

SKILLS

- **Programming Languages:** Python, SQL, Bash
- **Data Science, ML Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, XGBoost, LightGBM
- **Deep Learning, NLP:** PyTorch, TensorFlow, Hugging Face Transformers, LangChain, LangGraph, OpenAI API
- **MLOps and Deployment:** MLflow, DVC, Docker, FastAPI, Flask, Jenkins, CircleCI, GitHub Actions, Google Cloud Run, GCP (GCS, GKE), Airflow, Prometheus, Grafana, Dagshub, BentoML
- **Vector DBs and GenAI Stack:** ChromaDB, Pinecone, Tavily API, Wikipedia API, YouTubeTranscript API, Groq, Ollama, AstraDB
- **Dev Tools and Others:** Git, GitHub, VS Code, Streamlit, Gradio, REST APIs, PostMan, MCP
- **Soft Skills:** Communication, Problem-Solving, Collaboration, Rapid Prototyping

PROJECTS

Hotel Reservation Cancellation Prediction

- Built a modular **MLOps pipeline** to predict hotel booking cancellations with **88% accuracy**.
- Automated model training and deployment using **Jenkins, Docker**, and **Flask API** on **GCP Cloud Run**.
- Tracked experiments using **MLflow** and ingested data securely from **Google Cloud Storage**.
- Integrated **CI/CD** with robust exception handling and performance logging.

LangGraph Report Generator (Multi-Step RAG)

- Built a **LangGraph**-based agentic pipeline for generating structured AI reports from queries.
- Integrated **Groq + LLaMA3-70B** for fast generation with node-level control and memory.
- Created custom retrieval + summarization nodes using **LangChain**'s Runnable framework.
- Enabled dynamic branching and multi-language output via configurable routes.

Abstractive Text Summarizer (Pegasus)

- Fine-tuned **Pegasus model** using **Hugging Face** on dialogue data (Samsun) for abstractive summarization.
- Developed a full **MLOps pipeline** with YAML-based config, modular code, and reusable stages.
- Achieved **ROUGE-L score of 42+** with evaluation tracked using **Hugging Face metrics**.
- Included prediction script and **Streamlit** interface for summarizing unseen text.

IMDB Movie Review Sentiment Classifier (RNN)

- Trained a Simple **RNN** on IMDB dataset, achieving **85% accuracy** on sentiment classification.
- Deployed real-time predictions using a **Streamlit** web app with word cloud visualization.
- Handled text preprocessing with tokenizer, padding, and embedding layers in **Keras**.
- Used Sigmoid-activated Dense layer for binary output and binary cross-entropy loss.

INTERNSHIP

Machine Learning Intern

Dec 2024 — Jun 2025

Unified Mentor

Noida

- Built across 7 end-to-end classification/regression projects using real-world datasets.
- Applied preprocessing, feature engineering, and scaling to clean messy datasets, boosting model accuracy by **up to 25%**.
- Tuned hyperparameters with **GridSearchCV**, improving **RF/XGBoost** models to **94%+ accuracy** in multiclass tasks.
- Engineered metrics pipelines using confusion matrices, ROC AUC, MSE, and R2 for robust model evaluation.
- Achieved highest test accuracy of **98.16%** in Thyroid Disease prediction using ensemble techniques.
- Technologies: **Python, Pandas, Scikit-learn, XGBoost, Matplotlib, StandardScaler, LabelEncoder, ROC curves**.

CERTIFICATIONS

- Machine Learning, NLP, Transformers - Udemy
- Deep Learning Specialization – DeepLearning.AI
- Google Data Analytics – Coursera
- MLOps Fundamentals – Udemy
- Generative AI (LangChain) – Udemy

EDUCATION

Amity University, Greater Noida

Jul 2023 – Jul 2025

Master of Computer Applications (MCA)

AlmaBetter (Data Science and AI)

Dec 2024 – Present

Certification Program in Data Science and AI