Loan Support Chatbot for Smart Bank: Project Report

Introduction:

This report details the approach taken to develop a loan support chatbot for Smart Bank, utilizing a Large Language Model (LLM). The chatbot is designed to assist customers with loan eligibility checks, product information, application guidance, FAQs, and personalized recommendations.

Meeting Project Requirements:

Strategies for Training Data:

- A traditional RAG-based approach was considered but deemed computationally expensive and potentially prone to losing context due to data splitting.
- Instead, the entire conversation context was fed into the system message section of the prompt itself, leveraging the large context window (8000 tokens) of the LLM model compared to the typical conversation size (around 3000 tokens).
- Use Few-shot prompting by providing some examples on how the chatbot should behave to improve model performance and reliability.

Maintaining Chat History:

- StreamlitChatMessageHistory and RunnableWithMessageHistory APIs provided by Langchain were utilized to manage chat history on the client-side, eliminating the need for manual implementation.

Cost-Effectiveness Evaluation:

- Avoiding RAG-based processing reduced computational costs significantly.
- Streamlit facilitated a single UI, minimizing the need for separate frontend and backend maintenance.
- API-based LLM inference ensured minimal computationally intensive processes within the application.

Additional Features:

- Instructions were embedded within the system message to ensure domain-relevant responses from the LLM.
- Deployed the application here