

# AJAY ARAKH

Data Scientist | Machine Learning Engineer

 arakhajay42@gmail.com |  9579852657 |  Pune, Maharashtra, India

---

## PROFESSIONAL SUMMARY

Results-driven Data Scientist with 4.3 years of experience in leveraging data, machine learning, and advanced AI techniques to deliver impactful business solutions. Skilled in the end-to-end lifecycle of model development from data collection, cleaning, and feature engineering to building, deploying, and monitoring predictive and generative models in production. Experienced in developing RAG pipelines, intelligent chatbots, and automation systems that enhance decision-making and business efficiency. Proficient in MLOps, cloud platforms (Azure, Databricks), and CI/CD workflows. Strong collaborator with proven ability to translate business requirements into data-driven solutions while staying updated with the latest trends in AI and applied ML.

---

## PROFESSIONAL EXPERIENCE

### Unit Manager - Analytics DMS COE

Bajaj Finserv Ltd | Mar 2021 – Jun 2025

#### Roles & Responsibilities

- Analyzed complex business problems in credit card collections and developed data-driven solutions
- Developed and deployed financial chatbot using Langchain, langgraph frameworks and OpenAI LLM to implement the RAG's flow for generating contextual responses from LLM.
- Developed Risk of Flow models for Credit Card collections across PDD (Pre Due Date), PreX, and Write-off customer segments
- Built end-to-end solutions for variety of business problems with strategic focus on recovery efficiency

- Conducted in-depth business problem analysis and delivered ML insights aligned with operational priorities
  - Presented analytical findings to cross-functional stakeholders including product, risk, and collections teams
  - Understood and analyzed complex business problems in collections and personal loan segments
  - Developed Collections Risk of Flow models to predict customer movement across delinquency buckets
  - Built Logistic Regression models to predict customer repayment behavior, delinquency risk, and default likelihood
  - Monitored model performance using Population Stability Index (PSI) and Characteristic Stability Index (CSI)
  - Identified data drift and ensured model stability over time through continuous monitoring
  - Automated reporting processes, reducing daily manual work by 2 hours
  - Supported risk and collection teams in integrating model insights into business operations
  - Developed and deployed FinBot, an LLM-based customer support chatbot using LangChain
  - Developed Multi-Source Advanced RAG Chatbot and Agentic ai application for Helping out campaign creation for marketing
  - Ensured compliance and audit readiness through comprehensive documentation
- 

## PROJECTS

### 1. FinBot - LLM-Powered Financial Chatbot

#### Project Description

Developed an AI chatbot capable of understanding and interpreting financial documents, customer FAQs, and policies using advanced natural language processing. The chatbot leverages Retrieval-Augmented Generation (RAG) methodology to provide accurate, context-aware responses to customer inquiries.

#### Roles & Responsibilities

- Designed and architected the chatbot system for financial document understanding

- Integrated LangChain and Langgraph framework with OpenAI models for NLP processing
- Implemented RAG pipeline to retrieve and generate accurate responses
- Deployed the chatbot for customer support operations
- Integrated with existing customer support workflows

## Tech Stack Used

- LLM Framework: LangChain, Langgraph, HuggingFace Models, OpenAI LLM
  - Architecture Pattern: Retrieval-Augmented Generation (RAG)
  - AI/ML Domain: Generative AI, NLP
  - Languages: Python
  - Deployment: Production environment
- 

## 2. Multi-Source Advanced RAG Chatbot (LangGraph & LangSmith)

### Project Description:

Developed a multi-source RAG system using LangGraph for complex query orchestration. The system features a comprehensive evaluation framework to ensure high-fidelity responses across disparate data sources.

### Roles & Responsibilities

- Implemented Ragas and DeepEval metrics (Faithfulness, Answer Relevancy, Context Recall) to quantify retrieval accuracy across SQL and Vector databases.
- Integrated LangSmith for full-trace observability, enabling the identification of bottlenecks in the query routing logic and multi-step retrieval process.
- Established automated evaluation pipelines in LangSmith using custom datasets to benchmark performance before each production deployment.

