

VEMULA GOWTHAM

Hyderabad, India | +91-7095960394 | vemulagowtham7@gmail.com
[linkedin.com/in/vemula-gowtham-624206286](https://www.linkedin.com/in/vemula-gowtham-624206286) | github.com/Gowtham12345292

PROFESSIONAL SUMMARY

Engineering graduate with about a year of hands-on experience building production-grade AI and machine learning solutions. I've built LLM-powered RAG pipelines and agentic workflows with LangChain and LangGraph, fine-tuned transformer models with Hugging Face, and deployed ML models on 5M+ records on Azure (Microsoft Fabric) using PySpark and Python. I build REST APIs with FastAPI, track experiments in MLflow, and communicate results through Streamlit dashboards to business stakeholders. Strong across the full AI/ML lifecycle — from data ingestion and feature engineering through model development, evaluation, and production deployment. Demonstrated portfolio of real AI/ML projects on GitHub.

TECHNICAL SKILLS

Languages: Python, SQL, PySpark

GenAI & LLMs: LLMs, RAG Pipelines, LangChain, LangGraph, Agentic Workflows, Prompt Engineering, Embeddings, Vector Databases (ChromaDB), Hugging Face Transformers, Model Fine-tuning, NLP, Text Classification, Tokenization

Machine Learning: Scikit-learn, LightGBM, Random Forest, Gradient Boosting, Regression, Classification, Model Evaluation, Feature Engineering, Hyperparameter Tuning, TensorFlow, PyTorch (familiar)

Engineering: FastAPI, REST APIs, Object-Oriented Programming, MLflow, Jupyter Notebooks, Git, GitHub, Agile

Data: Pandas, NumPy, Structured & Unstructured Data, Data Pipelines, Feature Engineering

Cloud & MLOps: Azure (Microsoft Fabric), MLflow, Docker (learning), AWS (in progress)

Visualization: Streamlit, Matplotlib, Seaborn

PROFESSIONAL EXPERIENCE

Data Science Intern

Jul 2025 – Apr 2026

Northern Tool + Equipment (US-based Company)

Hyderabad, India

- Designed and deployed end-to-end ML pipelines on Microsoft Fabric (Azure) using PySpark and Python to process **5M+** customer records, integrating structured and unstructured data for real-time inference.
- Built supervised ML models (Random Forest, LightGBM) for classification and propensity prediction, delivering a **~3 percentage-point** lift in buy rate and **2.12x** top-decile lift across **545K+** customers, validated with holdout testing and cross-validation.
- Built a brand-specific predictive model on 10 years of transaction data, delivering **17x** lift for VIP segments and **464% ROI** (\$98.8K revenue on \$17.5K spend).
- Ran iterative model evaluation, hyperparameter tuning, and feature engineering across **5M+** records using Python OOP-based pipelines in Pandas and SQL to improve accuracy.
- Collaborated with data engineers and business teams to translate requirements into production-ready AI/ML solutions, communicating findings to non-technical stakeholders.

Generative AI Intern

Jan 2025 – Mar 2025

Innomatics Research Labs

Hyderabad, India

- Built LLM-powered RAG pipelines for document Q&A using LangChain, ChromaDB, and prompt engineering, systematically evaluating outputs to cut model errors by **65%**.
- Fine-tuned a BERT transformer model with Hugging Face on 50K+ samples, reaching **87% F1-score** on text classification, served as a real-time REST API.

AI/ML PROJECT PORTFOLIO

ACIGE — Autonomous Customer Intelligence and Growth Engine | *Python, LightGBM, FastAPI, MLflow, Streamlit*

- End-to-end deployed ML system on **500K+** records — OOP-based pipeline covering data ingestion, feature engineering, model training, and evaluation reaching **PR-AUC of 0.74**, flagging **\$2.4M** in revenue-at-risk.
- Served via FastAPI REST API backend with automated MLflow experiment tracking and a Streamlit dashboard for business stakeholder reporting.

RAG Document Q&A System | *LangChain, LangGraph, ChromaDB, Embeddings, LLMs*

- Built LLM-powered RAG chatbot for document question-answering — document ingestion, chunking, embedding, vector storage (ChromaDB), and retrieval-augmented generation with LangChain and LangGraph.

Telugu Sentiment Analysis System | *Python, BERT, Hugging Face Transformers, TensorFlow*

- Fine-tuned BERT transformer on 50K+ samples reaching **87% F1-score** and **89% accuracy**, deployed as a real-time REST API.

EDUCATION

B.Tech, Electronics and Communication Engineering

2020 – 2024

Anurag Engineering College, Kodad, India

Professional Certification in Data Science

Jun 2024 – Jun 2025

Innomatics Research Labs, Hyderabad, India

CERTIFICATIONS

- Advanced Generative AI Certification — Innomatics Research Labs
- Advanced Data Science with Python — NASSCOM FutureSkills Prime
- Data Science Certification — Innomatics Research Labs