical data handling and reduced latency by 28% in pilot trials. 	
Domain-Specific QA for Pavement NDE <i>BERT, GPT-2, FAISS</i>	Aug 2024 – Nov 2024
<ul style="list-style-type: none"> – Created a technical QA engine for the Federal Highway Administration using fine-tuned LLMs. – Curated and embedded pavement-engineering datasets with FAISS for real-time retrieval, achieving 86% answer accuracy on domain queries. 	
ArXivIntelliRAG – Academic Research Assistant <i>Python, FAISS, MiniLM, SciBERT, LangChain, LLaMA 3.2</i>	Aug 2024 – Nov 2024
<ul style="list-style-type: none"> – Developed a RAG pipeline using Python, FAISS, and MiniLM to enable semantic search across 10K+ arXiv papers, increasing contextual retrieval accuracy and streamlining research workflows. – Integrated SciBERT-based ranking and LangChain-driven summarization with LLaMA 3.2, enhancing top-3 relevance by 40% and reducing manual analysis effort for domain-specific literature reviews. 	