

Minautee

+917857955843 minautee2001@gmail.com [in Minautee](#) [G Minautee](#) [k Minautee](#)

Professional Experience

Fractal Analytics *AI Engineer*

Bangalore, KA Oct 2023 - present

- **Optimized ETL pipelines** using Airflow to extract data from Snowflake tables with 20+ key metrics, **calculating z-scores and anomaly counts**. Integrated **real-time Slack and email alerts**, reducing anomaly detection and action time by **80%**, and **accelerating stakeholder notifications** for improved decision-making.
- **Revamped Slack alerting system** for the **Top 60 daily alerts**, implementing **real-time stakeholder tagging** and detailed threshold values across **30+ analytics**. Introduced a **T-3 day historical alert count**, improving alert management efficiency by **60%** and enhancing stakeholder accountability.
- **Architected and deployed** a complete data model for a business vertical, building **10+ fact tables** in Snowflake using **DBT**. Orchestrated workflows via **Airflow**, achieving **70% faster data delivery**. Collaborated directly with clients to **align technical solutions** with business goals, ensuring seamless integration and actionable insights.
- **Co-engineered a content-based tagging system (GAIT)** using **Generative AI (GenAI)** on **AWS Sagemaker**, developing a comprehensive taxonomy for content categorization. Improved tagging accuracy to **70%** and **reducing manual efforts** by 80%, streamlining the content management process.
- **Enhanced tag prediction accuracy** by building a **critic agent** leveraging **parallel processing** with **ThreadPoolExecutor** over **300+ data points**. Generated detailed **JSON-formatted outputs** with reasoning and keyword analysis for misclassified tags, leading to a **10% boost** in tagging efficiency by refining the existing taxonomy.
- **Integrated a client-specific GPT API endpoint**, fine-tuned with company data, improving content tagging accuracy by **30%** while ensuring **data security**. Conducted a comparative study using **Anthropic** and **Claude** through **AWS Bedrock**, evaluating LLM performance to inform future development strategies.
- **Authored weekly articles** on **Generative AI (GenAI) use cases**, improving team knowledge sharing and fostering collaboration on **GenAI applications**, leading to increased engagement and adoption across the team.
- **Developed** a dynamic PDF query system using **Ollama** to generate responses from documents uploaded via a **Streamlit interface**. **Integrated** open-source **FAISS** library for efficient vector searches, enabling seamless handling of large datasets.
- **Led a live session** on "**The Art of Prompting**" for **100+ participants**, providing hands-on insights into effective prompt engineering techniques, which boosted team proficiency in working with LLMs and GenAI tools.

Projects

Website Scraping and RAG

Jan 2025

- Scraped website data using **Crawl4AI** and stored embeddings in **Supabase** for efficient retrieval.
- Built an agentic **RAG system** with **Pydantic AI**, enhancing structured responses and contextual accuracy.
- Developed an interactive **Streamlit** frontend, enabling users to query and extract useful insights dynamically.

Food Recommendation System

Dec 2024

- Preprocessed and enriched recipe data by cleaning reviews, performing sentiment analysis, and tagging nutritional values for better recommendations using **NLP**.
- Implemented a hybrid recommendation system using **content-based filtering** (TF-IDF & cosine similarity) and **collaborative filtering** (SVD) to personalize recipe suggestions.
- Built an interactive **Streamlit** app, allowing users to receive real-time food recommendations based on inputs like meal type, preference, dietary needs, etc.

WeCure Management System using Azure

Dec 2023

- Designed an **ETL pipeline** on **Azure Data Factory** to load source data to the destination using **Medallion architecture**, reducing the processing time of ad-hoc queries by **30%**.
- Enhanced data quality using **Medallion architecture**, processing datasets exceeding **5GB** in size with **PySpark** on **Databricks**. Applied performance tuning techniques like caching and partitioning to improve job efficiency.

Technical Skills

AI/ML & Deep Learning: Generative AI (GenAI), Transformers, GPT, LangChain, TensorFlow, PyTorch, scikit-learn, NLP, Model Fine-Tuning

Programming & Data Engineering: Python, SQL, API Integration, ETL, DBT, Airflow, PySpark, Hadoop, Docker

Cloud & Big Data: AWS, Azure, Snowflake, Databricks

Data Science & Visualization: Exploratory Data Analysis (EDA), Power BI, Statistical Modeling

Education

MIT World Peace University

Pune, MH 2019-2023

B.Tech Electrical Engineering GPA: 9.5/10

Online Courses & Certifications

- AWS Certified Solutions Architect (Feb. 2024 - Feb. 2026) - [AWS](#)
- Google Data Analytics Professional Certificate (May. 2023) - [Coursera](#)
- Databricks: Interactive Learning with AI for Medallion, Delta Live Tables & Unity Catalog (Jul. 2024) - [Fractal Hive](#)