Applied Data Science Department

Hospital Readmission Prediction System With LLM

Project Advisor: Dr. Simon Shim

Jaya Lakshmi Gunji
Naveen Yadav Gongati
Neha Dabeeru
Priyanka Akula
Somna Sattoor

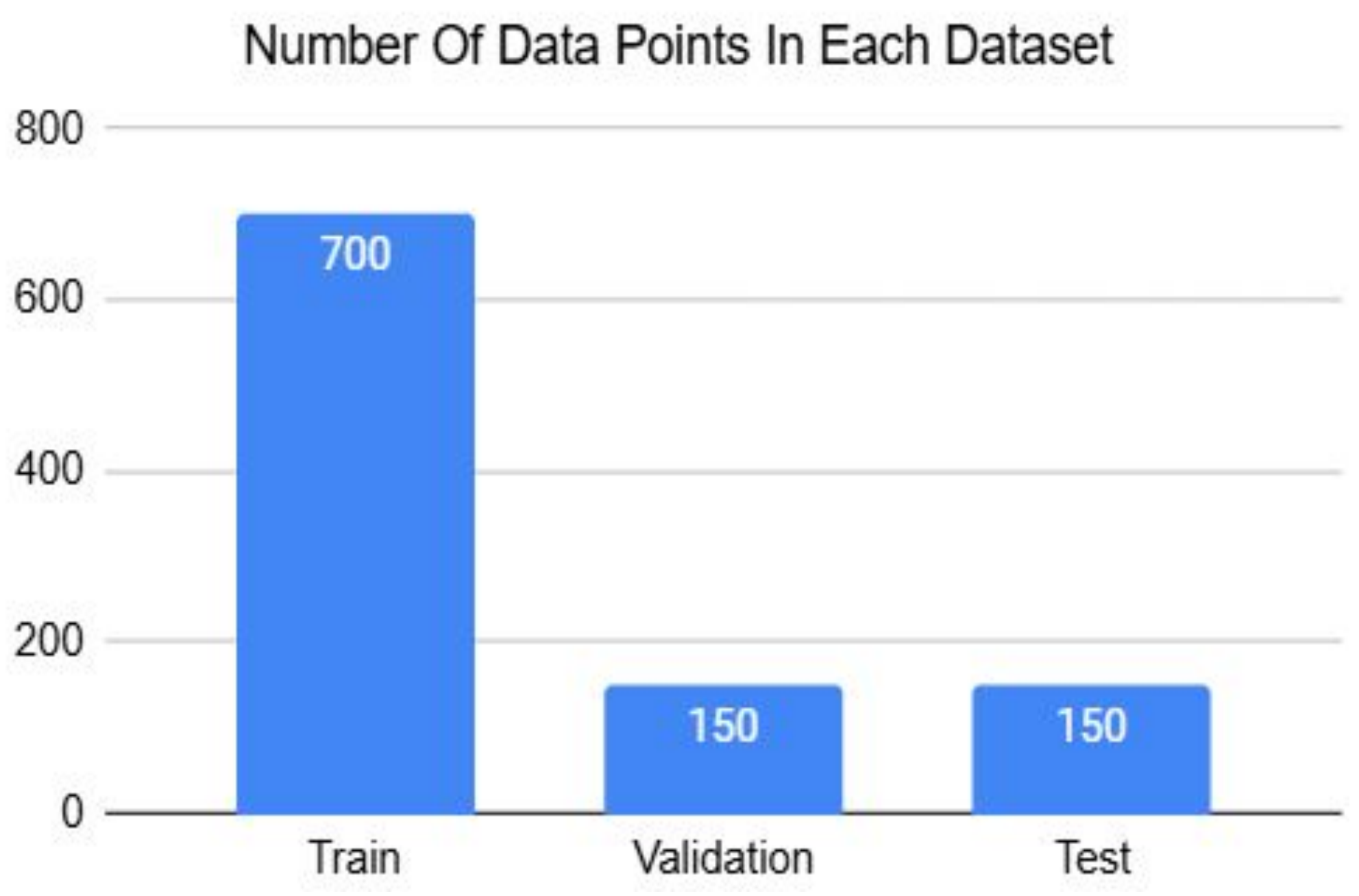
Introduction

Hospital readmissions are a growing concern, straining healthcare systems and negatively impacting patients' recovery journeys. Every unexpected readmission means additional stress for patients, families, and medical teams, along with increased healthcare costs. To address this, **MedPredict - Hospital Readmission Prediction System** leverages cutting-edge technology to predict the likelihood of a patient being readmitted within 30 days. Powered by advanced machine learning models and tools like Python, FastAPI, and React, MedPredict transforms complex patient data from the MIMIC-III dataset into clear, actionable insights. By helping healthcare providers focus on high-risk cases and take proactive measures, MedPredict not only reduces readmission rates but also improves the quality of care and resource management, paving the way for a more patient-centered and efficient healthcare system.

