DIYA GOSWAMI

+91 6909371878 | diya13goswami@gmail.com | LinkedIn | GitHub | Bengaluru, India

EDUCATION

Vellore Institute of Technology
Bhopal
B. Tech Computer Science and Engineering
8.98 CGPA

ACHIEVMENTS

- Smart India Hackathon 2024 Finalist
- JHU and VITB Health Hackathon 2025 Finalist

PUBLICATIONS AND CONFERENCES

- International Conference on Data Computation and Communication, Nov 2024: Presented and published research work on the utilization of ML in alerting cardiac issues
- International Conference on Emerging Trends in ESM, April 2025: Presented a device for the dermatological industry for real time skin type detection

CO-CURRICULAR

- Research Team, Eureka Club, Core Member: Participated and contributed in numerous new innovative research related work along with organizing multiple events
- Fine Arts Club, Active Member: Participated in collaborative and engaging activities showcasing creativity, communication and networking skills

SUMMARY

Analytical and data-driven, with experience in statistics, ML, and real-world data workflows. I enjoy building practical pipelines and predictive models that genuinely help solve problems.

EXPERIENCE

Intern, YugaYatra Retail (OPC) Private Limited | Remote
December 2025- Present

• Analyzed and prepared operational and customer data through cleaning, querying, and structuring to generate insights that supported business decisions and process improvements.
• Designed intuitive and visually engaging interfaces using AI-assisted tools, helping translate analytical insights into easily understandable outputs for stakeholders.
• Collaborated with cross-functional teams to align data insights, interface design, and business requirements for ongoing product enhancements.

PROJECTS

EchoRetail: Customer Insights & Retail Analytics Platform
August 2025- October 2025

• Built a retail analytics pipeline using 10,000 synthesized transaction records and more than 10k customer reviews, enabling scalable analysis without privacy concerns.
• Implemented NLP-driven insights, including sentiment analysis, aspect categorization, and BERTopic-based theme modeling to uncover customer pain points and satisfaction patterns.
• Designed a query-driven analytical layer using RAG and vector databases (ChromaDB), enabling natural-language access to insights such as product performance, review trends, and seasonal buying behavior.

CardiaSynth: Data Engineering & Predictive Modeling for Cardiac Risk
November 2024-May 2025

• Engineered a multi-model synthetic data generation workflow (CTGAN, VAE, Table Diffusion) to create balanced clinical datasets and resolve real-world class imbalance issues.
• Conducted extensive EDA, feature engineering, and comparative model evaluation across SVM, XGBoost, Logistic Regression, CNN, Ridge, and KNN, achieving 85.85% accuracy and 0.86 F1-score.
• Performed error analysis using confusion matrices, reducing false negatives by 25% and improving reliability of early-stage cardiac risk assessment.

SkinSight: Real-Time Image Analytics for Dermatological Classification
June 2023 – April 2024

• Utilized a diverse dataset of oily, dry, and normal skin images across varied lighting, texture, and complexion conditions to build a robust preprocessing and classification pipeline using CNN and a fine-tuned ResNet-50, achieving 80.38% accuracy.
• Optimized image analytics for edge deployment through data normalization, ROI extraction with Haar Cascade, and TensorFlow Lite quantization, enabling sub-second inference on Raspberry Pi.
• Performed model validation, visualization, and statistical reporting to ensure consistent performance across real-world environments and user variability.

SKILLS

EDA, Feature Engineering, Statistical Analysis, A/B Testing, Data Visualization, Predictive Modeling, Clustering, SQL Querying & Data Warehousing Basics, NLP Analytics, Tableau, PowerBI, Excel, LangChain, LLM APIs, ML and DL