Shraddhesh Bhalerao

+91-9359216428 | shraddheshbhalerao10@gmail.com | LinkedIn | GitHub

I'm a passionate data science graduate with a strong interest in machine learning and data analytics. I have hands-on experience working on projects using Python, RAG, and LangChain to build AI chatbots and tools that generate useful responses from documents. I enjoy turning data into useful insights and am eager to learn and grow in areas such as real-time data analysis and generative AI.

Education

Symbiosis Institute of Technology, Pune
B. Tech, Computer Science & Engineering — *June 2021 – June 2025*

Atomic Energy Central School - No.1, Tarapur
Junior College — *Apr 2019 – June 2021*

Skills

- **Languages:** Python | R | SQL
- **Tools/Libraries:** VS Code | Git | RAG | LangChain | Vector Database | PowerBI | Google Workspace
- **Cloud & APIs:** Microsoft Azure | Graph API | OpenAI API | Slack API

Internship

Data Analytics Intern — *Feb 2025 –Present*
GBK Collective, New York

- Identified **key drivers** of customer satisfaction by analyzing survey data and resolving **missing values and outliers**.
- Built a **framework** to **optimize segmentation variables and algorithm selection** using VarSelLCM, PCA, and clustering methods like PAM and K-Means, **guiding** consultants to **the best-fit solution**.
- Developed an **AI Segmentation Assistant** using GPT-4o mini via **OpenAI API** to analyze and summarize segment profiles, reducing **manual cluster evaluation time by 7x.**
- Optimized a **market reach simulation package** by rewriting R loops in **C++ with Rcpp**, achieving **5x faster compilation** and enabling **quicker result** turnaround.

Projects

AI Research Assistant — *Dec 2024 – Feb 2025*

- Built a **RAG-based AI assistant** for summarizing scientific documents using **LangChain**, **LLM**, and a **vector database**, improving access to key insights.
- Improved document processing with LangChain's **text splitting and embedding**, enhancing search speed and response quality.
- Designed a **real-time chatbot interface** for research queries, scaling down document review time and improving research speed.

Sleep Disorder Prediction

- Achieved 89% accuracy in predicting types of sleep disorders by **designing and training** classification models such as **Decision Tree** and **Random Forest** using a dataset of 400 patient lifestyle and health profiles.
- Improved model **precision and recall** by 5% over baseline Decision Tree by engineering features like systolic/diastolic blood pressure and converting **categorical data** into **numerical format** for better sleep disorder classification.
- **Analyzed** 13 lifestyle and medical factors such as BMI, stress, and activity levels to **identify top predictors** of sleep disorders, revealing that gender, occupation, and BMI had the highest impact on model performance.

Certifications

- Coursera – IBM Data Science Professional Certificate
- Coursera – Generative AI Engineering with LLMs Specialization

Achievements

- Gold Medal - SIT PROTA Table Tennis
- Academic Volunteer - Z.P School, Pasthal