

Rishik Kasula

Charlotte, NC · (704) 421-0449 · contact.rishikkasula@gmail.com
rishikkasula.github.io · linkedin.com/in/rishikkasula · github.com/RISHIKKASULA

SUMMARY

Master of Science (M.S.) in Computer Science (Data Science) from UNC Charlotte (May 2026, GPA 3.80), building and shipping ML, data, and AI/LLM systems end-to-end. Author of three public, CI-tested projects: a streaming data-quality gate for Apache Kafka, a calibrated card-fraud scoring service, and an XBRL-graded LLM extraction harness. Background in Kafka pipelines, ETL (Extract-Transform-Load), local LLM deployment, and Power BI analytics.

TECHNICAL SKILLS

Machine Learning & AI: Python, scikit-learn, Classification, SVM, Feature Selection, Optimization (Genetic Algorithms, Particle Swarm), Probability Calibration, Model Evaluation, Bootstrap Confidence Intervals

AI & LLMs: LLMs, Retrieval-Grounded Generation, Prompt Engineering, LLM Evaluation Harnesses, Hugging Face, PrivateGPT, Local LLM Deployment, OpenAI API

Data Engineering: Apache Kafka, ZooKeeper, ETL (Extract-Transform-Load), SQL, Real-Time (Streaming) & Batch Ingestion, Change Data Capture (CDC), Data Quality & Validation, Avro, AWS Glue/S3, DuckDB

Engineering & Tooling: FastAPI, Docker, CI (Continuous Integration / GitHub Actions), Git, pytest, mypy, Power BI, Linux

EDUCATION

University of North Carolina, Charlotte

Charlotte, NC

Master of Science (M.S.) in Computer Science — Data Science Concentration

GPA: 3.80 | Graduated May 2026

Coursework: Artificial Intelligence, Big Data Analytics, Knowledge Discovery in Databases, Visual Analytics, Database Systems

Vellore Institute of Technology (VIT-AP)

Amaravati, India

B.Tech in Computer Science (Networking & Security)

CGPA: 8.31/10 | May 2024

Coursework: Data Structures, DBMS, Object-Oriented Programming, Cloud Computing, Foundations for Data Analytics

WORK EXPERIENCE

Graduate Research Assistant — Static Code Analysis & Data Pipelines

Charlotte, NC

UNC Charlotte, College of Computing & Informatics

Jan 2026 – May 2026

- Designed and integrated an information-theoretic feature-weighting module into the composite scoring layer of *RefExp*, a faculty-led static-code-analysis research tool built on Joern code property graphs.
- Automated the tool end-to-end with a single-command pipeline runner and terminal reporting, and fixed 3 query and scoring bugs affecting 60+ evaluation rows.
- Contributed ~3,500 lines of git-verified Python across scoring, clustering, and tooling; prototyped an LLM-based explanation module on a feature branch.

Instructional Assistant — Software Engineering for AI-Enabled Systems

Charlotte, NC

UNC Charlotte, College of Computing & Informatics

Sep 2025 – Dec 2025

- Supported a graduate course on software engineering for AI-enabled systems (syllabus-documented role), spanning AI system design, data quality, model evaluation, and testing.
- Mentored student teams across the ML lifecycle — data collection, cleaning, labeling, feature engineering, training, and evaluation.

Intern

India

Inspirisys Solutions Ltd.

Jan 2024 – May 2024

- Engineered an Apache Kafka + ZooKeeper ingestion pipeline for real-time querying over banking-IT data, applying ETL (Extract-Transform-Load), Change Data Capture (CDC), and PII handling.
- Deployed PrivateGPT in local open-source and OpenAI-API configurations, and developed Power BI dashboards tracking banking-IT KPIs.

PROJECTS

tickflow — Streaming Data-Quality Gate for Kafka-Compatible Streams (v0.9)

github.com/RISHIKKASULA/tickflow

- Built and released (v0.9, MIT, pre-1.0 with a published roadmap) an open-source contract-enforcement layer for Apache Kafka streams: per-topic Avro contracts and semantic rules enforced inline, violations quarantined with self-describing envelopes; 194 tests at 99.4% core-module coverage, mypy strict, CI on every push.
- In a seeded fault-injection benchmark, replaying with the gate on vs. off cut violated bars from 1,076 to 0 of 15,061, with gated output bit-identical to the ground-truth projection (fixture-scale).

fraudscore — Calibrated Card-Fraud Scoring Service

github.com/RISHIKKASULA/fraudscore

- Shipped a Python service (scikit-learn, FastAPI, DuckDB, Docker, CI) that prices every fraud decision — reviewing a transaction when $P(\text{fraud}) \times \text{amount}$ exceeds the review cost.
- Under leak-free chronological evaluation, an amount-aware rule saved \$84.87 per 10k transactions (~13% cheaper than the best single-threshold baseline); a plain logistic model beat gradient boosting.

filinglens — Evaluation Harness for LLM Extraction from SEC Filings

github.com/RISHIKKASULA/filinglens

- Released an evaluation harness that grades local-LLM financial-figure extraction against each company's own XBRL filings — a 600-call deterministic grid with cluster-bootstrap confidence intervals; mypy strict, coverage-gated CI.
- Measured a model-size cliff on headline-figure extraction with statements in context: an 8B local model reached 87.8%, well clear of both smaller models, whose scores (50.5% and 44.4%) the confidence intervals do not separate; hand-labeled 230 failures into a six-label taxonomy.

Feature Selection using Hybrid GA-PSO — Soft Computing, VIT-AP

- Implemented a hybrid Genetic Algorithm-Particle Swarm Optimization feature-selection pipeline (Python, scikit-learn) on remote-sensing data (Indian Pines hyperspectral, 224 bands; Massachusetts Roads), scored by linear-SVM accuracy and Cohen's kappa; converged faster to lower-error feature subsets than GA or PSO alone.

CERTIFICATIONS & LEADERSHIP

Certifications: AWS Academy Data Engineering (Apr 2025, 40 hrs); DeepLearning.AI Mathematics for ML & Data Science (Feb 2025); IBM Python for Data Science, AI & Development; IBM/Cognitive Class Docker Essentials (Apr 2025)

Leadership: Organizing committee, Mana Charitable Trust (2 yrs); Team Lead, VTAPP — 15+ students at VIT-AP's tech fest.