A Weekly Update on

DiaCare: An Intelligent Diabetes Management Application with LLM-Augmented Chatbot and ML-Based Early Risk Prediction

19 February, 2025

Group Information

Group-01 CSE299 (Section-17)

1. Saif Mohammed - 2121913042
2. Nazibul Islam Nabil - 2222456642
3. Humayra Rahman Nipa - 2121128042
4. Umme Suraia Haque Setu - 2031278642

Weekly Update Brief

- ▶ **Machine Learning:** Completed EDA Preprocessing, including data cleaning, handling categorical data, feature engineering, and train-test split.
- ▶ **LLM Chatbot (Part-1):** Processed BADAS Guideline 2019 PDF, implemented document chunking, embeddings, and retrieval for basic Q&A.
- ▶ **LLM Chatbot (Part-2):** Explored token counting, cosine similarity, retrieval, and RAG chain implementation.
- ▶ **Front-End Update (Flutter):** Integrated chatbot and DiaTrack input fields for improved user interaction.
- ▶ **Overall LLM:** Conducted experiments and evaluations on text embeddings and retrieval strategies for enhanced chatbot performance.
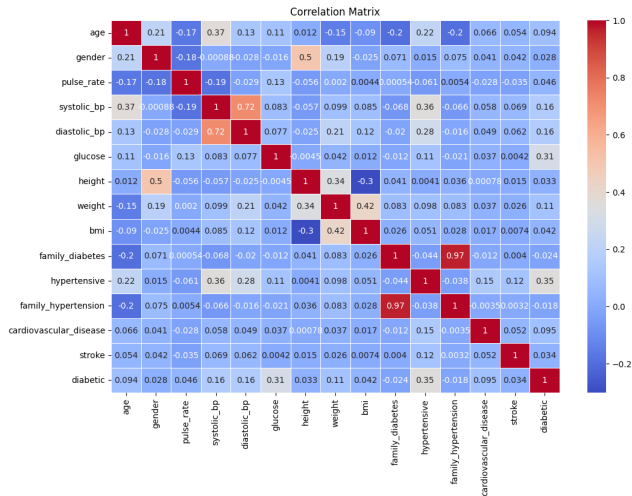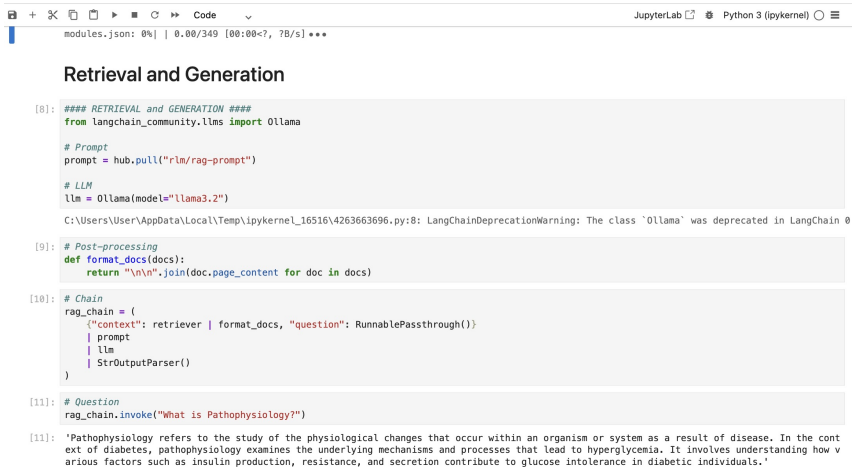
Machine Learning (Preprocessing)

Figure: Feature Engineering (Correlation Matrix)

LLM RAG Chatbot using LangChain (Overview)

```
modules.json: 0%| | 0.00/349 [00:00<?, ?B/s] •••
```

Retrieval and Generation

```python
[8]: #### RETRIEVAL and GENERATION ####
     from langchain_community.llms import Ollama

     # Prompt
     prompt = hub.pull("rlm/rag-prompt")

     # LLM
     llm = Ollama(model="llama3.2")

     C:\Users\User\AppData\Local\Temp\ipykernel_16516\4263663696.py:8: LangChainDeprecationWarning: The class `Ollama` was deprecated in LangChain 0
```

```python
[9]: # Post-processing
     def format_docs(docs):
         return "\n\n".join(doc.page_content for doc in docs)
```

```python
[10]: # Chain
      rag_chain = (
          {"context": retriever | format_docs, "question": RunnablePassthrough()}
          | prompt
          | llm
          | StrOutputParser()
      )
```

```python
[11]: # Question
      rag_chain.invoke("What is Pathophysiology?")
```

```
[11]: 'Pathophysiology refers to the study of the physiological changes that occur within an organism or system as a result of disease. In the cont
      ext of diabetes, pathophysiology examines the underlying mechanisms and processes that lead to hyperglycemia. It involves understanding how v
      arious factors such as insulin production, resistance, and secretion contribute to glucose intolerance in diabetic individuals.'
```

