# Vinay Ram Gazula

Newark, NJ | 908-552-1879 | gazulavinayram@gmail.com | vinaygazula.dev | linkedin.com/in/vinayramg

## Professional Summary

MS **Data Science** graduate with strong research background and multiple publications focusing on machine learning and explainable AI. Holds over 4 years of experience designing, building, and maintaining scalable data pipelines to enhance data quality and drive business growth. Proficient in Python, SQL, Excel, and BI tools like Tableau, Power BI and Looker. Excels at collaborating with data scientists, engineers, and analysts in cross-functional teams.

## Education

**New Jersey Institute of Technology** | Newark, NJ — May 2025
*Master of Science in **Data Science*** — GPA: 3.89/4
Coursework: Applied Statistics, Data Analytics with R Program, Data Visualization, Advanced Database Systems Design

**SRM University AP** | Amaravati, India — May 2021
*Bachelor of Technology in Mechanical Engineering* — GPA: 8.51/10

## Experience

**Data Scientist** — Jan 2024 — May 2025
**New Jersey Institute of Technology** — Newark, NJ

- Collaborated with multiple interdisciplinary research teams to extract actionable insights from complex scientific datasets
- Conducted analyses on institutional undergraduate student data, developing regression models in R to predict GPA
- Performed ablation studies to evaluate the influence of academic, demographic, and socioeconomic factors on student performance
- Designed and implemented "SolarFlareNet"—a transformer framework for forecasting solar flare occurrences achieving 90.7% accuracy
- Automated the end-to-end ETL pipeline for SolarFlareNet using Azure Data Factory to ingest solar data via the DRMS Python package into Azure SQL Database, ensuring data quality and accelerating model iteration speed by 50%
- Integrated explainable AI algorithms (LIME, SHAP, Anchors, PDP, ALE) into SolarFlareNet to interpret/explain black-box model predictions
- Published and presented research findings at AL/ML conferences FLAIRS 2024 and IEEE ICTAI 2024

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

- Implemented robust data models for generating 15+ key KPIs improving accuracy and alignment with business reporting needs
- Developed Tableau dashboards enabling real-time KPI tracking for 10+ stakeholders and improving reporting accuracy
- Designed and implemented scalable ETL pipelines integrated with data quality checks to ingest and process 10 TB of raw data using PySpark, reducing processing times by 40% and accelerating access to business insights
- Leveraged Alteryx and AWS Glue for ETL processes, driving a 20% increase in transformation speed and data accuracy.
- Reduced 30% cloud storage and I/O costs by implementing Apache Parquet snappy compression and Amazon S3 life cycle policies
- Proposed a data transformation plan utilizing DBT and Aiflow to achieve a 15% increase in transformation efficiency

## Projects

**EDA on Global Soil Respiration Data** | *MySQL, Tableau* — Dashboard

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

**University Ranking** | *Excel, Tableau* — Dashboard

- Extracted public university ranking data, performed data cleaning and aggregations on MS Excel and performed advanced visualization techniques to generate comprehensive dashboards using Tableau

**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

## Technical Skills

| | | |
|---|---|---|
| Languages | : | Python (Pandas, Numpy, scikit-learn, matplotlib, dash), SQL, R (rvest, dplyr, ggplot) |
| Databases | : | PostgreSQL, MySQL, Oracle (PL/SQL), MongoDB |
| Cloud | : | AWS (S3, Lambda, Athena, Redshift, Aurora, RDS, DynamoDB, QuickSight)<br>Azure (Data Factory, Data Lake Storage, Synapse Analytics, Blob Storage)<br>GCP (Cloud Storage, BigQuery, Dataflow, Dataproc, Bigtable) |
| Big Data | : | Trino, Databricks, Snowflake, Apache Spark, DBT |
| Analytics | : | Tableau, Looker, Power BI, MS Excel |

## Certifications

- Google Data Analytics Professional Certificate — Coursera — 2023