Minautee

+917857955843 ✉ minautee2001@gmail.com 🔗 Minautee 🐙 Minautee k Minautee

Profile

Aspiring data engineer skilled in Python, SQL, and data analysis, with a focus on optimizing language models and fine-tuning techniques. Passionate about leveraging cutting-edge AI methods to create context-aware outputs and drive impactful real-world solutions. Eager to contribute and grow in dynamic, fast-paced environments.

Education

MIT World Peace University   **Pune, MH** *2019-2023*
B.Tech Electrical GPA: 9.5/10

Technical Skills

Programming & Data Engineering:

Python | SQL | LaTeX | PySpark | Docker | Hadoop | Databricks | ETL | GIT | PostgreSQL | API Integration | Snowflake | DBT | Airflow

Cloud & AI Technologies:

Azure | AWS | Generative AI (GenAI) | Prompt Engineering | Langchain | TensorFlow | PyTorch

Data Analysis & Visualization:

Power BI | Excel | Exploratory Data Analysis (EDA) | Model Fine-Tuning | Natural Language Processing (NLP)

Professional Experience

Imagineer - Fractal Analytics   **Bangalore, KA** *10/2023 - present*

- Developed and implemented an end-to-end data pipeline for Realtor.com, leveraging dbt to create tables in Snowflake and configuring tasks in Airflow to orchestrate workflows within a DAG. Optimized the pipeline to ensure smooth data flow from source systems to the dashboard, reducing processing time and enhancing data accuracy for real-time decision-making.
- Revamped alerting systems for data anomalies across multiple verticals by integrating time-delta-based metrics, extending monitoring periods, and implementing owner tagging for accountability via Email and Slack. These improvements enabled stakeholders to proactively address anomalies, ensuring accurate dashboards, enhancing data reliability, and fostering trust through timely and actionable insights.
- Developed and implemented a cutting-edge content-based tagging system (GAIT) using Generative AI (GenAI) for Merck Sharpe & Dohme across U.S. markets, with plans to expand globally. Integrated client-specific GPT API endpoint (GPTeal) to ensure secure and compliant tagging of diverse content types. Utilized a comprehensive taxonomy that reduced manual tagging efforts by 73% while significantly improving content categorization accuracy.
- Improved tag prediction accuracy through prompt optimization, generating JSON-formatted outputs with reasoning and keyword analysis for misclassified tags. Additionally, used a critic agent to fine-tune taxonomy definitions, resulting in a 10% increase in future tag predictability, achieving 80% accuracy compared to earlier results.
- Contributed to the CDTGenAI team by developing GenAI applications in the cloud domain and sharing knowledge through newsletters. Authored weekly articles on GenAI use cases and created a RAG-based Q&A chatbot for the team. Used Ollama to generate responses from PDFs uploaded via the Streamlit interface, preprocessing text into chunks and storing them in FAISS for similarity searches.

Projects

Cycling Race using AWS   *02/2024*

- Developed a real-time analytics dashboard on AWS for cyclists, providing better competition insights using Power BI.
- Deployed Amazon Kinesis and RDS to process GPS and registration data, resolving anomalies and loading into DynamoDB.
- Integrated DynamoDB with Power BI via a Cdata connector for real-time visualizations.

WeCure Management System using Azure   *12/2023*

- Implemented a data consolidation system on Azure, reducing ad-hoc queries processing time by 30%.
- Utilized Databricks to enhance data quality using medallion architecture by processing and analyzing datasets exceeding 5GB in size using Spark on Databricks, implementing performance tuning techniques like caching and partitioning to improve job efficiency.
- Leveraged Azure Synapse Analytics for ad-hoc queries, boosting efficiency by 40% while ensuring 100% data integrity and compliance.

Online Courses & Certifications

- AWS Certified Solutions Architect (Feb. 2024 - Feb. 2026) - AWS
- Microsoft Certified: Azure Data Engineer Associate (Dec. 2023 - Dec. 2024) - Microsoft Azure
- Google Data Analytics Professional Certificate (May. 2023) - Coursera
- Consulting Bootcamp (Aug. 2022) - Strategy & Co.
- Databricks: Interactive Learning with AI for Medallion, Delta Live Tables & Unity Catalog (Jul. 2024) - Fractal Hive
- Financial Markets, (Jul. 2020) - Coursera