

V.Madhuvamsi

🐙GitHub / ✉madhuvamsi135@gmail.com / 🔗LinkedIn / 📞+91 8688361638

Education

- | | |
|---|-----------------------|
| Indian Institute of Technology Guwahati , MTech in Mechanical | July 2023 – July 2025 |
| • GPA: 8.69/10.0 | |
| Andhra University college of Engineering , BTech in Mechanical | Aug 2019 – May 2023 |
| • GPA: 8.33/10.0 | |

Experience

- | | |
|---|----------------------|
| Data science intern , Afame Technologies (virtual) | May 2024 – June 2024 |
| <ul style="list-style-type: none">Completed a Data Science internship at Afame Technologies, working on a Sales prediction project using Python. Performed feature engineering, EDA, and feature scaling and normalization, applying a Linear Regression model that achieved a MAPE of 0.093.Further optimized the model with GridSearchCV, improving performance with an RMSE of 1.46 and an R^2 value of 0.92, indicating strong predictive accuracy. | |
| Data Analytics Intern , YBI Foundation (virtual) | May 2024 |
| <ul style="list-style-type: none">Completed a Bank Customer Churn Prediction Modeling project during an online data analytics internship, utilizing data encoding, feature scaling, and techniques like random over/under sampling.Applied a Support Vector Machine classifier, achieving an initial accuracy of 75%, which improved to 92% after hyperparameter tuning using GridSearchCV.Enhanced recall from 0.74 to 0.97 and F1-score from 0.76 to 0.92 through model optimization. | |

Projects

Graph DB with Langchain – Q&A Application | GitHub

- Developed a Q&A application using LangChain and Groq LLM to query a Neo4j graph database with natural language. Implemented dynamic Cypher query generation using the Gemma2-9b-It model to retrieve relevant information from a movie dataset.
- Integrated LangChain's GraphCypherQChain to streamline query translation and execution. Designed an efficient workflow to process user input, generate queries, fetch results, and deliver accurate responses.

RAG Document Q&A with NVIDIA NIM | GitHub

Feb 2025

- Built a RAG-based Document Q&A system using NVIDIA NIM microservices & APIs, LangChain, and Streamlit for efficient document retrieval and response generation.
- Integrated NVIDIA's ChatNVIDIA model (Llama-3.3-70b-instruct) and NVIDIAEmbeddings to process and store vector embeddings in a FAISS vector database.
- Implemented document loading, chunking, and retrieval mechanisms to fetch relevant content dynamically.
- Developed a user-friendly Streamlit interface for seamless question answering, ensuring transparency by displaying retrieved document chunks.

Blog Generation system using AWS | GitHub

Feb 2025

- Built an end-to-end GenAI-powered blog generation system using AWS Lambda, AWS Bedrock, and S3. Integrated API Gateway to trigger Lambda, which invokes the Llama3-8B Instruct model for dynamic blog creation.
- Automated storage in S3 with timestamped filenames and logged execution details using CloudWatch.
- Designed a serverless, scalable, and efficient AI-driven content pipeline leveraging AWS services.

GenAI Content Summarization app | GitHub

Feb 2025

- Built a Streamlit-based GenAI app using LangChain and Groq's API to summarize YouTube videos and website content.
- It extracts transcripts via YouTubeLoader and fetches web data with UnstructuredURLLoader, then generates concise summaries using the Gemma2-9b-It model.

Chat with SQL DB using Langchain | GitHub

Jan 2025

- Developed a Streamlit-based GenAI app that enables users to query SQL databases (SQLite/MySQL) using natural language.
- Utilizing LangChain's SQLDatabaseToolkit and Groq's Llama3-8b-8192 model, the app translates user queries into SQL, retrieves data, and responds conversationally.
- It features dynamic database connections, real-time query execution, session-based chat history, and caching for efficiency.

Text Summarization with Transformers | GitHub

Jan 2025

- Developed an end-to-end NLP project for text summarization using Hugging Face's Transformers library, leveraging the Pegasus model fine-tuned on the Samsung/samsum dataset.
- Implemented a modular pipeline for data ingestion, transformation, training, and evaluation, using ROUGE scores to assess summarization quality.
- Deployed the model with FastAPI, enabling real-time summarization of dialogues through a user-friendly web interface.

Chicken Disease Classification | GitHub

Nov 2024

- Developed a deep learning model for chicken disease classification using fecal images, incorporating data ingestion, transformation, model training, and evaluation pipelines.
- Deployed the solution across multiple platforms, including Flask and Streamlit web apps, AWS EC2, and Azure, while leveraging MLOps tools like MLflow for tracking experiments, DVC for pipeline tracking.

Student Performance prediction | GitHub

August 2024

- Developed a student performance prediction project using Machine Learning with complete ML lifecycle from data ingestion to model deployment.
- Deployed the solution on AWS Beanstalk, EC2, and as a Flask web app, optimizing CI/CD pipelines with GitHub Actions to reduce manual intervention by 90%, and utilized Docker for deployment in ECR.
- Implemented comprehensive logging, exception handling, and created efficient training and prediction pipelines.

Technologies and Skills

DS/ML/AI: Machine learning algorithms, Deep learning, NLP, GenAI, LLMs

Tech Stack: Python, PyTorch, TensorFlow, Sklearn, Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, Docker, Github, NLTK, SpaCy, Streamlit, Flask, AWS Bedrock, Sagemaker, HuggingFace, PySpark, GCP, Transformers, Training and fine-tuning LLMs.

Languages: Python, SQL

Data Science: Hypothesis Testing, feature engineering, data preprocessing, model evaluation, supervised and unsupervised learning.

MLOps Tools: Airflow for work orchestration, MLflow, DVC, Github actions(CI/CD) for automating model deployment, DagsHub.

LLM Tools and Frameworks: Langchain, Llama Index, Ollama, Langsmith, Langflow, Vector databases, RAG, Graph Databases, ChromaDB, FAISS, Pinecone, AstraDB, OpenAI GPT, Claude, Groq APIs, Agentic AI frameworks.

Soft Skills: Problem-solving, Predictive analytics, Teamwork, Communication, Networking

Certifications

- AI for Real-World Applications – IIT KGP & AI4ICPS
- MLOps Bootcamp - udemy
- Python & MySQL Bootcamp – udemy
- Hands on Generative AI training – AGIE AI