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 as a Data Engineer in retail, healthcare and finance domains. Over 1 year of experience as a Data Scientis working with scientific data and building predictive models. Proficient in designing, building, and maintaining scalable data pipelines to enhance data quality and drive business growth. Expert in data modeling, data analysis, data warehousing, cloud technologies, statistics and ML/DL algorithms. Proven track record in cross-functional collaboration, effectively communicating with data scientists, engineers, and analysts in cross-functional teams.

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 data 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

Data Engineer June 2019 — Aug 2023
Impetus Technologies Bengaluru, India

- Developed end-to-end data warehousing solution to track customer behavior and enable targeted ad campaigns increasing revenue by 10%
- Designed and implemented scalable ETL pipelines integrated with data quality checks to ingest and process 4 TB of raw data using PySpark, reducing processing times by 40% and accelerating access to business insights
- Leveraged AWS Lambda and Glue for ETL processes, driving a 25% 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
- Implemented robust data models for generating 10+ key KPIs improving accuracy and alignment with business reporting needs
- Accelerated the migration of data from Snowflake warehouse to S3 for a data lake solution, leveraging Athena for ad hoc analysis and Redshift Spectrum with materialized views for BI dashboards, resulting in 50% faster analytics reporting
- 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, Pandas, Tensorflow, PyTorch, scikit-learn, Streamlit), SQL, R, Bash
Databases	:	PostgreSQL, MySQL, Oracle (PL/SQL), MongoDB
Cloud	:	AWS (S3, Glue, Lambda, Athena, Redshift, Aurora, RDS, DynamoDB, Firehose, SageMaker) Azure (Data Factory, Data Lake Storage, Synapse Analytics, Blob Storage) GCP (Cloud Storage, BigQuery, Dataflow, Dataproc, Bigtable)
Big Data	:	Trino, Databricks, Snowflake, Apache Spark, DBT
Visualization	:	Tableau, Looker, Power BI, Dash (Python), Apache Superset, Excel
CI/CD	:	Git, GitHub, GitLab, GitHub Actions, Docker, Kubernetes, Terraform, Jenkins

Projects

TradeForecast | *Python, PyTorch, PyTorch Lightning, yfinance, Polars, scikit-learn* Report

- Developed three deep-learning architectures (LSTM, CNN+LSTM, Transformer) for multi-horizon timeseries forecasting of stock price
- Orchestrated training using "ReduceLROnPlateau" learning rate scheduler in PyTorch for faster convergence, and implemented hyperparameter tuning using grid search
- Implemented feature engineering by ingesting OHLCV data via yfinance and adding temporal variables (Day of week, Fiscal Quarter) and technical indicators (MA, MACD, RSI, ATR)

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

Education

New Jersey Institute of Technology | Newark, NJ 2023 — 2025
*Master of Science in **Data Science*** GPA: 3.89/4
Coursework: Applied Statistics, Machine Learning, Deep Learning, Data Analytics with R Program, Data Visualization