SIDDHANT MAHAJAN

+917710983852 | sidumahajan@gmail.com | LinkedIn | GitHub

EDUCATION

VIT Bhopal University (Aug, 2022 – Expected Jul, 2026)

B. Tech CSE with Specialization in AIML | CGPA: 8.72

SKILLS

Languages: Strong grasp on Python, C++; familiarity with Java, SQL in MySQL.

Experience: Data Structures and Object-Oriented Programming, Machine learning and Deep learning using Scikit-learn, Pandas, TensorFlow, Rapids; Basic Data Analytics using MS Excel, Dataiku, Tableau, Version Control systems (Git).

PROJECTS

DocuSummary (July, 2025)

Python, Transformers, VLMs, Generative AI

- Engineered a Retrieval-Augmented Generation (RAG) application to streamline data extraction and contextual query answering from PDF documents using LLMs.
- Leveraged Qwen 2.5 Instruct, Qwen 3 Embeddings, and the InternVL3 Vision-Language Model for high-accuracy textual data extraction and description.
- Implemented model quantization techniques to optimize the RAG pipeline, reducing hardware dependency and enhancing inference speed.

Brain Cancer Detection using Fine-tuned CNNs (May, 2025)

Python, TensorFlow

- Engineered a Retrieval-Augmented Generation (RAG) application to streamline data extraction and contextual query answering from PDF documents using LLMs.
- Leveraged Qwen 2.5 Instruct, Qwen 3 Embeddings, and the InternVL3 Vision-Language Model for high-accuracy textual data extraction and description.
- Implemented model quantization techniques to optimize the RAG pipeline, reducing hardware dependency and enhancing inference speed.

Lungs Disease Detection through Audio Recordings (April, 2025)

Python, TensorFlow, Librosa for Audio

- Created a CNN-RNN model to accurately map and identify 8 classes of diseases using audio samples.
- Extraction of audio features using MFCC coefficients, allowing for feature extraction using single dimension CNN filter. The output is then passed as a time series data to RNN layers, providing robust pattern identification and recognition capabilities.

CO – CURRICULARS AND EXTRA – CURRICULARS

CO – CURRICULARS

Sabudh Foundation – STPI, India (January 2025 – June, 2025)

- Gained hands-on experience applying machine learning, deep learning, and Natural Language Processing to solve real-world data problems.
- Utilized Dataiku for comprehensive data wrangling, model training, and deployment, improving the efficiency of data science workflows.

Certificates

- OCI AI Foundations – Oracle Education Academy
- Machine Learning with TensorFlow – FreeCodeCamp
- Applied Machine Learning in Python – Coursera
- Dataiku Developer Certificates – Dataiku

EXTRA – CURRICULARS

National Service Scheme - VIT Bhopal University

- Spearheaded planning for women's empowerment events during 'International Women's Day', resulting in increased local participation and heightened awareness of gender equality issues among villagers.
- Part of NSS – VIT Bhopal 10,000 trees plantation drive, managing a team of 10 volunteers along with coordinating with other 25 teams to make the drive successful.

ADDITIONAL INFORMATION

Languages: English (Professional working proficiency), Hindi and Marathi (Native professional proficiency)