

AMAN TRIVEDI

Data and AI Engineer

+91-76655-44777 | amantrivedi2003@gmail.com | [LinkedIn](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data and AI Engineer at AYE Finance and **M.Sc. (Data Science & AI)** candidate at **BITS Pilani** (Digital), building statistically-grounded ML models and interpretable scorecards behind credit and collections decisions for a regulated NBFC. **Own the collections-prioritization model** (XGBoost, benchmarked against an interpretable WOE/IV scorecard) ranking overdue accounts by recovery likelihood, **live for pilot testing** ahead of wider rollout. Also design the **large-scale AWS data pipelines** (4B+ records) that feed these models end-to-end. 2nd Runner-up, AWS Agentic AI Hackathon; HackerRank SQL 5★.

SKILLS

- **Machine Learning & Statistics:** Classification, clustering, WOE/IV scorecards, SHAP explainability, Hypothesis testing, survival analysis, imbalanced-classification (SMOTE), feature engineering.
- **GenAI & Applied Analytics:** RAG pipelines, fine-tuning, LLM-generated reporting, Natural-language-to-query analytics.
- **Tools & Libraries:** Python, XGBoost, scikit-learn, PyTorch, Transformers, statsmodels, LangChain/LangGraph, AWS Bedrock, embeddings, Qdrant
- **Data Analysis & BI:** SQL, Spark SQL, PostgreSQL, Power BI, query performance profiling & tuning, exploratory data analysis, dashboarding
- **Data Engineering & Pipelines:** PySpark, AWS Glue, CDC, Medallion Architecture, ETL/ELT pipeline design & ownership, Kafka, Airflow, data warehousing & modeling
- **Engineering & Cloud (AWS):** FastAPI, REST APIs, Docker, Redis, Lambda, Redshift, EMR, Step Functions, RDS

EXPERIENCE

Data and AI Engineer | AYE Finance

Oct 2025 – Present

- Built the **Collections Prioritization model** (XGBoost) ranking overdue accounts by recovery likelihood, paired with an **LLM-generated reporting layer** auto-narrating segment trends; live for a pilot group in real-user testing ahead of wider rollout.
- Built an interpretable statistical scorecard (Logistic Regression, WOE/IV) benchmarked against it; handled class imbalance via SMOTE/cost-sensitive thresholding notably improving minority-class recall and added PSI-based drift monitoring with MLflow versioning.
- Fine-tuned an open-source LLM (QLoRA) for a lending-domain task, benchmarked against prompted baselines on accuracy, latency, and cost.
- Added a **natural-language-to-query layer** to the **Lending Analytics Platform** (5,000+ employees, live production data), turning plain-English questions into LLM-generated trend summaries.
- Design, build, and own **8+ ETL/ELT pipelines (PySpark, Lambda, Glue)** processing **4B+ rows across 5+ source systems**; shipped the Data Warehouse 3 months ahead of deadline.
- Tuned PostgreSQL queries (indexing, query plans) for the MIS dashboard, cutting report latency from 2 hours to 15 minutes.
- Lead 6–7 engineers across data, ML, and analytics initiatives — sprint planning, code review, stakeholder delivery.

IT Intern – Data Warehousing | AYE Finance

Dec 2024 – Sep 2025

- Led end-to-end design of a data warehouse and ETL pipelines (10M+ records) across LOS, LMS, and CMS systems; presented the architecture to the CTO and delivered an RBI compliance app 3× faster than planned with a 10-member team. Awarded "Cha Gaye – February" for delivery impact.

AI Intern | Accodyn Technologies

Aug 2024 – Dec 2024

- Built an ML-based risk detection model achieving 93% accuracy for a lending application; contributed full-stack ML features and UI/UX.

Technical Intern | RK Agrotech

Jan 2024 – Jun 2024

- Built an inventory and lead management dashboard (HTML/CSS/JS + FastAPI) backed by PostgreSQL, with automated data ingestion from Google Forms/Sheets.

EDUCATION

M.Sc. in Data Science & AI (Pursuing) | BITS Pilani (Digital)

2026 – 2028

B.Tech. in Computer Science & Engineering | MIT-WPU, Pune

2021 – 2025

PROJECTS

Credit Scorecard & Survival-Causal Study | *Python, scikit-learn, statsmodels, lifelines*

- Built an interpretable Logistic Regression scorecard (WOE/IV binning) on a public lending dataset, benchmarked against a gradient-boosted model on AUC/KS/Gini; handled class imbalance via SMOTE and cost-sensitive thresholding meaningfully improving minority-class recall over the baseline
- Ran a Cox proportional-hazards survival analysis on time-to-default, and a propensity-matched / diff-in-diff causal analysis estimating the effect of loan-term modifications on default risk.

LoanRoute Pipeline | *Python, XGBoost, FastAPI, SHAP, Docker, MLflow, Evidently* | [Live demo](#)

- Built and deployed a live loan-applicant prioritization pipeline: automated data ingestion, feature engineering, an XGBoost model, SHAP-based explainability, and a FastAPI inference service.
- Added a data-quality validation gate before scoring and PSI-based drift monitoring (Evidently) on incoming feature distributions, with structured logging for inference latency and volume.
- Containerized the service (Docker) and added MLflow experiment tracking and model versioning to support safe, auditable retraining.

ACHIEVEMENTS & RECOGNITION

- 2nd Runner-up, AWS Agentic AI Hackathon
- 6+ internal awards for delivery excellence and impact.
- HackerRank SQL 5★