

Rohit Reddy Bollam

+1 (201)-680-88 | rbollam@stevens.edu | [Linkdin](#) | [GitHub](#) | [Leetcode](#)

EDUCATION

Stevens Institute of Technology
Masters of Science in Information Systems

Hoboken, USA
Jan 2025 - May 2026

Gokaraju Rangaraju Institution of Engineering and Technology
Bachelor's in Technology in Computer Science

Hyderabad, India
Aug 2020 - July 2024

EXPERIENCE

Data Engineer Intern
Snicare Software Solution Pvt Ltd

June 2024 – Nov 2024
Hyderabad, India

- **Azure Cloud Infrastructure:** Engineered **end-to-end ETL pipelines** on **Azure** to automate the daily ingestion of thousands of records from **REST APIs** and **RDBMS**, improving data availability for stakeholder reporting.
- **Recruitment Market Intelligence:** Architected scalable data models and schemas that powered a **Market Analysis Dashboard**, enabling recruiters to identify shifting company requirements and market trends in real-time.
- **Data Quality Integrity:** Performed extensive **Exploratory Data Analysis (EDA)** and data cleaning using **Python** and **SQL**, identifying anomalies and ensuring 100% data reliability for downstream analytical use cases.
- **Pipeline Optimization:** Streamlined legacy transformation scripts to eliminate processing bottlenecks, reducing data latency and ensuring critical dashboard updates were completed with 100% consistency before business hours.
- **Cross-Functional Collaboration:** Partnered with **Data Analysts** to translate complex business requirements into optimized technical schemas, bridging the gap between raw data sources and actionable recruitment insights.

PROJECTS

End-to-End E-Commerce ETL Pipeline | *Python, PostgreSQL, SQLAlchemy, Pandas, PyTest, SQL*

- Architected a production-ready ETL pipeline processing **100,000+ transaction records** across 9 relational datasets, implementing incremental loading logic to optimize processing time for daily updates.
- Designed a normalized PostgreSQL database schema (Star Schema concept) with automated constraints and Foreign Key enforcement, ensuring **100% referential integrity** between customers, orders, and geolocation data.
- Implemented an automated **Data Quality (DQ) framework** that validates schema conformity and null-checks prior to loading, generating automated quality reports to identify data drift or corruption.

Credit Risk Scoring System with Explainable AI | *Python, LightGBM, SHAP, LIME, Pandas, Scikit-learn, MLflow*

- Developed a high-performance credit risk assessment model by benchmarking 5 algorithms (including XGBoost and CatBoost), achieving a **75.6% AUC-ROC** on imbalanced financial data using LightGBM.
- Implemented Explainable AI (XAI) using **SHAP** and **LIME** to interpret local and global model predictions, ensuring model transparency and compliance with financial regulations (GDPR/ECOA)
- Engineered a robust data processing pipeline handling missing values and outlier detection for a dataset of **100,000+ records**, utilizing Stratified K-Fold cross-validation to prevent data leakage.

E-commerce Chatbot | *Python, PyTorch, NLTK, Scikit-learn, Flask, NLP*

- Engineered a context-aware chatbot using **Deep Neural Networks (PyTorch)** to automate customer support, successfully handling product queries and recommendation requests with high accuracy.
- Optimized the **NLP pipeline** using NLTK for tokenization and stemming, applying a Bag-of-Words (BoW) model to train a custom feed-forward neural network for intent classification.
- Deployed the inference engine via a **Flask REST API**, enabling real-time integration with the web platform and reducing average customer query response time to **<200ms**.

TECHNICAL SKILLS

Languages: Java, Python, SQL

Developer: Git, GitHub, Data Structures and Algorithms, Flask, MLflow, FastAPI, Docker

Data Engineering: Pandas, NumPy, ETL Pipelines, Data Quality, DAX, Excel, Power BI, Tableau, PostgreSQL, SQLAlchemy

Machine Learning: Scikit-Learn, TensorFlow, Keras, PyTorch, XGBoost, LightGBM, CatBoost, SHAP, LIME, NLTK