Sanya Shourya

📞 +91 9263483611   ✉ sanyashourya1803@gmail.com   📍 Bihar, India   in LinkedIn   GitHub

Education

VIT Bhopal University, Bhopal, India *2022 – Present*
B.Tech in Computer Science and Engineering (AI & ML), CGPA: 8.82

Technical Skills

Programming Languages: Python, Java, MySQL
Libraries & Tools: TensorFlow, OpenCV, Pandas, Streamlit, Excel, Git
Concepts: Machine Learning, Deep Learning, EDA, NLP, Data Visualization

Projects

Sales Data Analysis using SQL and Python *2024*
- Designed and implemented a pipeline to clean and preprocess 10,000+ sales records using Python (Pandas) in Jupyter Notebook.
- Loaded transformed data into a MySQL database with 5+ linked tables, enabling structured analysis and storage.
- Executed 20+ advanced SQL queries to extract insights such as top 3 categories and peak revenue months.
- Used dotenv and SQLAlchemy to securely manage credentials and build a modular codebase with 100% script automation.

Next Word Prediction using LSTM *2025*
- Built a word-level next word prediction model using TensorFlow and LSTM, trained on a dataset of 5,000+ motivational and philosophical quotes.
- Applied custom tokenization and generated 10,000+ n-gram sequences; limited vocabulary to 3,000 tokens to optimize memory usage.
- Designed a 3-layer model (Embedding, LSTM, Dense with Softmax) and trained it for 30 epochs using the Adam optimizer and categorical cross-entropy loss.
- Achieved 66% prediction accuracy; model can generate up to 15 semantically relevant words based on a user-provided input phrase.

Conversational PDF Q&A using RAG + Streamlit *2025*
- Developed a Conversational Retrieval-Augmented Generation (RAG) app that enables users to ask questions over uploaded PDFs using Streamlit.
- Integrated LangChain with HuggingFace embeddings and Groq's LLM (Gemma2-9b-It) to generate context-aware, standalone answers.
- Built a dynamic session-based chat system with memory using RunnableWithMessageHistory and message persistence.
- Incorporated PDF parsing, chunked embedding with ChromaDB, secure .env-based API handling, and a user-friendly web UI.

Extracurricular Activities

- Attended multiple technical workshops, coding competitions, and seminars hosted by student clubs.
- Participated in guest lectures on Machine Learning, Cloud Computing, Cybersecurity, and Tech Innovation.
- Preparing for the JLPT N5 (Japanese Language Proficiency Test) to expand linguistic and cultural competency.

Languages

English: Proficient   **Hindi:** Native/Bilingual   **Japanese:** Basic