**Chatbot(Domain Specific: medical and finance)**

**Abstract**

This study presents the development of a transformer-based chatbot system designed to address user queries in two critical domains: healthcare and finance. Leveraging the powerful contextual understanding of transformer architectures, the chatbot effectively interprets and responds to user

In the **medical domain,** the chatbot is capable of diagnosing diseases based on symptoms provided by the user and delivers informative responses related to disease conditions, precautions, and treatments. In the **finance domain,** the chatbot answers a range of queries related to personal finance, banking, and investment topics, offering reliable and context-aware guidance generated through transformer-based deep learning.

This dual-domain chatbot demonstrates the versatility and effectiveness of transformers in handling distinct areas of expertise

**Introduction**

With the rapid advancement in Natural Language Processing (NLP), transformer-based models have become the cornerstone for building intelligent conversational systems. Their superior ability to understand contextual relationships in language has paved the way for more human-like, responsive, and versatile chatbots. This study introduces the Multifunctional Fine-Tuned Retrieval-Based Chatbot, a robust conversational agent developed using RoBERTa and BART transformer models, aimed at delivering domain-specific, context-aware responses in areas such as healthcare, finance, and general conversation.

Unlike generative chatbots, this system follows a retrieval-based architecture, where user queries are matched against a curated dataset using cosine similarity. The chatbot retrieves the most semantically relevant responses by generating high-quality sentence embeddings through RoBERTa-based sentence-transformers. This approach ensures fast, accurate, and reliable response generation across multiple functional domains.

To enhance domain specificity and robustness, the chatbot is fine-tuned on specialized datasets and further strengthened through data augmentation using NLP-Aug, allowing it to handle diverse linguistic variations and phrasings. Additionally, TextBlob is integrated to perform sentiment analysis, categorizing user queries as positive, negative, or neutral, and enabling the chatbot to tailor its replies accordingly.

**Model Analysis**

A wide comparative analysis was conducted on various sentence transformer models—including all-mpnet-base-v2, multi-qa-mpnet-base-dot-v1, all-distilroberta-v1, all-MiniLM-L12-v2, multi-qa-distilbert-cos-v1, paraphrase-albert-small-v2, and others—to identify the best-performing embeddings for retrieval accuracy and inference efficiency. The insights from this evaluation guided the model selection for optimal system performance.

This multifunctional chatbot demonstrates the effectiveness of combining **transformer-based language models, retrieval mechanisms,** and **sentiment analysis tools** to build a scalable and adaptable solution suitable for a wide range of real-world applications—from **customer support** to**research assistance**and**automated domain-specific information retrieval.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model Name** | **Language Support** | **Embedding Size** | **Best\_Score** | **Response Time** | **Notes** | |
| paraphrase-multilingual-mpnet-base-v2 | 50+ languages | 768 | 0.8703 | Fast | |  | | --- | | Produces high-quality multilingual embeddings; performs well in cross-lingual semantic matching tasks. | | |
| paraphrase-albert-small-v2 | English only | 768 | 0.8633 | Moderate | Lightweight and fast model; suitable for limited-resource environments but offers lower accuracy. | |
| paraphrase-multilingual-MiniLM-L12-v2 | 50+ languages | 384 | 0.9801 | Fast | Delivers an excellent balance of speed and accuracy; ideal for multilingual, high-performance tasks. | |
| paraphrase-MiniLM-L3-v2 | English only | 384 | 0.9595 | Very Fast | Extremely lightweight model; offers very fast inference with slightly reduced accuracy. | |
| distiluse-base-multilingual-cased-v1 | 50+ languages | 512 | 0.8338 | Moderate | Good cross-lingual performance with moderate speed; suitable for multilingual use cases | |
| distiluse-base-multilingual-cased-v2 | 15+ languages | 512 | 0.79203 | Moderate | Faster inference than v1 with comparable multilingual capabilities; slightly reduced accuracy. | |
| all-mpnet-base-v2 | English only | 768 | 0.72351 | Moderate | Provides high semantic accuracy for domain-specific queries; less suited for latency-sensitive tasks. | |
|  | | | | | |
|  | |

Each model was analysed in terms of **accuracy, inference time, embedding quality, language support, and computational efficiency**. This analysis informed the selection of paraphrase-multilingual-MiniLM-L12-v2 for final deployment, as it provided the best trade-off between **semantic understanding and real-time performance**, essential for a responsive chatbot experience.

