

Sri Anvesh Gandhari

anvesh.g@savemymails.com | +1(216)200-0234 | Open to Relocate | LinkedIn

SUMMARY

Data Scientist with **4.5+ years of experience** delivering applied machine learning, analytics, and GenAI systems in **retail and finance** enterprise environments. Core expertise includes **machine learning, deep learning, Python, SQL, cloud-based ML platforms, data engineering, model deployment, MLOps, and LLM systems**. **Successfully delivered credit risk models covering 95% of high-risk accounts** and deployed **LLM and RAG pipelines processing 50,000+ documents monthly**. Experienced in **Generative AI, LLM orchestration, RAG architectures, prompt engineering, transformer models, Hugging Face ecosystems, vector databases, and scalable inference**, building production-ready AI solutions that drive **critical business decisions**.

SKILLS

Programming & Development: Python, SQL, PySpark, Pandas, NumPy, JupyterLab, FastAPI, Scikit-learn, PyTorch, TensorFlow

Machine Learning & Predictive Modeling: Supervised & Unsupervised Learning, Gradient Boosting (XGBoost, LightGBM), Random Forest Techniques, Time Series Modeling, Feature Engineering, Hyperparameter Optimization, Model Validation & Selection

Artificial Intelligence & Generative AI: Large Language Models (GPT-4-class, LLaMA variants), Retrieval-Augmented Generation (RAG), LangChain Frameworks, Transformer-Based Architectures, Hugging Face Models, Prompt Engineering, Vector Embeddings, FAISS

Data Engineering & MLOps: End-to-End ML Pipelines, ETL/ELT Workflows, Model Deployment & Performance Monitoring

Cloud Platforms: AWS (SageMaker, EC2, S3), Docker, Kubernetes

Data Management: Relational Databases (PostgreSQL), NoSQL Data Stores, Analytical Data Warehousing Concepts

Data Visualization & Reporting: Tableau, Power BI, Matplotlib, Advanced Excel, Seaborn, Plotly

EXPERIENCE

Morgan Stanley | Data Scientist

Jul 2025 – Present | USA

- Designed and deployed **supervised and unsupervised credit and loan risk models** leveraging **gradient-boosting techniques (XGBoost, LightGBM)** to analyze client portfolios, covering **95% of high-risk accounts** and improving **decision efficiency for credit analysts**.
- Created **LLM-powered insights pipelines** using **GPT-4** to summarize memos, **BERT** to classify unstructured compliance documents, and **LangChain** to orchestrate workflows, processing **tens of thousands of records monthly** and reducing manual review workload.
- Configured ML-driven predictive insights for operational risk using **Hugging Face Transformers** and **LangChain** to analyze regulatory notes and internal reports, enabling auditors to reduce **preparation time significantly** while improving **workflow efficiency**.
- Applied **PyTorch** and **TensorFlow** to model sequential transaction patterns and behavioral signals, supporting **exploratory risk analytics** and validating signals alongside traditional ML models for **risk assessment**.
- Implemented a **RAG pipeline** using **LangChain** with **FAISS** as the vector database to index and query regulatory and transaction documents, automating context retrieval for LLM-based summaries and processing **50,000 documents monthly** with high retrieval accuracy.

Accenture | Data Scientist

Jan 2020 – Dec 2023 | India

- Developed retail customer purchase and return-risk classification models using **Python, NumPy, Pandas, Scikit-learn**, and **XGBoost**, achieving **0.81 accuracy** and **0.76 precision** on validation data to support merchandising and fraud review decisions.
- Built time-series demand forecasting pipelines for retail SKUs using **Prophet** and **ARIMA**, reducing average forecast error by **120 units per SKU per month**, improving replenishment planning across regional fulfillment centers.
- Engineered customer segmentation and recommendation workflows using **collaborative filtering**, enabling personalized product recommendations across **300K+ active customer profiles** in targeted retail campaigns.
- Integrated an **LLM-assisted analytics layer** to generate natural-language summaries of customer feedback trends and model insights, cutting business insight preparation time from **multiple analyst hours to a few minutes per reporting cycle**.
- Established model explainability techniques including **SHAP** and **feature attribution** to identify key drivers of demand variability, pricing sensitivity, and product returns, improving stakeholder trust and accelerating model approval cycles.
- Orchestrated **NLP pipelines** using **spaCy** and **transformer-based text embeddings** to analyze customer reviews and support tickets, processing **40K–60K records monthly** and reducing manual theme analysis time from **3–4 hours to under 30 minutes per batch**.
- Deployed scalable ML models using **FastAPI, Docker**, and **MLflow**, integrating predictions into downstream retail systems and supporting recurring decision workflows with **sub-300 ms average inference latency**.
- Executed automated data validation and model monitoring checks using **Python** and **SQL** alongside deployed ML services, detecting data drift and pipeline issues early and reducing incident triage time from **hours to minutes**.
- Constructed **SQL-driven Power BI dashboards** integrating transactional data, forecasts, and model outputs, enabling leadership teams to track **KPIs such as sales growth, forecast accuracy, and demand volatility** across **multiple retail categories** in **near real time**.

ACADEMIC PROJECT

Fraud Detection System for Financial Transactions

- Programmed a supervised fraud detection model on public credit-card transaction data, addressing severe class imbalance through resampling and threshold tuning to **reduce false-positive alerts by 20%** while maintaining strong fraud recall.
- Implemented a lightweight inference pipeline on **AWS**, using **LangChain** to integrate an LLM-based explanation component that generates concise, human-readable summaries for flagged transactions, improving model interpretability.

Skills: Python, Pandas, SQL, Imbalanced Learning, Classification Models, Model Evaluation, AWS, REST APIs, LangChain, LLM-based Explainability

Customer Churn Prediction with Business Insights

- Architected a customer churn prediction model using a public telecom dataset, conducting structured **EDA** and **feature engineering** to improve churn prediction performance by **15%**, enabling early identification of high-risk customers for retention-focused analysis.
- Trained and evaluated multiple ML models (**Logistic Regression, Random Forest, XGBoost**) and executed batch inference workflows on **AWS (EC2 + S3)** to generate churn risk scores and analytical reports for academic business insight evaluation.

Skills: Python, EDA, Feature Engineering, Pandas, SQL, Logistic Regression, Random Forest, XGBoost, Model Evaluation, AWS (EC2, S3)

EDUCATION

Master of Science in Computer Science | Cleveland State University, Cleveland, Ohio, United States.