Data Collection and Preprocessing

Hospital Readmission Prediction System is built on a foundation of robust and reliable data. It utilizes the MIMIC -III clinical database, a comprehensive source of patient records that includes demographic details, medical diagnoses, treatment histories, medications, and clinical notes. This rich dataset provides a complete view of patient care, enabling accurate and meaningful predictions of hospital readmissions.

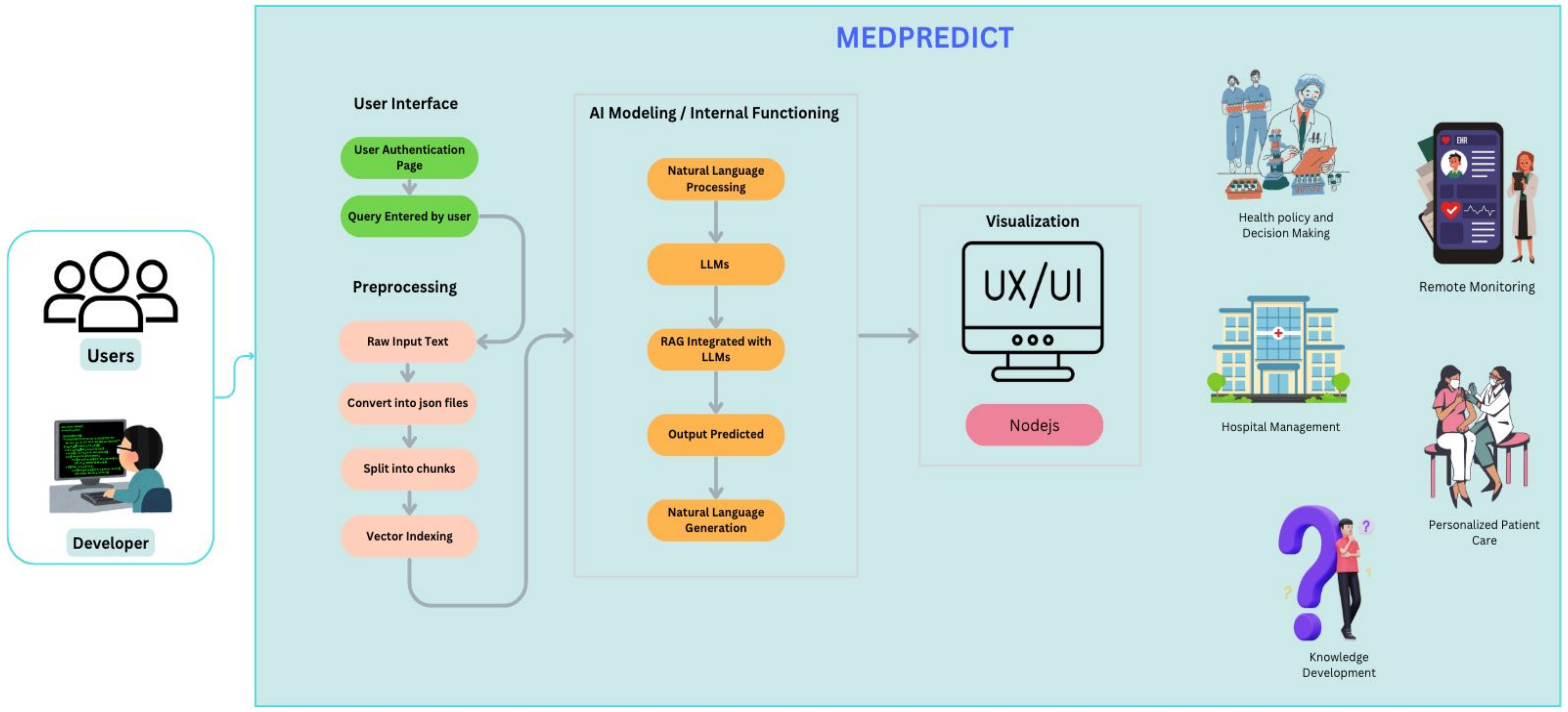
The collected data was processed into a structured JSON file, enhanced with summaries and additional context. To ensure balance, synthetic data was generated, and the dataset was reformatted into a question-and-answer format for NLP tasks. It was then split into training, validation, and test sets. Different LLM models were trained using the training set, evaluated based on performance, and the best model was selected for chatbot development.



