

Ankit Kumar

IIT Bombay | Data Scientist | LLMs | SQL | GenAI

Location: Gurugram, Haryana | Phone: +91 8923003333 | Email: ankitgeotek@gmail.com
LinkedIn: [linkedin.com/in/kaushal-ankit](https://www.linkedin.com/in/kaushal-ankit) | GitHub: github.com/ankitgeotek | Portfolio: mlwithankit.com

Summary: Data Scientist with expertise in **Machine Learning, NLP and Generative AI**. Proficient in **LLMs and Web Development**. Experienced in building **AI-powered document automation tools, scalable APIs, and cloud-based ML solutions**.

Skills:

- **Programming Languages:** Python
- **Web Development:** React.js, Streamlit
- **Large Language Models (LLMs):** LangChain, LangGraph, PyTorch, Hugging Face Transformers
- **Database Management:** SQL, Faiss, Milvus (Vector DB)
- **Python Libraries and tools:** Pandas, Matplotlib, Seaborn, Scikit-learn, NLTK, spaCy, FastAPI
- **Cloud & DevOps:** AWS (IAM, EC2, S3, Bedrock), Git, GitHub

Professional Experience

Miimansa AI | Data Scientist | Sep 2024 – Present

Agentic AI-Powered Financial Analysis Chatbot

- **Objective:** Developed AI chatbot for financial analysts to enable quick decision-making from complex PDFs, reducing analysis time from hours to minutes.
- **Approach:** Built MultiModal LLM pipeline for PDF extraction, Milvus vector database with 4-level indexing, & FastAPI backend.
- **Features:** Implemented LangGraph-based agentic workflows for intelligent query routing, parallel retrieval for complex charts/tables, and HyDE method for comparison analysis.
- **Intelligence:** Integrated Gemini with LangGraph agents for autonomous decision-making and tool table analysis for insights.

RAG Assisted technical document Authoring tool

- **Objective:** Automated **technical document writing** for industries like **construction, legal, finance, and healthcare**, reducing manual effort and turnaround time from **weeks to hours**.
- **Approach:** Integrated **LLMs (Claude) with RAG (Cohere, FAISS)** for **accurate retrieval & structured document generation**.
- **Optimization:** Used **multi-threading & refined prompt engineering** to improve response accuracy and efficiency.

AI-Powered Document Quality Enhancement System

- **Objective:** Extended the document authoring tool with advanced post-processing capabilities to ensure enterprise-grade document quality and compliance standards.
- **Approach:** Built ML-powered validation pipeline using NLP for grammar correction, terminology checks, and compliance review. Integrated a **ReAct agent** combining **LangChain, LangGraph**, and the **Pint** library for standard unit conversion.
- **Features:** Automated **citation formatting**, de-duplication with semantic checks, and structure optimization using templates.
- **Impact:** Reduced manual review time by **95%** and improved document accuracy by **40%** with intelligent error detection.

Tata Consulting Engineers Ltd. | Assistant Manager | Aug 2019- Aug 2021

Automation of Tall Chimney Design | Automation

- **Objective and Implementation:** Built an automation platform for RCC Chimney design, transitioning from manual to data-driven workflows with Python algorithms and MySQL for calculations, validation, and reporting.
- **Impact:** Achieved a 40% reduction in design time and eliminated errors through automated data processing, rule-based validations, and performance tracking.

Business Intelligence & Data Analytics Initiative

- **Objective & Implementation:** Led data analytics exploration to transition engineering workflows into insight-driven processes, creating dashboards for performance tracking, resource utilization, and decision-making.
- **Skills & Research:** Acquired expertise in Power BI; conducted POCs for automated reporting and business intelligence solutions to optimize management strategies.

Award and Achievements

- Kudos Achievement **Award by CTO of TCE** for handling a high volume of **mathematical calculations, design charts** and delivering user friendly web based tool.
- Obtained 481 Rank in the **GATE-2017** out of 1.3 lakh+ candidates.

Education

- M. Tech. 2017-19 IIT Bombay
- B. Tech. 2012-16 MMMUT Gorakhpur

Self Project Links

- **Emotion Recognition:** Fine-tuned a BERT model to classify customer emotions from e-commerce product reviews (96% accuracy) using Kaggle data, enabling businesses to analyze customer behavior and enhance service quality.
- **Car Price Prediction:** Developed a car price prediction model using regression techniques, achieving an R^2 of 0.83 with Random Forest, utilizing one-hot encoding and min-max scaling for feature preprocessing.