## Tech Stack Used

- Evaluation: Ragas, DeepEval, G-Eval
- Monitoring: LangSmith (Tracing, Alerts)
- Orchestration: LangGraph, LangChain
- Knowledge Base: SQL Server, ChromaDB, MCP Tools

### 3. Agentic AI Campaign Creation & Strategy Suite

#### Project Description

Designed a multi-agent AI system for marketing automation to generate data-driven campaign strategies and multi-channel creative content. The application utilizes a "Human-in-the-loop" workflow to refine strategic outputs and features a robust evaluation layer to ensure brand alignment and content quality.

#### Roles & Responsibilities

- Orchestrated a multi-agent team using LangGraph where specialized agents collaborate to perform market research, draft tailored copy, and design cohesive visual campaign strategies.
- Implemented an Agentic Evaluation Framework to measure Strategic Planning Index (SPI) and Component Synergy Score (CSS), ensuring high-quality alignment between strategy and generated content.
- Utilized LangSmith to monitor production performance in real-time, tracking agent decision-making paths, success rates for creative generation, and token costs.
- Established Online Evaluation triggers in LangSmith to automatically flag low-confidence campaign drafts for manual review and expert feedback.

#### Tech Stack Used

- AI Framework: LangChain, LangGraph (Multi-agent orchestration)
- Monitoring & Observability: LangSmith (Tracing, Cost tracking, Feedback loops)
- Agentic Metrics: Strategic Planning Index (SPI), CSS, Brand Alignment Score
- Deployment: Azure Cloud (App Service, Azure Container Registry)
- LLM: OpenAI GPT-4o / Azure OpenAI (Strategy & Content Generation)

### 4. Allocation Models for PLCS and SAL

#### Project Description

Developed machine learning models to intelligently route delinquent customers to appropriate collection channels (field agents or telecalling teams). The model segments customers into five categories based on allocation probability to optimize collection resource allocation and improve efficiency.

## Roles & Responsibilities

- Analyzed customer behavior patterns to determine optimal allocation strategy
- Built and trained Logistic Regression models for customer classification
- Segmented customers into five categories based on allocation probability
- Performed rank order validation for model performance evaluation
- Collaborated with collections team to integrate model outputs into routing workflows
- Validated model effectiveness through real-world deployment

## Tech Stack Used

- ML Algorithm: Logistic Regression
- Programming Languages: Python, SQL
- Data Processing: Python, Databricks
- Databases: SQL Server, Databricks SQL
- Evaluation Metrics: Rank Order Validation, Classification Performance Metrics

## 5. Credit Card Risk Models for PDD, PreX, Write-off

### Project Description

Designed sophisticated Risk of Flow models to predict customer movement across different delinquency segments (Pre Due Date, Pre-Closure, and Write-off buckets). The system monitors 2 million accounts monthly to prioritize outreach strategies and enhance recovery efficiency.

## Roles & Responsibilities

- Designed Risk of Flow modeling framework for multiple delinquency segments
- Built predictive models to forecast customer transitions across PDD, PreX, and Write-off stages
- Monitored and managed 2 million customer accounts monthly
- Prioritized outreach strategies based on model predictions
- Contributed to 6% improvement in recoveries from early-stage delinquencies
- Implemented continuous monitoring and performance tracking
- Provided actionable insights to risk and collections teams

## Tech Stack Used

- ML Algorithm: Risk Flow Modeling, Statistical Analysis
  - Programming Languages: Python, SQL
  - Cloud & Data Platforms: Azure Databricks
  - Databases: SQL Server, Databricks SQL
  - Data Processing: Spark SQL, Python
  - Tools: Jupyter Notebook, SSMS, Excel
  - Monitoring: PSI, CSI, Performance Metrics
  - Scale: 2 million+ accounts monthly
- 

## 6. Digital Payment Propensity Model

### Project Description

Developed a machine learning model to predict whether PLCS (Personal Loan Cross Sell) and SAL (Salary Account Loan) customers will adopt digital payment methods for loan repayments. This project aimed to enable targeted communication strategies and optimize collection operations while increasing digital payment adoption and reducing operational costs.

