

GANGUMALLA HEMA MALINI

hemamalini07g@gmail.com • +917396248629 • [linkedin.com/in/hemamalini07](https://www.linkedin.com/in/hemamalini07)
github.com/hemamalini0708

Data Science with hands-on experience in Python, Machine Learning, NLP, and Generative AI. Skilled in end-to-end ML workflows, prompt engineering, and deploying AI solutions. Strong communication and collaborative skills with proven ability to design innovative projects using LLMs and core ML libraries.

SKILLS

- Programming & Development:** Python, SQL, PostgreSQL
- Machine Learning & Deep Learning:** Scikit-Learn, TensorFlow, Keras, PyTorch, Random Forest, CNNs, RNNs, LSTM, Transfer Learning (ResNet, VGG16)
- Generative AI & NLP:** Hugging Face, LangChain, RAG Pipelines, LLM Applications (OpenAI API, LLaMA, Mistral), Prompt Engineering
- Data Science & Statistics:** Data Analysis, Feature Engineering, Model Evaluation, Hyperparameter Tuning, Cross-Validation
- Deployment & Tools:** FastAPI, Flask, Streamlit, Git, REST APIs, Jupyter Notebook
- Other:** Communication Skills, Documentation, Collaboration, Problem Solving, Metrics Tracking

WORK EXPERIENCE

- Data Science Intern**
Vihara Tech

Jan 2025 - Dec 2025
Hyderabad, India

 - Designed and deployed end-to-end ML pipelines including preprocessing, feature engineering, model training, and evaluation.
 - Built classification models with hyperparameter tuning and cross-validation; improved reliability using SMOTE.
 - Developed predictive solutions and deployed interactive applications using Streamlit and FastAPI.
 - Collaborated with stakeholders to document workflows and communicate insights effectively.

PROJECTS

- Sentiment Analysis (Bi-Directional RNN)** [🔗](#)
 - Designed and tuned RNN model in TensorFlow/Keras; deployed via Flask REST API.
 - Implemented a Bi-Directional RNN with embedding and masking layers in TensorFlow/Keras, trained on 7K/2K/1K train-validation-test splits with Adam optimizer over 20 epochs.
 - Deployed a Flask web application for real-time sentiment inference with confidence scoring and consistent preprocessing.
- AI PDF Q&A Chatbot (RAG + LLM)** [🔗](#)
 - Designed and deployed chatbot using LangChain, Hugging Face embeddings, ChromaDB, and Groq-hosted LLaMA 3.3 70B; implemented knowledge graph-based retrieval.
 - Implemented a complete RAG workflow including document chunking, vector embedding, semantic retrieval, prompt construction, and response generation.
 - Enabled source-grounded answers with page-level references from uploaded PDF documents.
 - Deployed the application using Streamlit.
- Credit Risk Prediction** [🔗](#)
 - Developed classification pipeline on 150K+ records; achieved 89.1% accuracy with Random Forest; deployed via Streamlit.
 - Applied SMOTE, feature engineering, model evaluation, and hyperparameter tuning techniques.
 - Compared multiple algorithms including Logistic Regression, KNN, Naive Bayes, Decision Tree, and Random Forest.
 - Achieved 89.1% accuracy and AUC of 0.818 using Random Forest and deployed the solution as a real-time prediction application.
- NLPSQL Engine** [🔗](#)
 - Built offline NL-to-SQL engine using Ollama and pgvector; integrated embeddings for query validation and optimization.
 - Converted natural language queries into validated SQL statements using schema-aware retrieval and local LLM inference.
 - Implemented query validation, self-healing SQL correction, and query logging functionalities.
 - Designed a scalable retrieval workflow capable of handling datasets exceeding 100K records.

EDUCATION

- Bachelor of Science in Computer Science**
ASN Degree College for Women
- Jan 2020 - Sep 2023
GPA: 8.98

CERTIFICATIONS

- SQL** [🔗](#) by HackerRank
- Feb 2026
- Machine Learning with Python** [🔗](#) by IBM
- Sep 2025
- Excel** [🔗](#) by Coursera
- Dec 2023
- Data Scientist** [🔗](#) by Vihara Tech