Figure: Basic Retrival & Generation from textbook

LLM RAG Chatbot using LangChain (Indexing)

```
[57]: # Run
      chain.invoke({"context":docs,"question":"What is Pathophysiology?"})
```

```
[57]: 'The text refers to "Pathophysiology" as the main topic for section 1.2 on two separate occasions, describing it as:\n\n\n- The "hall mark" of
      Type 1 diabetes.\n- A description of the main pathophysiologic defects in type 2 diabetes.\n\nIn both cases, Pathophysiology is described as
      the study of the normal functions and abnormal function of living organisms and their parts, particularly in relation to disease.'
```

```
[58]: from langchain import hub
      prompt_hub_rag = hub.pull("rlm/rag-prompt")
```

```
[59]: prompt_hub_rag
```

```
[59]: ChatPromptTemplate(input_variables=['context', 'question'], input_types={}, partial_variables={}, metadata={'lc_hub_owner': 'rlm', 'lc_hub_re
      po': 'rag-prompt', 'lc_hub_commit_hash': '50442af133e61576e74536c6556cefe1fac147cad032f4377b60c436e6cdcb6e'}, messages=[HumanMessagePromptTem
      plate(prompt=PromptTemplate(input_variables=['context', 'question'], input_types={}, partial_variables={}, template="You are an assistant for
      question-answering tasks. Use the following pieces of retrieved context to answer the question. If you don't know the answer, just say that y
      ou don't know. Use three sentences maximum and keep the answer concise.\nQuestion: {question} \nContext: {context} \nAnswer:"), additional_kw
      args={})])
```

RAG chain

```
[60]: from langchain_core.output_parsers import StrOutputParser
      from langchain_core.runnables import RunnablePassthrough

      rag_chain = (
          {"context": retriever, "question": RunnablePassthrough()}
          | prompt
          | llm
          | StrOutputParser()
      )

      rag_chain.invoke("What is Pathophysiology?")
```

```
[60]: 'According to the document, Pathophysiology refers to:\n\n"Pathophysiology: Type 1 diabetes: Marked impairment of insulin production due to c
      ellular-mediated autoimmune destruction of beta cells is the hall mark. Some of type 1 diabetes cases are of idiopathic in nature. Type 2 dia
      betes: Insulin resistance and б-cell failure represent the main pathophysiologic defects in type 2 diabetes."\n\nAdditionally, it mentions th
      at these eight pathways comprise "the ominous octet" for the development of glucose intolerance in type 2 diabetic individuals.'
```

Figure: Retrieval & Generation using Cosine Similarity
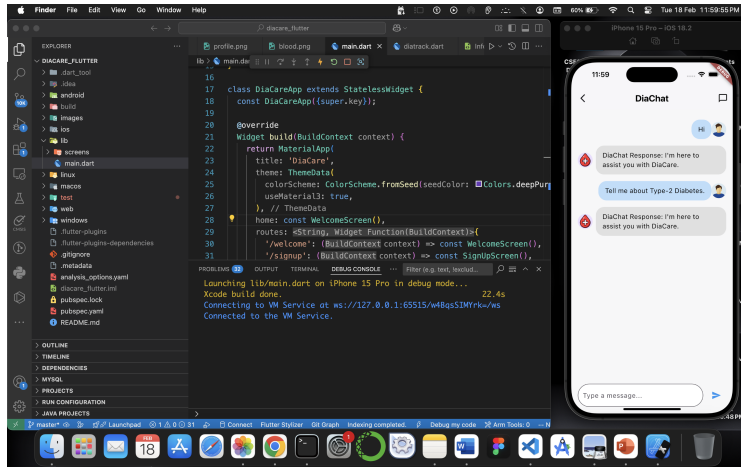
FrontEnd using Flutter (Dart)

Figure: Chatbot Interface using statefulWidget Class

Achievements

- Completed preprocessing for machine learning model development.
- Successfully implemented basic retrieval and generation from the textbook.
- Integrated OllamaEmbeddings for improved text representation.
- Built an initial RAG pipeline using LangChain and CromaDB.
- Debugged and fixed issues related to Chatbot ForntEnd.

Technology Stack

- **Programming Language:** Python
- **Framework:** LangChain
- **Libraries:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn
- **Embedding Model:** OllamaEmbeddings
- **Vector Database:** chromadb
- **LLM:** HuggingFace LLM
- **Frontend:** Flutter (Dart)
- **Document Processing:** PyMuPDF, PyPDF

Work Distribution (This Week)

- **Saif Mohammed - 2121913042**
  - ML EDA, Preproessing
  - LLM Chatbot LangChain Pipeline (Indexing)

- **Nazibul Islam Nabil - 2222456642**
  - LLM Chatbot Integration
  - LLM Chatbot Prototyping

- **Humayra Rahman Nipa - 2121128042**
  - FrontEnd Development
  - Chatbot Integration

- **Umme Suraia Haque Setu - 2031278642**
  - FrontEnd Development
  - DiaTrack Input Fields (Health Data)

References

A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2023.

LangChain Engineer, "LLM Part: Learn RAG From Scratch – Python AI Tutorial," YouTube. [Online]. Available: *https://www.youtube.com/*. [Accessed: Feb. 19, 2025].

P. Lewis, E. Denoyer, and S. Riedel, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *arXiv preprint arXiv:2005.11401*, May 2020. [Online]. Available: *https://arxiv.org/abs/2005.11401*.

GPT-4 Tutorial, "RAG Prototyping: GPT-4 Tutorial: How to Chat With Multiple PDF Files (1000 pages of Tesla's 10-K Annual Reports)," YouTube. [Online]. Available: *https://www.youtube.com/*. [Accessed: Feb. 19, 2025].