Diya Goswami

CONTACT

diya13goswami@gmail.com
+91- 6909371878
Agartala, India
LinkedIn
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
- Presentation and Publication of SkinSight in ETESM 2025
- Core Member, Eureka Club, VIT Bhopal
- Open-Source Contributor, GitHub

PROJECTS

EchoRetail: Retail Feedback Captured and Analyzed by AI
August 2025- October 2025

- Architected GAN-based data generation system (PyTorch, 500 training epochs) producing 10,000 synthetic retail transactions with realistic correlations, enabling privacy-compliant analytics and reducing data collection costs.
- Deployed production-ready RAG system using Google Gemini LLM, ChromaDB vector store, and LangChain framework to query more than 10,000 customer feedbacks in natural language, reducing insight generation time from hours to seconds.
- Engineered comprehensive NLP analytics suite combining aspect-level sentiment analysis (5 aspects per review), BERTopic clustering (more than 10 auto-discovered themes), accompanied exploratory data analysis and temporal trend visualizations, providing 360-degree customer intelligence dashboard

CardiaSynth: Multi-Model Synthetic Data Generation for Improved Diagnostics
November 2024-May 2025

- Built synthetic cardiac data pipeline using CTGAN, VAE, and Table Diffusion models to address class imbalance, generating balanced training datasets that improved ML model performance by 15-20% across 5 algorithms
- Trained ensemble of 5 models (SVM, XGBoost, CNN, Ridge, KNN) achieving 85.85% peak accuracy, 83.7% secondary accuracy, and average F1-score of 0.86 with validated precision/recall improvements
- Reduced false negative rates by more than 25% through confusion matrix optimization across 3 synthetic datasets, enhancing cardiac risk detection and clinical diagnostic reliability

SkinSight: Intelligent Skin Type Detection System
June 2023- April 2024

- Developed production-ready skin type detection system combining CNN, ResNet-50, and Haar Cascade models on Raspberry Pi 5, achieving 80.38% accuracy with consistent 76-81% per-class performance (Dry: 56/73, Normal: 98/120, Oily: 100/123) in real-time edge inference
- Architected clinical-grade user interface with hardware-software integration including LED ring light for controlled illumination, one-way mirror display, and TensorFlow Lite model quantization enabling sub-second inference latency on resource-constrained edge devices

SKILLS

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

CERTIFICATIONS

- IBM Blockchain Fundamentals and Developers
- FacePrep Mastering Data Structures and Algorithms
- FutureSkills Generative AI Fluency
- Google Data Analytics