

Pratiksha Dhavane Data Scientist

pratiksha.dhavane1998@gmail.com | 8793378794 | In/Pratiksha-Dhavane | Pune, Maharashtra | Portfolio

Professional Summary

Data Scientist with **4+ years of experience** in ML, NLP and enterprise GenAI, specializing in AI agents and agentic workflows across finance, insurance, and pharma. Proven at designing production grade ML systems, tool-driven agents, and reliable GenAI applications, translating ambiguous business problems into scalable, high-impact data products.

Core Competencies

- **Programming & ML:** Python, Pandas, NumPy, Scikit-learn, XGBoost, Feature Engineering, Time Series Forecasting
 - **GenAI & Agentic AI:** LLMs, AI Agents, Agentic Workflows, Tool Calling & Routing, RAG, NL2SQL, Embeddings, Prompt Engineering, Verification & Guardrails
 - **Frameworks & Orchestration:** LangChain, LangGraph, OpenAI APIs, AWS Bedrock
 - **Cloud & Data Engineering:** AWS S3, Glue, Athena, Lambda, OpenSearch Serverless
 - **Databases & Dev Tools:** SQL, MongoDB, Git, Postman
 - **Deployment:** Flask, FastAPI, Streamlit
-

Professional Experience

Senior Analyst | Proclink Consulting LLP, Hyderabad | Jan 2024 – Present

- Developed multiple enterprise GenAI solutions including **NL2SQL** engines using LangChain, Claude Sonnet, and AWS-native services (e.g.: AWS Bedrock).
- Built an email preprocessing and categorization pipeline to clean, parse and structure raw email threads for NLP analysis.
- Built a cross-sell **propensity model** using XGBoost, improving conversion rates by **30%** across multiple insurance segments.
- Designed a PyCaret-based **revenue forecasting pipeline**, improving accuracy by **25%** across 6+ product lines and enabling reliable financial planning.
- Automated ML retraining and data workflows using AWS Glue + MWAA, reducing manual analyst effort by **70%**.

Data Scientist | Mphasis, Pune | Apr 2023 – Nov 2023

- Built ML-driven **insurance underwriting models** using Decision Trees and Random Forest, improving risk stratification accuracy by **15%**.
- Automated eligibility and premium recommendation pipelines, cutting manual review time by **40%**.

Assistant System Engineer | Tata Consultancy Services, Pune | May 2021 – Mar 2023

- Developed NLP sentiment analysis system for **1M+ hospitality reviews** using Word2Vec, TF-IDF, SVM, and Random Forest; achieved **>85% accuracy** and deployed via Flask API.
 - Built internal credit underwriting models using bureau, GST, and financial features, improving loan recommendation consistency.
 - Automated underwriting workflows, reducing operational bottlenecks and turnaround time.
-

Key Projects

Conversational AI NL2SQL

Tech: AWS Bedrock, Claude 4.5 Sonnet, Opensearch Serverless, Rerankers, LangChain, Athena, Lambda

- Built a production-grade NL2SQL engine enabling natural-language analytics on complex credit card datasets.
- Designed a hybrid retrieval pipeline using Opensearch vector + lexical search and a reranker, improving schema relevance by **~40%**.
- Implemented multi-step SQL generation, validation, and correction using Claude Sonnet and LangChain orchestration.

AI Agent Chatbot - Graph-Based Control Flow with Verification, Safety & Short-Term Memory

Tech: Python, LangChain, LangGraph, LLMs, Streamlit

- Built a production-oriented AI agent using LangGraph with explicit graph-based control flow across decision, search, synthesis and verification nodes, with multi-provider model fallback across Gemini and Groq. Implemented agentic control flow with tool routing, model fallback and response verification.
- Implemented hybrid short-term memory combining a sliding window of recent turns with periodic LLM summarization, enabling context-aware routing without unnecessary search.
- Designed answer verification for grounding, hallucination detection, bounded retries and immediate abort on critical failures, with full observability per query including routing decision, retries, confidence and latency.

First-Party Fraud Classification Model

Tech: XGBoost, Python, SMOTE, KS Evaluation

- Developed a fraud classification model for a dataset with 3.6% bad rate, using 1700+ bureau variables reduced to ~300 via XGBoost importance.
- Tuned model using KS, precision-at-deciles, and calibrated probability thresholds.
- Built a complete feature engineering → training → scoring pipeline optimized for imbalanced classification.

Education

B.Tech | *Vishwakarma Institute of Technology, Pune* | 2016 – 2020 | **CGPA: 8.52**

HSC 82.77% | **SSC 92.40%**

Languages

English | Hindi | Marathi