The use of these **pre-trained models** greatly accelerated the development process by providing powerful linguistic representations trained on massive corpora. This eliminated the need for training from scratch and enabled **domain-specific fine-tuning,** which significantly enhanced the chatbot’s ability to interpret queries in specialized fields like symptom-based diagnosis or financial consultation

**Dataset Description and Analysis**

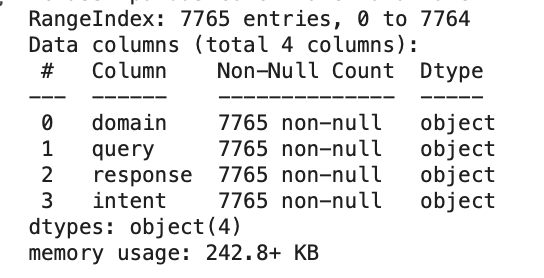
The effectiveness of a retrieval-based conversational agent is highly dependent on the diversity, structure, and semantic richness of its underlying dataset. In this study, a well-organized, multi-domain, and multi-intent dataset was employed to train and evaluate a transformer-based chatbot capable of responding to user queries across healthcare, finance, and general-purpose domains.

**A.Dataset Structure**

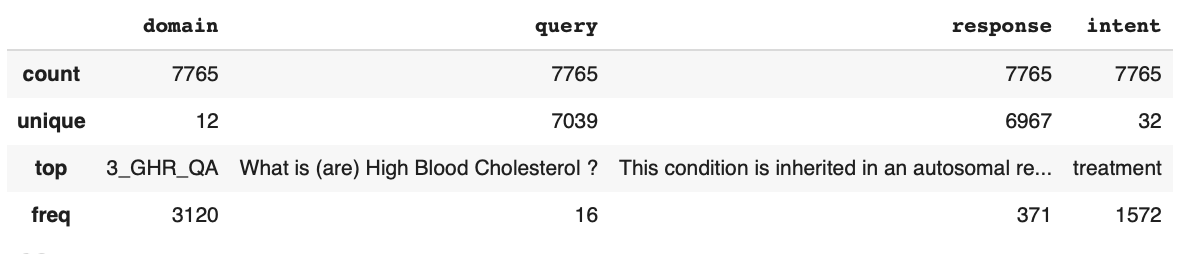
The dataset comprises **7,765 instances**, each represented across four textual fields: domain, query, response, and intent. All entries are non-null, and the dataset is free from missing or malformed records. The structure is as follows:

**B.Statistical Summary**

* **Total Records**: 7,765
* **Unique Queries**: 7,039
* **Unique Responses**: 6,967
* **Unique Intents**: 32
* **Unique Domains**: 12
* **Data Type**: All columns are of string (object) type
* **Memory Footprint**: ~243 KB

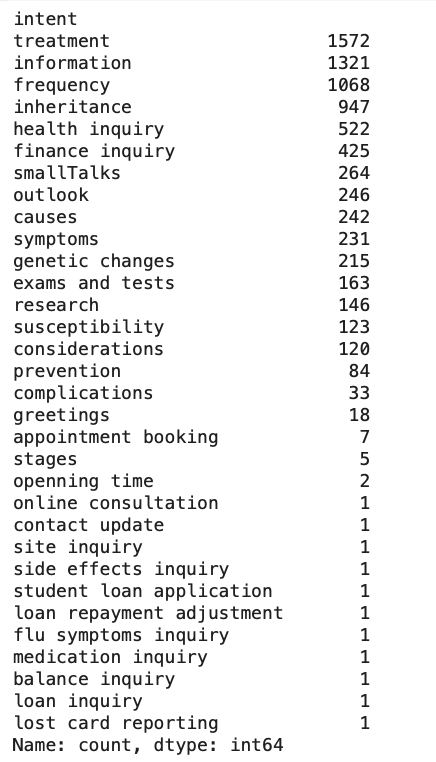


| **Attribute** | **Description** |
| --- | --- |
| domain | Denotes the context or category of the query (e.g., healthcare, finance) |
| query | Represents the user’s natural language question or input |
| response | Corresponding system-generated or pre-defined response |
| intent | Captures the functional goal or purpose behind the user's input |
|  |  |

****

**C.Intent Distribution**

The intent field plays a central role in guiding response retrieval and classification.

****

**D.Domain Analysis**

The dataset encompasses a total of **12 distinct domain categories**, each representing a specific area of user queries. Notably, the **3\_GHR\_QA** domain, which includes Genetic and Health-Related Questions and Answers, is the most prevalent—accounting for **3,120 out of 7,765 entries**. This strong representation significantly strengthens the chatbot’s domain expertise in healthcare-related conversations.