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. Over 1.5 years of experience as a Data Scientist working with scientific data and building predictive models. Proficient in data modeling, data analysis, data warehousing, cloud technologies, statistics and ML/DL algorithms. Proven track record in collaboration, effectively communicating with data scientists, engineers, and analysts in cross-functional teams.

Education

New Jersey Institute of Technology — Newark, NJ
*Master of Science in **Data Science*** — GPA: 3.89/4
Coursework: Introduction to Big Data, Advanced Database Systems Design, Machine Learning, Data Visualization

SRM University AP — Amaravati, India
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 performance
- Designed and implemented "SolarFlareNet"—a transformer framework for forecasting solar flares 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 model predictions

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
- 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)

Data Engineer Playground | *Docker, Airflow, Trino, Spark, MinIO, PostgreSQL, Project Nessie, Unity Catalog* — Github

- Built a fully containerized multi-service environment to prototype end-to-end ETL workflows, from data ingestion in MinIO to batch or stream processing with Spark and workflow orchestration via Airflow. Enabled interactive SQL analytics through Trino with connectors for Postgres DB, Nessie Catalog and Unity Catalog

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 (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)
Big Data	:	Trino, Databricks, Snowflake, Apache Spark, DBT
Visualization	:	Tableau, Looker, Power BI
Data Modeling	:	Normalization (3NF), OBT, Star Schema, Snowflake Schema, Data Vault
CI/CD	:	Git, GitHub, GitLab, GitHub Actions, Docker, Kubernetes

Certifications

- Google Data Analytics Professional Certificate — Coursera — 2023