**CodeMate AI Summary Report : DataQuest**

**Approach Taken:**

* Developed the core functionality using Langchain, Pinecone and OpenAI text embedding model.
* **Primary Feature** : Implements a Q/A bot with interactive and mobile responsive frontEnd that accepts PDF files.
* **Bonus Feature** : User is able to modify the complexity of the generated answers (Easy,medium,Hard).
* Uses Nextjs and mantine v7 library to implement mobile responsive front-end

**Technology Used :**

* OpenAI's text embedding model (text-embedding-ada-002)
* Langchain framework for QA chains
* Pinecone vector database
* PDFParse for parsing PDF files
* Multer for directory and file manipulation
* NextJs for frontEnd
* Mantine v7 library

**Challenges Faced:**

* Understanding the workings of Langchain and Pinecone within a short amount of time was a bit difficult, but it was informative.
* Currently, Free tier of pinecone database doesn’t support deleting vectors via namespaces or metadata filters, had to implement other ways to do the same.
* Took some time to debug the hallucinations faced by Langchain vector stores
* Using multer and other file system manipulation tools to handle different PDF formats took some time to implement.

**Future Improvements**

* Can increase the accuracy of Generated answer (Less to No hallucinations).
* Plans to implement more question configurations like specific topics to focus, number of concurrent responses,ect..
* Can make the frontEnd more seamless.