

Salman Farcy Peer

New York, NY (open to relocate) | +1 (240) 726 2921 | pfarcy46@umd.edu | [LinkedIn](#) | [GitHub](#) | salmanfarcy.com

PROFESSIONAL SUMMARY

Data Scientist with 3+ years of experience building predictive machine learning models for credit risk, such as probability of default (PD), at a Fortune 500 company. Pursuing an MS in Data Science, specializing in production grade RAG systems and scalable machine learning, with rigorous LLM evaluation and agentic workflows, to deliver traceable, trustworthy AI for regulated finance.

WORK EXPERIENCE

HCLTech

Data Scientist

Chennai, India
Sep 2022 - Jun 2025

- Improved the **lead-to-disbursal ratio by 9%** by building and deploying three **credit risk acquisition models (XGBoost, LightGBM, logistic regression)** into production through the ACTICO Business Rule Engine (H2O, PMML).
- Achieved an average **KS of ~0.32 and Gini of ~0.46** on the validation set through decile analysis, probability calibration, and rank ordering across score bands, ensuring reliable risk separation before deployment.
- Engineered **1,700+ features across ~100K loan applications** by unifying five data sources (credit bureau, bank aggregator, web analytics, SMS, and email) through Weight of Evidence (WOE) encoding and Spark pipelines on Databricks.

University of Maryland | Graduate Assistant

Feb 2026 - Present

- Built a grounded RAG assistant over 5,000+ de-identified counseling records using semantic chunking, vector embeddings, and metadata filtering, achieving **92% retrieval accuracy** with FERPA compliant, cited answers that eliminate hallucinations.
- Cut analyst review time from hours to seconds by deploying the assistant across **40 to 50 clinics**, saving an estimated **150 analyst hours per month** on program planning and quality improvement reporting.

TECHNICAL PROJECTS

FilingIQ: Hybrid RAG system built over SEC 10-K Filings | [\[Live Demo\]](#), [\[GitHub\]](#)

Python, RAG, ChromaDB, SQLite, hybrid search, cross encoder reranking, FastAPI, Docker

- Built a dual engine Hybrid RAG system whose deterministic router sends numeric questions to exact SQL cell lookups and prose to semantic search, **scoring 0.92 faithfulness and 97% routing accuracy** on a held out eval harness.
- Enforced grounding via inline cell level citations and automated refusal that **declined unsupported queries at 0.80 precision**, deployed to production on Hugging Face Spaces with FastAPI.

Sentinel AI: Autonomous Financial Crime Adjudication Agent on Real Bitcoin and Card Fraud Data | [\[Live Demo\]](#), [\[GitHub\]](#)

Python, LangGraph, NetworkX, XGBoost, OpenAI API, Pydantic, FastAPI, Docker

- Built a LangGraph ReAct agent adjudicating detection model alerts across fraud and AML, **clearing 38% of false positives** at equal recall over 633 held out cases, every fact sourced from deterministic tools enforced by a mechanical citation checker.
- Designed a hash sealed evaluation harness scoring final verdicts and investigative steps at **91% overturn precision and 88% step recall**, plus a recall cost curve quantifying review hours saved against true positives lost.

HomeCredit: Credit Default Prediction | [\[GitHub\]](#)

Python, credit risk modeling, feature engineering, model stacking, hypothesis testing

- Built an end to end default prediction pipeline (EDA, cleaning, preprocessing, feature engineering, training, evaluation) reaching **ROC AUC 0.963** on a held out test set to support underwriting risk screening.
- Engineered **500+ features** across delinquency, credit vintage, amount financed, and demographics, then **built a multi level stacked ensemble** that improved generalization over a single model baseline on **~300K loan applications**.

EDUCATION

University of Maryland, College Park

Master's in Data Science | CGPA: 3.9/4.0

College Park, USA
Aug 2025 - May 2027

Sri Sivasubramaniya Nadar College of Engineering

Bachelor of Engineering in Electronics and Communication Engineering | CGPA: 8.47/10

Chennai, India
Jun 2019 - Jun 2022

TECHNICAL SKILLS

Programming & Databases: Python, SQL, PostgreSQL, MySQL

Machine Learning: scikit-learn, XGBoost, PyTorch, TensorFlow, pandas, NumPy, SciPy, feature engineering, model stacking, imbalanced data, calibration, model evaluation & monitoring

GenAI / LLM: RAG, LangChain, LangGraph, vector databases (ChromaDB), LLM evaluation, agentic workflows, Hugging Face

Big Data & Cloud: Spark / PySpark, Databricks, Delta Lake, Apache Airflow, Kafka, AWS (S3, EC2, SageMaker, Glue)

Visualization & Tools: Power BI, Tableau, Matplotlib, Seaborn, Git/GitHub, Docker, FastAPI, REST APIs