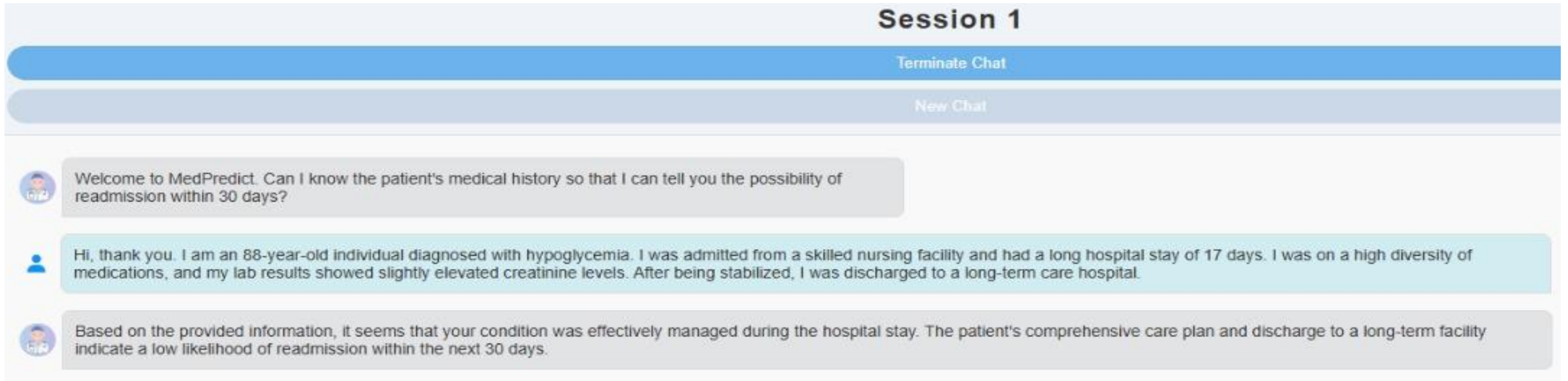
Methodology

The MedPredict chatbot was built by carefully selecting the most effective language model for processing medical text. Several advanced models, including **Clinical BERT**, **Clinical XLNet**, **PubMedBERT**, and **Clinical BigBird**, were evaluated using key performance metrics. After thorough testing, **PubMedBERT** stood out for its exceptional accuracy and reliability, making it the ideal choice. This model was fine-tuned and seamlessly integrated into the chatbot, ensuring it delivers precise and meaningful healthcare insights tailored to user queries.

