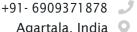
# Diya Goswami

## **CONTACT**

diya13goswami@gmail.com



Agartala, India



LinkedIn in

Github (

## **EDUCATION**

B. Tech

Computer Science and Engineering with specialization in Health Informatics

Vellore Institute of Technology Bhopal September 2022 - July 2026

8.98 CGPA

**Higher Secondary** Auxilium Girls' School March 2021- April 2022

95.2 Percentage

High School Auxilium Girls' School March 2019- April 2020

96.2 Percentage

# **CO-CURRICULARS**

- Smart India Hackathon 2024 Finalist
- Health Hackathon JHU & VITB Finalist
- Presented and Published research work on ML in cardiac disease prediction in ICDCC 2024
  - Core Member, Eureka Club, VIT Bhopal Open-Source Contributor, GitHub

# **PROJECTS**

#### EchoRetail: Retail Feedback Captured and Analyzed by AI

August 2025- Ongoing

- An Al-driven retail analytics system that generates synthetic transaction datasets using GANs and simulates realistic customer behavior.
- Implemented LLM-powered feedback analysis with ChromaDB and Gemini embeddings, enabling natural language queries on customer sentiment and trends.
- Built aspect-based sentiment analysis pipeline to extract opinions on price, quality, delivery, and service, revealing granular insights for business strategy.
- Applied topic modeling with BERTopic to cluster feedback into themes and visualized evolving trends in customer preferences and complaints.
- Conducted exploratory data analysis (EDA) on synthetic datasets, uncovering revenue patterns, category performance, and customer segmentation.

#### Al-Augmented Cardiac Risk Prediction using Synthetic Data Generation **Techniques**

November 2024-May 2025

- Built a synthetic data pipeline using CTGAN, VAE, and Table Diffusion to address dataset imbalance in heart disease prediction.
- Trained multiple models (SVM, XGBoost, CNN, Ensemble) achieving 85.85% accuracy with KNN on VAE data, 83.7% with Ridge Regression on Diffusion data, and consistent improvements across precision/recall (avg. 0.86 F1-score).
- Validated results with confusion matrices, showing reduced false negatives and improved diagnostic reliability using synthetic datasets.

# SkinSight: Intelligent Skin Type Detection System

June 2023- April 2024

- Built a real-time CNN, ResNet-50 and Haar Cascade-based skin type detector on Raspberry Pi 5 + Logitech C920, achieving 80.38% accuracy (Dry: 56/73, Normal: 98/120, Oily: 100/123).
- Optimized with TensorFlow, Keras, OpenCV, and quantized models for low-latency edge deployment; integrated LED ring light and oneway mirror UI for clinical-grade usability.
- **Applications** personalized dermatology, cosmetic recommendations, and telemedicine; demonstrated at ETESM-2025.

## **SKILLS**

Java, C++, Python, SQL, Machine and Deep Learning, Generative Al algorithms, TensorFlow, PyTorch, Scikit-learn, Pandas, NumPy, Keras, Matplotlib, Tableau, LLMs, LangChain, LangGraph, ChromaDB/PineconeDB

## CERTIFICATIONS

- IBM Blockchain Fundamentals and Developers
- FacePrep Mastering Data Structures and Algorithms
- Google Data Analytics