Summary

In building the Sales Performance Analysis System, I took a thoughtful approach by combining the LLaMA model and Hugging Face Transformers. This choice was driven by my need for strong natural language processing capabilities while keeping an eye on limited computational resources.

Key Approaches

LLaMA Model Integration: I chose the LLaMA model (LLaMA 3.1) because it delivers excellent performance with lower resource demands compared to larger models like GPT. This made it a perfect fit for my setup.

Hugging Face Transformers: I used Hugging Face Transformers for their powerful text embedding features, which helped in efficiently analyzing the sales data and retrieving relevant feedback.

A web app built with Streamlit, a command-line script for direct queries, and a RESTful API using Flask. This variety allows users to choose how they want to interact with the system.

Data Preparation: I focused on preprocessing the sales data and creating embeddings to ensure that the models could respond accurately to user queries.

Why LLaMA Over GPT

I go for LLaMA instead of GPT for several reasons:

Efficiency: LLaMA requires fewer resources, making it more suitable for my limited hardware capabilities.

Context Understanding: It performs well with context-rich queries, which is essential for analyzing detailed sales performance.

Community Support: LLaMA has a growing community, providing plenty of resources and support, which is helpful for troubleshooting and learning.

Conclusion

Overall, my approach was about finding the right balance between powerful analysis tools and resource management. By using LLaMA and Hugging Face Transformers, I aimed to create a user-friendly system that effectively meets the needs of sales professionals, all while keeping things efficient and manageable.

Github Repo: https://github.com/Wasiue03/Performance-System