

# Devansi Mahatar

Email: devansimahatar@gmail.com | LinkedIn | GitHub | Phone: 8249537214

## EDUCATION

VIT Bhopal University, B.Tech, CSE (Specialization in E-Commerce Technology) 2022-2026	CGPA: 8.85/10
DAV Public School, Pokhariput, CBSE XIIth (June 2022)	82%
Blossoms School, Bhubaneshwar, CBSE Xth (April 2020)	97.2%

## WORK EXPERIENCE

**Data Science Intern | Sabudh Foundation (Remote)** Jan 2025 - June 2025

- Designed and deployed LLM-powered document analysis system, reducing manual insight extraction time by 60%.
- Implemented Retrieval-Augmented Generation (RAG) pipeline using Pinecone and LangChain for semantic search across multi-document datasets.
- Engineered preprocessing workflows including chunking, embeddings, ranking optimization, and structured validation.
- Created SQL-based agent and Streamlit interface enabling automated querying and real-time insight visualization.

## PROJECTS

### 1. InvoiceShield - Invoice Anomaly Detection & ML Pipeline ([LINK](#))

- Architected **end-to-end invoice processing pipeline** using OCR (88% field extraction accuracy) and engineered features for anomaly detection.
- Applied **DBSCAN** clustering to detect pricing patterns and duplicates, improving anomaly detection precision by 25%.
- Integrated **FastAPI**-based ML inference pipeline integrated with **PostgreSQL** for near real-time anomaly scoring.
- Established automated data ingestion, preprocessing, and validation workflows to ensure structured consistency and reliable anomaly analysis.

### 2. ChurnX - Customer Churn Prediction & Retention Modeling ([LINK](#))

- Analyzed 7,043 customer records using **SQL** to calculate churn rates across tenure, contract type, and monthly charges.
- Performed **data cleaning, EDA and feature engineering** in Python to identify **key churn drivers** and **high-risk segments**.
- Optimized **Logistic Regression** and **Random Forest** models using GridSearchCV, achieving ROC-AUC score of 0.86.
- Built **Power BI dashboard** to visualize churn trends, 30-35% revenue at risk, and retention-focused customer segments.

## TECHNICAL SKILLS

**Machine Learning:** Feature Engineering, Cross-Validation, Hyperparameter Tuning, Logistic Regression, Random Forest, Decision Trees, DBSCAN, Linear Regression, K-means

**Generative AI:** LangChain, Pinecone (Vector Database), Semantic Search, Prompt Engineering

**Programming:** Python (Pandas, NumPy, Scikit-learn), SQL (Joins, Subqueries, Window Functions)

**Tools & Platforms:** PostgreSQL, MySQL, Streamlit, Git, Jupyter Notebook, VS Code, GCP

## ACHIEVEMENTS & LEADERSHIP

- National Semi-Finalist in Flipkart GRiD 7.0
- Advanced to the 3rd round of Arthneeti 4.0 business case study held by IIM Nagpur
- Completed 2nd round (Technical interview), Google STEP internship

## LEADERSHIP EXPERIENCE

**Meraki - The Fine Arts Club | Decor Team Lead** July 2024 - May 2025

Led a team of 10 members in planning and executing décor operations across 5+ college events, coordinating timelines and responsibilities to ensure smooth execution.

## CERTIFICATIONS

- GenAI Academy (Google Cloud x Hack2Skill) - Vertex AI, Gemini APIs, Imagen, Streamlit, Multimodal RAG
- Dataiku Machine Learning Practitioner
- Dataiku Generative AI Practitioner
- Statistics Foundations Professional - Wolfram Research
- HackerRank SQL (Intermediate)

**Interests:** Reading, Mandala Art, Travel and Cultural exploration

**Languages:** Fluent in Hindi, English, Odia; Conversational Proficiency in Tamil