Ganesh Sharma

Mumbai, Maharashtra
ganesh27sharma09@gmail.com
+91 90760 99303
https://www.linkedin.com/in/ganeshsharmaa/
https://github.com/GaneshSharmaa

Professional Summary

Final-year Computer Science (AI/ML) student with hands-on experience in Python, data analysis, and machine learning. Skilled in exploratory data analysis (EDA), data preprocessing, and building ML-based projects.

Education

B.E. in Computer Science & Engineering (AI/ML)
Bharat College of Engineering
2022 – 2026 (Expected)
CGPA: 6.69

Certifications/Courses

- CS50: Introduction to Programming with Python — Harvard (edX)
- Introduction to Data Science in Python — University of Michigan (Coursera)
- Machine Learning Crash Course — Google Developers
- DevOps Fundamentals (Git, Docker, FastAPI) — FreeCodeCamp

Technical Skills

Programming: Python, C/C++
Data Analysis: Pandas, NumPy, Matplotlib
Machine Learning: Linear Regression, Logistic Regression, Classification, Model Evaluation (Accuracy, Precision, Recall, F1-score)
Tools & Frameworks: Scikit-Learn, FastAPI, Git, GitHub, Docker, Jupyter Notebook, VS Code

Projects

SIH Hackathon Project: Unified Multi-modal RAG framework
Python, Pandas, NumPy, Matplotlib, FastAPI, Streamlit

- Engineered a high-performance, 100% offline Retrieval-Augmented Generation (RAG) system to ingest and query diverse data formats including PDFs, DOCX, images (OCR), and audio recordings.
- Developed a Unified Semantic Search pipeline by aligning text and visual embeddings into a shared vector space using OpenAI's CLIP, enabling cross-modal retrieval (e.g., text-to-image search).
- Integrated OpenAI's Whisper for automated speech-to-text transcription, enabling semantic indexing of voice recordings with precise timestamp-based citations.
- Designed a transparent citation engine that links generated responses to source metadata, reducing model hallucinations and improving information reliability in enterprise workflows.
- Optimized local LLM performance using Quantization (GGUF/AWQ) to run Llama 3 8B on consumer-grade hardware, ensuring data privacy and zero-latency dependence on external APIs.

Credit Risk Prediction model
Python, Pandas, NumPy, Matplotlib, FastAPI, Streamlit

- Developed an end-to-end ML pipeline to classify loan applicants as High or Low Risk.
- Engineered features and handled preprocessing including encoding, scaling, and imbalance mitigation.
- Deployed model as a REST API (FastAPI) and built an interactive frontend (Streamlit).
- Implemented consistent training–inference feature schema to ensure reliable production predictions.

MITRA AI Chatbot
Python, PyTorch

- Built a conversational AI chatbot by integrating a locally hosted open-source language model.
- Focused on understanding model integration, prompt handling, and response flow.

Plagiarism Detector Application
Python, FastAPI, Scikit-Learn

- Developed a REST API using FastAPI to accept text input and return similarity scores.
- Implemented a machine learning-based text comparison model and evaluated performance using confusion matrix and accuracy.