VINAY RAM GAZULA

☎ 908-552-1879 | ✉ gazulavinayram@gmail.com | 🌐 vinaygazula.dev | in linkedin.com/in/vinayramg

PROFESSIONAL SUMMARY

Data Engineer with over 3 years of experience designing scalable data pipelines, data warehouses using AWS, Apache Airflow, Apache Spark, Databricks, and Snowflake. Proficient in advanced data modeling techniques, lambda and kappa architectures, open table formats (Apache Iceberg, Delta Lake). Effective at collaborating with data scientists, analysts, and cross-functional teams. Strong research background in data science, with multiple publications focussing on AI/ML.

EXPERIENCE

New Jersey Institute of Technology **Sept 2023 — Present**
Research Assistant | NJIT Engineering Education Research *Newark, NJ*

- Analyzed institutional student data with machine learning and developed regression models (Multiple Linear Regression, Random Forest, XGBoost, LightGBM) in R to predict GPA and identify academic performance factors

Teaching Assistant | CS-634 Data Mining Course

- Evaluated weekly homework, quizzes, and projects with prompt, equitable feedback, while mentoring students in Python coding to enhance their project outcomes

Research Assistant | Data and Knowledge Engineering Lab

- Integrated xAI tools (LIME, SHAP, Anchors, PDP, ALE) into "SolarFlareNet"—a transformer framework for space weather research—successfully re-trained the model to maintain 90.7% accuracy, published findings at FLAIRS & ICTAI

Impetus **July 2020 — Aug 2023**
Data Engineer *Bengaluru, India*

- Achieved 30% cost reduction in Amazon S3 by leveraging Parquet snappy compression and optimizing life cycle policies
- Collaborated with stakeholders to design data models for KPI generation and efficiently handled 10+ ad hoc requests, ensuring alignment with business needs
- Developed idempotent ETL pipelines to ingest and process terabytes of data using Apache Spark, cutting processing times by 40% for faster business insights
- Utilized AWS MWAA to orchestrate complex data workflows with data quality checks, enhancing pipeline reliability and operational efficiency
- Leveraged AWS Glue Data Catalog and AWS Lake Formation to standardize metadata management and enforce data governance policies, reducing integration complexities and accelerating data discovery

TECHNICAL SKILLS

Languages :	**Python** (PySpark, Polars, Pandas, NumPy, scikit-learn, PyTorch, TensorFlow), **SQL** (ANSI SQL, PL/SQL), **Scala** (Apache Spark), Go, Rust, Bash, R, C
Databases :	**PostgreSQL**, **MySQL**, **Oracle**, **MongoDB**, MS SQL
Cloud :	**AWS** (Athena, EC2, Glue, Lambda, RDS, Redshift), Azure (Data Factory, Synapse Analytics), GCP (BigQuery)
Big Data :	**Apache Spark** (Spark SQL, Dataset API, Dataframe API, Streaming), **Databricks**, Apache (**Iceberg**, **Flink**, **Kafka**, **Airflow**, Hive), **DBT**, Snowflake
Analytics :	**Tableau**, Apache Superset, Power BI, Excel
CI/CD :	Git, **GitHub**, **GitHub Actions**, **Docker**, **Kubernetes**, Terraform, Jenkins, GitLab

PROJECTS

EDA on Global Soil Respiration Data | *MySQL, Tableau* ⑂ Github

- Analyzed the Global Soil Respiration Database (SRDBv5) by cleaning and querying data in MySQL and creating interactive Tableau dashboards for visual insights

WebScraping Using R | *R, rvest, ggplot* ⑂ Github

- Developed an R-based web scraping tool using rvest to collect and update articles from "Parasites & Vectors", optimizing resource usage and performing data cleaning and exploratory analysis with regex and ggplot

AlgoTrade API | *Python, yfinance, Pandas, Tensorflow, ks-api-client* ⑂ Github

- Developed a fully automated NSE stock trading bot in Python by integrating real-time and historical data with yFinance, training ML models (including LSTM) for stock price prediction, and executing live trades via the Kotak Securities API

RESEARCH PUBLICATIONS

1. Interpretable Deep Learning for Solar Flare Prediction — IEEE ICTAI 2024 2024
2. An Interpretable Transformer Model for Operational Flare Forecasting — FLAIRS 2024 2024

CERTIFICATIONS

- Data Engineering Bootcamp — DataExpert.io 2025
- Google Data Analytics Professional Certificate — Coursera 2023

EDUCATION

New Jersey Institute of Technology Newark, NJ
*Master of Science in **Data Science*** GPA: 3.89/4
SRM University AP Amaravati, India
Bachelor of Technology GPA: 8.51/10