### Roles & Responsibilities

- Conducted exploratory data analysis on customer payment behavior patterns
- Engineered features to capture digital payment propensity signals
- Built predictive models for PLCS and SAL customer segments
- Analyzed model predictions to develop targeted communication strategies
- Collaborated with operations and collections teams on implementation
- Identified opportunities to reduce operational costs through digital payment adoption
- Designed interventions to increase digital payment uptake

### Tech Stack Used

- ML Algorithms: Classification Models, Logistic Regression
- Programming Languages: Python, SQL
- Data Processing: Python, Databricks, Spark SQL
- Databases: SQL Server, Databricks SQL

- Feature Engineering: Python, Advanced SQL
  - Tools: Jupyter Notebook, SSMS, Excel
  - Evaluation Metrics: Classification performance, ROC-AUC, Precision-Recall
  - Cloud Platform: Azure Databricks
- 

## 7. Reporting Automation

### Project Description

Built an automated workflow solution that ingests raw Excel/CSV files from Outlook or Databricks, processes them in Databricks using Python and SQL, and emails cleaned and enriched output back to respective teams. The solution significantly reduced manual effort and reporting errors while enabling timely daily insights.

### Roles & Responsibilities

- Designed end-to-end automation workflow architecture
- Built data ingestion pipeline to fetch files from multiple sources (Outlook, Databricks)
- Developed data processing scripts for cleaning and enrichment
- Implemented automated email distribution system
- Tested and validated automation across different teams
- Monitored workflow performance and handled edge cases
- Provided technical support to end-users

### Tech Stack Used

- Data Processing: Python, SQL, Spark SQL
  - Cloud & Data Platform: Azure Databricks
  - Databases: Databricks SQL, SQL Server
  - Data Formats: Excel, CSV
  - Integration: Outlook API, Email services
  - Workflow Orchestration: Python scripts, scheduled jobs
  - Impact: 10+ hours weekly manual effort saved, 90% reduction in reporting errors
-

# TECHNICAL EXPERTISE

## Programming Languages

- Python
- SQL (T-SQL, Databricks SQL, Spark SQL, MySQL)
- Advanced Excel

## Cloud & Data Platforms

- Azure (Cloud Computing, ML Services)
- Azure Databricks
- Jupyter Notebook
- Azure ai foundry
- Azure ML

## Databases & Data Tools

- SQL Server
- MySQL
- Databricks
- SQL Server Management Studio (SSMS)
- Data Warehousing concepts

## Machine Learning & Statistics

- Regression Models
- Classification Models
- Clustering Models
- Risk Modeling

## Model Monitoring Tools

- Population Stability Index (PSI)
- Characteristic Stability Index (CSI)

## Frameworks & Libraries

- Pandas, Numpy, Scikitlearn, Matplotlib etc
- LangChain, LangGraph
- HuggingFace Models, OpenAI etc
- RAG (Retrieval-Augmented Generation)
- Generative AI

## Methodologies & Practices

- Agile Development
- Waterfall Methodology
- Data-Driven Decision Making
- MLOps
- CI/CD Workflows
- Model Monitoring & Performance Tracking

## AI/ML Specializations

- End-to-end ML Lifecycle (Data Collection → Model Deployment → Monitoring)
  - RAG Pipelines
  - Chatbot Development
  - LLM Applications
  - Predictive Modeling
  - Business Problem Analysis
- 

## EDUCATION

BE Electronics and Telecommunication Engineering  
Sanjivani College of Engineering, Kopargaon | Pune University | 2019

Diploma in Electronics and Telecom  
Government College of Polytechnique, Aurangabad | MSBTE | 2013

---