Summary Report

Group – 2

University of Missouri – Kansas City

COMP\_SCI5588: Data Science Capstone

Yugyung Lee

March 18, 2024

# Summary Report

**Phase – 2:**

**1. User Authentication and Registration System:**

* Introduction: Aims to implement a user authentication and registration system using Flask and Firebase, enabling secure login, registration, and logout functionalities.
* Technologies Used: Flask, Firebase, Pyrebase.
* Implementation:
  + Utilizes Firebase for authentication and real-time database. Flask handles application routes and session management.
  + User authentication uses Firebase's methods, while registration includes storing additional user details in Firebase.

**2. Integrating Firebase with Streamlit for Chat Application and Data Export:**

* Introduction: Outlines the technical implementation of a chat application using Streamlit and Firebase, including a feature for exporting chat data to CSV format.
* Technologies Used: Streamlit, Firebase, Firebase Admin SDK, Firestore.
* Implementation:
  + Utilizes Streamlit for the chat application interface, with Firestore backend for chat history storage.
  + Implements a separate feature for exporting chat data to CSV format using Firebase Admin SDK.

**3. User Authentication and Chat Application Development:**

* Introduction: Aims to develop a user authentication system and a chat application using Flask, Firebase, and Streamlit.
* Technologies Used: Flask, Firebase Authentication, Firestore, Streamlit.
* Implementation:
  + Flask handles user registration, login, and session management, while Streamlit provides the frontend for the chat application.
  + Firebase Authentication ensures secure user authentication, and Firestore stores chat messages securely.

**4. Natural Language Processing**

* Introduction: To finetune an LLM to be able to answer user queries about universities.
* Technologies Used: Llama2, Langchain, Faiss, Firebase
* Implementation:
  + Llama2 functions as the core NLP engine, enabling the bot to comprehend user queries and generate human-like responses effectively.
  + LangChain enhances Llama2's capabilities by integrating external data sources and APIs, allowing the bot to access diverse linguistic resources for accurate information delivery.
  + Faiss acts as the vector database, employing efficient retrieval algorithms to swiftly fetch relevant information from the scraped data, thereby improving the bot's performance and user experience.

**Phase 3 -2:**

**1. Enhanced User Authentication and Registration System:**

* Introduction: Building upon the existing authentication and registration system, the focus was on refining the user experience and data management capabilities.
* Technologies Used: Flask, Firebase, Pyrebase.
* Implementation:
* The login/register page was redesigned for a more professional and user-friendly interface.
* Additional details such as name, university name, and email were incorporated for comprehensive user profiles.
* User details are securely stored in a real-time Firebase database, ensuring efficient data management.
* Upon login, users are redirected to the UniBuddy homepage, where personalized features are accessible.

**2. Dynamic Database Handling and Integration with Mistral AI:**

* Introduction: This phase focused on enhancing database handling and integrating the Mistral AI model for advanced natural language processing.
* Technologies Used: Firebase, Mistral AI, Langchain.
* Implementation:
* Each user's chat history is now automatically saved with their email ID as the database name, streamlining data organization.
* The Mistral AI model was successfully integrated into the UniBuddy homepage, enhancing the chatbot's capabilities.
* User prompts and model responses are captured and stored in respective databases, facilitating ongoing model training and refinement.