The system is designed to seamlessly integrate advanced AI technologies with user-friendly interfaces to deliver personalized healthcare insights. On the backend, **FastAPI** is used to manage efficient communication, while deep learning models built with **PyTorch** and fine-tuned using **Hugging Face Transformers** (PubMedBERT) handle medical text analysis and classification. Raw input data is preprocessed by splitting it into smaller chunks, converting it into JSON files, and embedding them using **Sentence Transformers**. **FAISS** enables fast similarity searches to retrieve relevant patient summaries. The frontend, built with **React** and styled with **CSS**, ensures an intuitive and responsive user experience, supported by **Firebase** for secure authentication. **Docker** and **Docker Compose** streamline deployment and scalability, ensuring the system remains robust and efficient. Insights are visualized through a sleek **Node.js** interface, empowering users with actionable outcomes for applications like healthcare policy, hospital management, remote monitoring, and personalized patient care. Rigorous API testing with **Postman** ensures the system's reliability and effectiveness.



The MedPredict chatbot provides a conversational interface designed to deliver personalized healthcare insights. As shown in the interface, users can engage in real-time sessions to input patient medical histories and receive AI-driven predictions or recommendations. The chatbot utilizes advanced natural language processing and retrieval-augmented generation (RAG) techniques to interpret input data and deliver precise, contextually relevant responses. The intuitive design ensures seamless user interaction, offering a reliable platform for understanding patient conditions and accessing follow-up care suggestions.



Welcome to MedPredict. Can I know the patient's medical history so that I can tell you the possibility of readmission within 30 days?

Hi, thank you. I am an 88-year-old individual diagnosed with hypoglycemia. I was admitted from a skilled nursing facility and had a long hospital stay of 17 days. I was on a high diversity of medications, and my lab results showed slightly elevated creatinine levels. After being stabilized, I was discharged to a long-term care hospital.

Based on the provided information, it seems that your condition was effectively managed during the hospital stay. The patient's comprehensive care plan and discharge to a long-term facility indicate a low likelihood of readmission within the next 30 days.

Analysis and Results

Various evaluation metrics, including **GLEU score**, **accuracy**, **F1 score**, and **recall**, are used to assess the performance of all model query engines. These metrics not only measure contextual relevance but also capture the semantic similarity between the correct response and the one retrieved by the query engine. The performance of **MedPredict** has been rigorously tested across different use cases, and it was found that the **PubMedBERT query engine** consistently retrieves responses that are both semantically and contextually richer compared to other models.

Summary/Conclusions

MedPredict showcases the potential of **LLMs** and **RAG** in overcoming challenges in data retrieval, improving the accuracy of readmission predictions, and enhancing contextual understanding within healthcare decision-making. It lays the groundwork for further advancements in predictive models across various healthcare areas. This research emphasizes how leveraging accurate and timely insights from **MedPredict** can significantly enhance decision-making in healthcare settings, leading to better patient care, more efficient resource allocation, and overall improved outcomes.

Key References

Ando, T., & Ando, K. (2023). Factuality analysis of SNS posts containing diverse symptom expressions for public health surveillance. *2023 9th International Conference on Systems and Informatics (ICSAI)*, Changsha, China, 1–5. https://doi.org/10.1109/ICSAI61474.2023.10423347

B. Saha, S. Lisboa, and S. Ghosh, "Understanding patient complaint characteristics using contextual clinical BERT embeddings," *arXiv (Cornell University)*, Jul. 2020, doi: https://doi.org/10.1109/embc44109.2020.9175577.

G. Althari and M. Alsulmi, "Exploring Transformer-Based Learning for Negation Detection in Biomedical Texts," *IEEE Access*, vol. 10, pp. 83813–83825, 2022, doi: https://doi.org/10.1109/access.2022.3197772.

Guo, Y., & Wan, Z. (2024). Performance evaluation of multimodal large language models (LLAVA and GPT-4-based ChaTGPT) in medical image classification tasks. *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*, 541–543. https://doi.org/10.1109/ichi61247.2024.00080

Acknowledgements

We would like to express our sincere gratitude to our Project Advisor, Dr. Simon Shim, for his unwavering support and invaluable guidance throughout the project.

We also extend our thanks to Dr. Lee Chang and the Department of Applied Data Science for providing the necessary infrastructure and structural support that helped us successfully complete the project.