Shaheen Nabi Data Scientist

✉ ishaheennabi333@gmail.com 📞 6005081860 📍 Bengaluru karnataka in linkedin.com/in/shaheennabi

○ github.com/shaheennabi ◉ https://huggingface.co/devshaheen ◈ https://leetcode.com/u/ishaheen333/

🪪 Profile

Data Scientist with **3 months of internship experience**, skilled in **taking ideas from concept to deployment**, building **production-ready systems using cloud platforms**. With a **deep understanding of algorithms** and a **strong theoretical foundation** in **ML, DL, CV, NLP, Generative AI**. I develop projects that leverage **industry-standard MLOps tools** to automate the entire lifecycle, from **versioning to post-production monitoring**. Having **practical expertise in LLMs, RAG agentic workflows, and model inference on the cloud**, I deliver **AI solutions directly to customers** and using **design principles and patterns** to write **modular, maintainable code** that ensures **scalability**.

🧠 Skills

Python: Scripting / OOPs • **Deep Learning:** Perceptron / ANN / Optimizers / Activation Functions •
Computer Vision: CNN / YOLO / Image Processing / Object Classification / Object Detection •
Natural Language Processing: Text-Preprocessing / RNN / LSTM / Embeddings / Transformers / BERT / GPT •
MlOps: DVC / Mlflow / DagsHub / Docker / Apache Airflow / GitHub Actions / CircleCI / Jenkins • **Databases:** MySQL /
MongoDB / Weaviate / Pinecone / FAISS • **Cloud:** (AWS | S3, EC2, ECR, Lambda, API Gateway, Sagemaker) •
Generative AI: Langchain / Llama-Index / OpenAI API / RAG / Agentic Workflows / Finetuning / RLHF •
Libraries/Frameworks: Sklearn / Keras / Tensorflow / PyTorch / OpenCV

💼 Professional Experience

Ineuron Intelligence, *Data Science Intern* 🔗   01/2025 – 03/2025 │ Bengaluru, Karnataka
Tech: Python, Pytorch, Jenkins, AWS, Taskflowai, OpenAI, SerperAPi, A100 GPU's, HuggingFace.
Internship description:
Developed a production-ready object detection system using **YOLOv5**, capable of accurately identifying and classifying over **100** plant and crop species, addressing the client's need for an automated agricultural monitoring solution.

Annotated **25,000 images**, later **open-sourced** on **Hugging Face**, contributing to the broader research community and enhancing model accessibility.

Designed and deployed a fully automated **AI agents pipeline** using the **TaskflowAI** framework, enabling efficient post-detection research by retrieving enriched **crop details** from various online sources.

Trained the model on **NVIDIA A100 GPUs**, ensuring high accuracy and performance for real-world deployment.

Maintained a **modular code structure** following best software engineering practices, ensuring scalability, reusability, and ease of maintenance while incorporating robust testing and evaluation mechanisms.

Deployed the solution using **Jenkins**, **AWS ECR**, and **EC2**, providing the client with a scalable, cloud-based infrastructure for real-time inference and decision-making.

Integrated an **SMTP service** to enable automated **email delivery** of **summarized reports**, allowing users to receive personalized 1 minute automated research report directly in their inbox upon entering their email on the UI.   *GitHub Link* 🔗

📂 Projects

Production-Ready TripPlanner Multi-AI Agents Project 🔗   01/2025 – 02/2025
Tech: Python, Taskflowai, OpenAI, GitHub Actions, ECR, EC2, Streamlit, SerperAPI.
Project Description:
Developed an intelligent AI-driven travel assistant, integrating three specialized agents to automate and optimize the travel planning experience.

Web Research Agent – Extracts real-time travel insights, including local attractions, events, weather conditions, and travel advisories using **Wiki API**, **Serper API, Weather API, and Amadeus API**, combined with advanced web scraping techniques.

Travel Planner Agent – Generates personalized itineraries, optimizes travel routes, and suggests appropriate clothing based on live weather forecasts, ensuring a seamless and well-prepared journey.

Report Generator Agent – Leverages **GPT-3.5 Turbo** to compile structured, actionable travel reports, enhancing trip planning with concise recommendations.

Tech Stack & Deployment – Built using the **TaskflowAI framework** for streamlined agent coordination and automation. Dockerized the application for **scalability and portability**, integrating a **CI/CD pipeline** with **GitHub Actions** for automated builds and deployments to **AWS Elastic Container Registry (ECR) and EC2**, ensuring efficient and reliable deployment. *GitHub Link* ⊘

Production-Ready Instruction Fine-Tuning of Meta LLaMA 3.2 3B 11/2024 – 01/2025
Instruct Project ⊘
Tech: Python, Unsloth, Qlora, Instruction-finetuning, SFT, HuggingFace, Transformers, TRL, Quantization, AWS
Project Description:
Embarked on an advanced NLP initiative to fine-tune the **LLaMA 3.2-3B Instruct model**, enhancing its ability to understand and respond to **Kannada-specific queries** for a large Kannada-speaking audience.

Primary **goal** was to fine-tune the model to improve its **contextual awareness** and generate culturally relevant responses in **Kannada**, ensuring a seamless and natural user experience.

Utilized the Kannada Instruct dataset (**charanhu/kannada-instruct-dataset-390k**) from Hugging Face, enabling the model to develop better instruction-following capabilities in low-resource languages.

Optimized GPU usage by implementing **4-bit QLoRA quantization** and fine-tuning the model on Google Colab **Tesla T4 GPUs**, efficiently addressing computational constraints.

Applied advanced **prompt engineering techniques** to evaluate the model's instruction-following ability, refining response accuracy and contextual awareness.

Combined the **base model** with the **fine-tuned** version in **16-bit floating point (FP16)**, ensuring efficient memory usage while maintaining model precision and performance.

Pushed the fine-tuned model to the **Hugging Face Hub**, making it publicly available for researchers and developers to leverage in Kannada language applications. *GitHub Link* ⊘

🧩 My Open-Source Work

Crops/plants 25k annotated object detection dataset ⊘
  Description:
Open-sourced a fully annotated dataset of **25,000 images** on **Hugging Face**, designed for detecting 100 different crops and plants.

Annotated using **Roboflow**, ensuring high-quality labels optimized for object detection and classification tasks.

Downloaded by hundreds of developers worldwide within months, actively being used for training AI models in agriculture and plant identification.

Adopted by government agricultural departments and a large community of developers, leveraging it to build advanced AI solutions for crop monitoring and precision farming. *HuggingFace Link* ⊘

db-crud-automated pypi package ⊘
  Description:
Developed db-crud-automated, a **PyPI package** designed to simplify and automate database operations across **MongoDB**, **MySQL**, and **Cassandra**, providing a unified and seamless database interaction experience.

Built an abstraction layer that eliminates the complexity of database connections and **CRUD operations**, enabling developers to interact with databases using minimal code.

Integrated bulk **data insertion**, schema validation, and indexing optimizations, ensuring efficient and scalable data handling for production-ready applications.

Implemented modular architecture, allowing users to extend support for additional databases with ease. *Pypi Package Link* ⊘

🎓 Education

Indira Gandhi National Open University, *Bachelor of Arts- English (BA)* ⊘ 2025 – 2028 | Srinagar

Ineuron Intelligence, *1 Year Course, Full StackData Science* ⊘ 10/2021 – 10/2022

Y.K Pora Higher Secondary School Anantnag, *High SchoolNon-Medical* ⊘ 04/2021 – 10/2023