Shubham Gajanan Tade

Pune, India | Shubhamgtade123@gmail.com | +91 7030409591 | GitHub | LinkedIn

Education

Dr. Babasaheb Ambedkar Technological University — Lonere, India
B.Tech. in Computer Science & Engineering | CGPA: 7.48 — 2020-2024

Technical Skills

- **Languages:** Python, SQL
- **Machine Learning:** Supervised Learning, Unsupervised Learning, Data Preprocessing, Model Evaluation
- **Deep Learning:** ANN, CNN, RNN, LSTM, GRU, Encoder – Decoder, Transformers
- **GenAI:** RAG, LLM, Vector Database, FAISS, Embeddings
- **Data Analytics:** Advanced Excel, Power BI, EDA, Data Cleaning, Correlation, Distribution
- **NLP:** Text Classification, Tokenization, Named Entity Recognition (NER)
- **Tools & Libraries:** LangChain, NumPy, Pandas, Seaborn, Sci-kit-learn, TensorFlow, NLTK, Git, Streamlit

Work Experience

AI ML Engineer | PandoAI Solutions Pvt. Ltd. (Healthcare Startup - caresila.com) — **Jul 2025 – Present**

- Cleaned and organized large-scale dataset of 10,000+ hospitals using Python and Excel, filling in missing details such as specialties, services, doctors and beds count, and geolocation information.
- Planned automated workflows to collect hospital details from **Practo** and official government health portals, compiling a database of over **1,00,000+ hospitals across India**.
- Linked **NABH** and **MJP_JAY** accreditation and Health Assurance data with hospital records to improve completeness and accuracy.
- Developed and launched the **Caresila Hospital Portal** using **Supabase** and **Vercel**, featuring verified hospital data, maps, and interactive pages for easy hospital discovery.
- Used **GenAI and automation strategies to resolve complex business challenges** to streamline workflows, optimize data preparation, and bring innovative ideas into production.

***Live Project**: https://caresila-hospital-portal.vercel.app/*

Projects

ST-GPT: Multi-Modal GenAI Companion | Generative AI & RAG — GitHub Repository
Python, LangChain, Streamlit, FAISS, Llama 3.3, Google Gemini

- Architected a **hybrid LLM system** combining **Llama 3.3** (via Groq) for high-speed chat and **Gemini Flash** for processing long documents.
- **Implemented RAG** (Retrieval-Augmented Generation) pipelines to synthesize insights from **2+ Hour** YouTube Transcripts and **100+ page** PDF documents, **utilizing FAISS for efficient semantic retrieval.**
- Engineered strict **System Prompts** (Guardrails) validated against **20+ adversarial queries** to prevent factual errors (hallucinations) and ensure highly engaging, context-aware interactions.

***Live Project**: https://study-buddy-st-ai.streamlit.app/*

Bank Loan Approval Classification | Machine Learning — GitHub Repository
Jupyter Notebook, Python, NumPy, Pandas, Seaborn, Scikit-Learn

- Designed predictive model improving bank loan approval **accuracy to 98% using 61,000+ financial records**.
- **Analyzed 61,000+ records**, removed outliers, balanced data, and standardized features to enhance the model.
- Prepared a stacking model with Decision Tree, RF, Logistic Regression, and others, **reaching 98% accuracy**.

Certification

Mastering in Data Science Feb 2025
3RI Technology, Pune

- Gained expertise in Python, ML, Deep Learning, and Data Visualization (Power BI, Excel) through capstone projects on predictive modeling and AI solutions.

Achievements

- **Won 2nd place in a college-level project** competition for developing a machine learning model.
- Led a team for an SIH-level IoT project during college and published a research paper based on the solution.