

Shreya Rahul Kale

+1(680)356-6752 | shreyakale0925@gmail.com | LinkedIn

SUMMARY

Experienced AI/ML Engineer with around 4 years of experience designing, training, and deploying machine learning and generative AI solutions using Python, SQL, deep learning frameworks, and cloud-based MLOps platforms for financial risk and fraud use cases. Built end-to-end pipelines spanning data preprocessing, feature engineering, model evaluation, CI/CD-driven deployment, and model governance, integrating LLMs, RAG workflows, and explainable AI. Partnered with cross-functional teams to translate business and regulatory requirements into scalable, production-ready AI systems with measurable performance improvements.

SKILLS

Programming Languages & Scripting: Python, SQL, Scala, Bash Scripting

Machine Learning & AI Techniques: Supervised Learning, Unsupervised Learning, Feature Engineering, Model Training, Model Evaluation, Hyperparameter Tuning, Classification, Regression, Clustering, Time Series Forecasting, Predictive Analytics, Risk Modeling, Anomaly Detection

Deep Learning, NLP & Generative AI: Neural Networks, CNN, RNN, LSTM, Transformers, NLP Pipelines, Text Classification, Named Entity Recognition, Embedding Techniques, Large Language Models (LLMs), Prompt Engineering, Retrieval-Augmented Generation (RAG), LLM Fine-Tuning

ML Frameworks, Libraries & GenAI Tools: Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, LightGBM, Hugging Face Transformers, OpenAI APIs, LangChain, Vector Databases

Data Engineering & Processing: Pandas, NumPy, PySpark, Apache Spark, Data Preprocessing, ETL Pipelines, Data Validation, Feature Stores

Databases & Data Stores: PostgreSQL, MySQL, Snowflake, MongoDB, Redis

Cloud Platforms, MLOps & CI/CD: AWS (S3, EC2, SageMaker, Lambda), Azure ML, GCP AI Platform, Docker, Kubernetes, MLflow, Model Deployment, Git, GitHub, GitLab, Jenkins, CI/CD Pipelines, Model Versioning, Experiment Tracking, Model Monitoring

Finance & Enterprise Analytics Tools: Financial Risk Analytics, Portfolio Analytics, Credit Risk Modeling, Fraud Detection Systems, Regulatory Reporting Analytics, Data Governance, Model Validation, Bias Detection, REST API Integration, Agile/Scrum Collaboration

EXPERIENCE

AI/ML Engineer

Northern Trust, USA

Mar 2025 – Present

- Identified manual risk-review delays and implemented high-performance predictive workflows using XGBoost, Random Forest, AWS SageMaker, CUDA acceleration, and Snowflake exposure data, reducing operational review effort by 45%.
- Designed enterprise GenAI document intelligence by applying LLM-based reasoning with Azure OpenAI GPT-4, Hugging Face Transformers, LangChain RAG, and FAISS, cutting regulatory and policy analysis effort by 50%.
- Engineered transaction anomaly classification using logistic regression and gradient boosting on AWS S3 financial logs with Pandas-driven feature extraction, improving early risk identification accuracy by 31%.
- Stabilized model lifecycle operations by orchestrating automated retraining and monitoring pipelines using Apache Airflow and MLflow governance, lowering false-positive risk alerts by 27% across compliance workflows.
- Enabled scalable analyst decision support by deploying containerized AI inference services with Docker and GitHub Actions CI/CD, boosting enterprise finance team productivity by 38% through faster, explainable insights.

Machine Learning Engineer

Persistent Systems, India

Apr 2020 – Jul 2023

- Analyzed lending inefficiencies and constructed end-to-end credit risk workflows using XGBoost with Scikit-learn pipelines, PostgreSQL finance datasets, AWS EC2 compute, and Power BI insights, improving default prediction accuracy by 29%.
- Designed customer risk segmentation and approval intelligence by applying Random Forest models with Pandas-driven feature engineering over loan and repayment histories stored in AWS S3, reducing manual underwriting effort by 42%.
- Engineered real-time transaction fraud intelligence by implementing Isolation Forest and Gradient Boosting using PySpark ML across Kafka streams and Spark processing layers, increasing fraud detection effectiveness by 34%.
- Optimized behavioral risk evaluation by deriving transaction velocity and deviation features using NumPy-based statistical modeling on banking aggregations, lowering false fraud alerts by 26% while protecting payment integrity.
- Deployed scalable risk and fraud scoring services by integrating trained models via Flask inference APIs, Dockerized delivery, and Git-managed ML workflows, accelerating decision turnaround by 18% across financial operations.

EDUCATION

Master of Science in Computer Science - Syracuse University, USA

May 2025

Bachelor of Technology in Information Technology - Savitribai Phule Pune University, India

Jun 2022

PROJECTS

Large Language Model-Powered Fraudulent Job Detection

- Initiated development of an LLM-powered fraud job listing detection model, applying prompt engineering and fine-tuning to achieve 95% test accuracy with strong NLP-driven classification performance.
- Engineered features from 17.8K job listings using NLP preprocessing and encoding, enhancing recall for fraudulent listings by 15% through class balancing and tuning while maintaining 91% precision and delivering 0.968 ROC-AUC.

Hit Song Prediction in Indian Popular Music

- Engineered supervised machine learning pipeline for Indian hit song prediction using Spotify Web API data, Librosa, Aubio, and custom pitch features, generating 150+ multimodal features.
- Trained and evaluated SVM, Random Forest, Naïve Bayes, and k-NN models with normalization, class balancing, and k-fold validation, identifying MFCCs, loudness, danceability as strongest predictors.

CERTIFICATIONS

- Forward Learners Program via McKinsey & Company
- Python3 Programming Specialization from University of Michigan via Coursera
- Java Programming by SEED Infotech
- 5-Day AI Agents Intensive Course with Google