ADITYA SANJAY PAWAR

+1 680-356-8183 | adityapawar1020@gmail.com | www.linkedin.com/in/aditya-pawar-825830133

EDUCATION

Syracuse University | School of Information Studies, Syracuse, NY **05/2025**
M.S. in Applied Data Science. GPA: 3.7/4.0
Relevant courses: *Database Management Systems, Applied Machine Learning, Building RAG Systems, Datawarehouse*

Savitribai Phule Pune University | Vishwakarma Institute of Technology | Pune, India **05/2022**
B. Tech Mechanical Engineering
Relevant courses: *Python, MATLAB, Data Analysis for computational fluid dynamics*

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, R, Matlab
- **Database Management:** MySQL, Snowflake DW, Chroma DB Vector Database, Azure Data Studio, SSMS
- **Machine Learning and AI:** Pytorch, Hugging Face Transformers, Fine-Tuning LLMs, Chatbot API (OpenAI, Gemini)
- **Data Visualization & Applications:** Tableau, Streamlit AI Application

WORK EXPERIENCE

Data Science Co-op | Quantum Integrators, New Jersey **01/2025 – Present**
- Building Gen AI projects for a pharmaceutical company, implementing RAG pipelines, Langchain using multiple Gemini APIs.
- These projects are being deployed on GCP using microservices such as Vertex AI, Agent Builder, Big Query etc.
- Load testing these applications for handling load, stress, spikes etc metrics using K6 evaluating the number of users it can handle.

Data Science Intern | Algoanalytics Pune, India **05/2022 – 09/2022**
- Formulated approaches for a Generative Q&A Pipeline for answering questions related to the National Stock Exchange.
- Used multiple approaches like testing keyword search ranking algorithms (BM25, TF-IDE), tweaking the parameters of the reader and the retrievers, and combining sparse and dense retrievers (BERT, RoBERTa, DistilBERT).
- Evaluated RAG and S2S pipeline methods where S2S performed better than RAG Pipeline.
- Applied Python, Haystack framework, Hugging Face models, transformers, FAISS, Pytorch, and Elasticsearch on this product.
- Engaged with other data scientists to produce actionable insights and learned innovative methods influencing the product.

Data Science Intern | Artificial Intelligence Shepherds, Munich, Germany **07/2021 – 12/2021**
- Created a database for use cases of AI in Industrial engineering, automotive engineering, and electrical engineering.
- Presented the use case database of the AI research to clients and their team while coordinating with my CEO.
- Built a classification neural network for a client to predict the type of rotor failure from the data collected from heat sensors, temperature sensors, and vibration sensors increasing the accuracy by ~15% using traditional ML Algorithms, Pytorch and statistics.

PROJECTS

Streamlit AI Travel assistant powered by OpenAI APIs *(link)* **08/2024 – 12/2024**
- Chatbot takes inputs in text, audio and image for a traveller to ask queries regarding language translation, image description.
- Chatbot responds in text/audio enabling a continuous conversation in 2 different languages removing the language barrier.
- The chatbot has an integrated RAG pipeline which retrieves information for relevant travel documents from a Vector DB.
- Other chatbot features include weather, location, map, currency converter, live translation and live conversation.

Snowflake ETL Data Warehouse Project *(link)* **07/2024 – 12/2024**
- Designed ETL pipeline using Snowflake, DB, GitHub to transform FudgeMart_v-3 database into a structured data warehouse.
- Conducted data profiling and dimensional modeling to analyze data quality and design schema for the business use cases.
- Wrote SQL scripts Snowflake notebooks and configured DBT transformations to automate ETL processes.
- Created a Tableau dashboard to visualize key insights and outcomes derived from the business case.

A/B Testing for reward placement in a game using statistics *(link)* **05/2024 – 08/2024**
- Used Python to analyse the user behaviour and psychology for 2 different versions of the same game for A/B testing.
- The data included the retention and the actions taken by the user on both the versions of the game.
- Python and statistics concluded that there is strong evidence that 7-day retention is higher when the gate is at level 30.

Natural Language Classification for text classification of Bank Complaints *(link)* **09/2023 – 12/2023**
Data modeling and analysis of movie production expenditure *(link)* **09/2023 – 12/2023**
Machine Learning optimization using R *(link)* **09/2023 – 12/2023**