

PROFESSIONAL SUMMARY

Data Engineer with over 3 years of experience designing scalable ETL/ELT pipelines, data warehouses, and utilizing frameworks like Apache Spark, Databricks, and Snowflake. Proficient in advanced data modeling techniques and effective at collaborating with data scientists, analysts, and cross-functional teams. Holds a strong research background in data science and machine learning, with multiple publications focused on explainable AI.

EXPERIENCE

New Jersey Institute of Technology Newark, NJ

Research Assistant | NJIT Engineering Education Research July 2024 — Present

- 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 May 2024 — Present

- 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 Sept 2023 — May 2024

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

Impetus Bengaluru, India

Data Engineer July 2020 — Aug 2023

- Optimized data lakes in Amazon S3 by applying lifecycle policies, converting partition files to Parquet, and sorting for run-length encoding, which resulted in a 25% reduction in storage and I/O costs
- Collaborated with analytics teams to design data models for KPI generation and efficiently handled 20+ ad hoc requests, ensuring alignment with business needs
- Developed robust ETL pipelines using Apache Spark on Databricks to ingest upstream master data and populate Apache Iceberg tables, cutting processing times by 40% for faster business insights
- Utilized Apache Airflow to orchestrate complex data workflows, enhancing pipeline reliability and operational efficiency in a dynamic cloud environment
- Leveraged AWS Glue Data Catalog and AWS Lake Formation to standardize metadata management and enforce data governance policies, reducing integration complexities and accelerating analytics workflows

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), 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

eComputer Store Database System | MySQL, Python, Streamlit, Pandas [🔗 GitHub](#)

- Developed a comprehensive MySQL database system for an e-commerce store by designing ER diagrams, data modeling, translating them into a relational schema, and building an interactive web interface with Python’s Streamlit library

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

PUBLICATIONS

- Interpretable Deep Learning for Solar Flare Prediction — [IEEE ICTAI 2024](#)
- An Interpretable Transformer Model for Operational Flare Forecasting — [FLAIRS 2024](#)

CERTIFICATIONS

- Data Engineering Bootcamp — [DataExpert.io](#) 2025
- Google Data Analytics Professional Certificate — [Coursera](#) 2023

EDUCATION

New Jersey Institute of Technology | Newark, NJ Sept 2023 — May 2025
Master of Science in Data Science GPA: 3.89/4

SRM University AP | Amaravati, India Aug 2017 — May 2021
Bachelor of Technology GPA: 8.51/10