Mini-Project

Developing a Context-Aware Chatbot for Academic Support

-Spoorti Basarkod Math(11038001)

Prof. Dr. Swati Chandna
Applied Data Science and Analytics

Introduction

This document outlines my journey to develop a chatbot using NLP techniques. The project aimed to create an interactive tool that responds to user queries effectively, starting with TF-IDF for initial implementations and later attempting to integrate BERT for enhanced understanding and response quality. The document highlights the approach, challenges, and learnings from using these technologies in a practical setting.

Dataset Generation

- The dataset was generated using a custom script that creates a structured array of typical questions and answers relevant to a specific academic course.
- This method ensures a focused dataset to train the chatbot, simulating real-world interactions that the chatbot would need to handle.
- Thereby providing it with a robust foundation of data to learn from and respond to user inquiries effectively.
- Tools used- Pandas.

TF-IDF Based Chatbot Implementation

- The chatbot initially used a TF-IDF vectorization approach to process and match user queries with relevant answers from the dataset.
- By transforming text data into a vector space model, the chatbot could efficiently identify key terms in queries and retrieve the most relevant answers based on cosine similarity measures.
- This method provided a baseline for the chatbot's ability to understand and respond to user queries.
- Pre-processing: Lemmatization, Tokenization and Cleaning.
- Chatbot Functionality:
    - Greeting detection: Checks if a user's input is a greeting (like "hello", "hi", or "hey") and responds accordingly.
    - TF-IDF Vectorization: Converts the questions from the dataset into a TF-IDF matrix to compare textual similarity.
    - Cosine Similarity: Used to find the most similar stored question to a user's input to fetch the corresponding answer.
- Callback Function
    - The update_output function handles user interactions.
    - It checks if the input is a greeting and responds accordingly.
    - If not a greeting, it processes the input using TF-IDF and cosine similarity to find and return the most relevant answer from the dataset.
    - If no relevant answer is found, it informs the user that the question is not understood.

Images of the working chatbot:

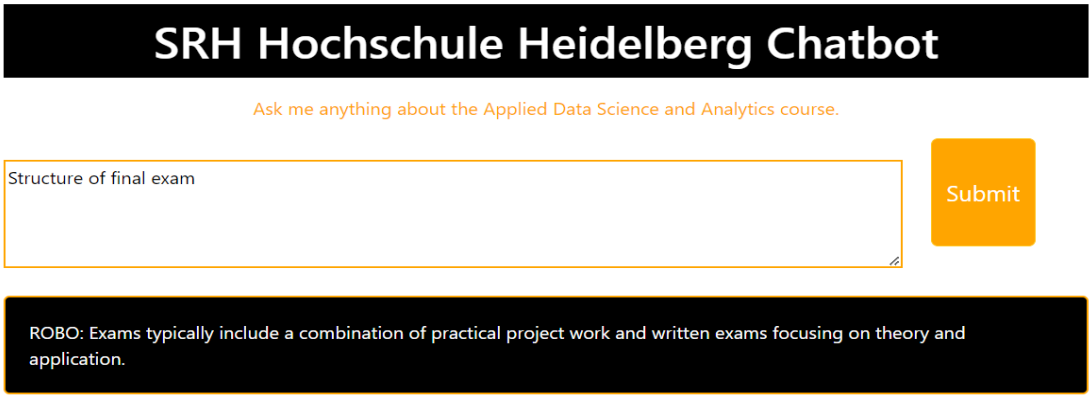

image 1: working chatbot.

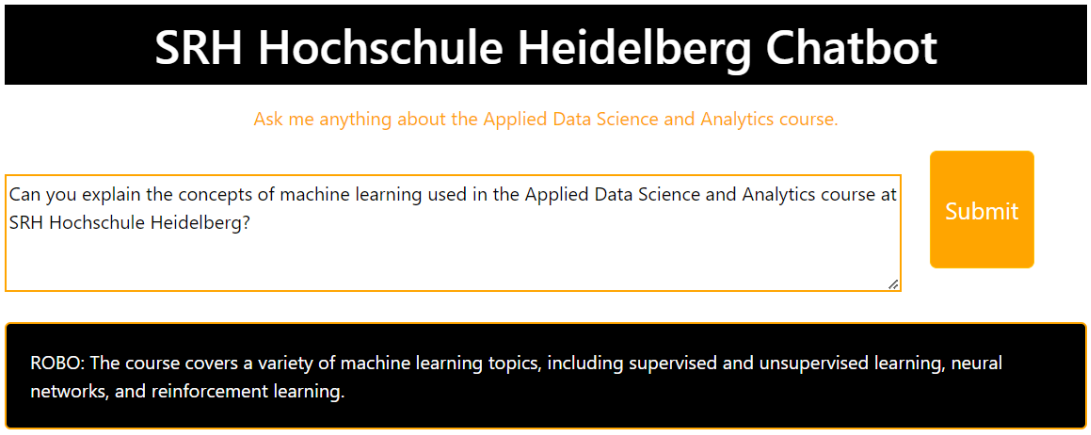

Image 2: checking a different question.

Attempt to Integrate BERT

- I attempted to integrate BERT to enhance the chatbot's performance by leveraging its advanced capabilities in understanding context within user queries.
- BERT's deep learning model processes words in relation to all the other words in a sentence, unlike traditional methods that treat words in isolation, aiming to generate more nuanced and contextually appropriate responses.
- This integration aimed to significantly improve the chatbot's ability to comprehend and process user interactions more intelligently.

Technical Challenges and Error Analysis

- The integration of BERT presented significant technical challenges, primarily due to its computational complexity and resource requirements.
- Issues included difficulties in preprocessing data to fit BERT's input requirements, managing the substantial computational resources BERT demands, and debugging incompatible library versions.
- These challenges were compounded by the high-dimensional output from BERT, which required sophisticated handling to map back to meaningful responses.
- Error messages and unexpected behavior during implementation provided critical insights into the limitations of applying such advanced NLP models in a constrained academic project setting.

Conclusion

The project remains in development, with the initial TF-IDF implementation operational and BERT integration ongoing. Future work includes addressing the technical challenges encountered, possibly through resource optimization or alternative model configurations. Consultations with experts and further research into BERT's application in chatbot technology will be crucial. This experience has provided valuable insights into the practical challenges of implementing advanced NLP techniques in real-world applications, paving the way for more effective and intelligent NLP-based interactive systems.