Pranay Shaurya

Bhopal, Madhya Pradesh, India
Pranayshaurya.pro@gmail.com — linkedin.com/in/pranay-shaurya-06106b252 — Pranay-Shaurya — 9334192924

PROFILE SUMMARY

Final-year B.Tech Computer Science student with strong foundations in Python, SQL, and data analytics. Experienced in building analytical data pipelines, performing data transformation using Pandas, and developing machine learning models. Familiar with relational databases, cloud platforms, and business intelligence tools. Seeking to build a long-term career in Data Engineering and AI by solving real-world business problems through data-driven solutions.

EDUCATION

Vellore Institute of Technology – Bachelor of Technology, CSE (Health Informatics) *Sep 2022 – 2026*
CGPA: 8.16/10.0

TECHNICAL SKILLS

Programming: Python, SQL, C++
Data Analytics: Pandas, NumPy, MySQL, Power BI
Machine Learning: Scikit-learn, TensorFlow
Databases: MySQL (Hands-on), PostgreSQL (Working Knowledge), ChromaDB
Cloud & Tools: Google Cloud Platform, Docker, Git, Jupyter Lab
Core Concepts: Data Structures & Algorithms, DBMS, OOPs, Probability & Statistics

EXPERIENCE

Generative AI Virtual Internship – Google Cloud *2025*

- Built and deployed applications using Generative AI tools on Google Cloud (Gemini, Vertex AI).
- Gained hands-on exposure to cloud-based model deployment and prompt engineering.
- Completed advanced learning paths in applied AI workflows and scalable cloud solutions.

PROJECTS

Retail Sales Analytics Pipeline
MySQL, Python (Pandas), Power BI

- Designed relational schema for customer, product, and transaction datasets.
- Developed analytical SQL queries using JOINs, CTEs, and window functions for revenue and cohort analysis.
- Implemented Pareto (80/20) analysis to identify high-value customer segments.
- Performed data cleaning and transformation using Pandas before loading into MySQL.
- Built interactive Power BI dashboards to track revenue trends, product performance, and customer retention.

Gen-AI Document QA System (RAG-based Data Pipeline)
Python, LangChain, ChromaDB, LLM

- Built end-to-end data pipeline to ingest, chunk, embed, and store documents in a vector database.
- Implemented similarity-based retrieval (Top-K) for contextual response generation.
- Designed modular architecture separating ingestion, embedding, and query layers.
- Improved retrieval precision through chunk-size and embedding optimization.

Alzheimer Disease Detection using CNN
Python, TensorFlow

- Preprocessed 6000+ MRI scans using normalization and augmentation techniques.
- Implemented CNN model with 5-fold cross-validation for robust evaluation.
- Tuned hyperparameters reducing overfitting by 25%.
- Achieved 92% classification accuracy; research accepted for publication in Springer.

CERTIFICATIONS

- Google Cloud – Generative AI Virtual Internship
- AWS Academy Graduate – Cloud Foundations
- Languages: English (Fluent), Hindi (Fluent)