IMPLEMENTATION IDEA FOR SOLVING NLP ASSIGNMENT

The implementation idea I used to solve the structured data assignment are as follows:

- **Data pre-processing:** The first step is to pre-process the data. This includes cleaning the data, removing stop words, and stemming or lemmatizing the words.
- **Feature extraction:** The next step is to extract features from the data. This can be done by importing PyPDF2 library.
- **Creating a function to get MCQs:** The get_mca_questions() function creates a list of MCQs. The function takes a context as its argument. The context is the text from the PDF file that has been split into sentences.
- **Evaluation metrics:** It can be difficult to evaluate NLP models, as there is no single metric that can capture all aspects of performance.

The challenges I faced to solve the structured data assignment are as follows:

- Designing an algorithm to generate questions with multiple correct answers.
- Optimizing the question generation process for scalability.

If I have extra time, there are several things I would do to improve the performance on the task. For example, I could:

- **Try different models:** The notebook only uses a few different models for each task. So I could try other models, such as BERT or RoBERTa, to see if they improve performance.
- **Read more about NLP:** There are many resources available online and in libraries that can help me to learn more about NLP.
- **Get feedback from others:** Ask a friend, colleague, or instructor to gain some inputs about NLP and ask for feedback.

REFERENCES:

1. https://medium.com/analytics-vidhya/part-1-introduction-to-natural-language-processing-nlp-a66ad8773b3
2. https://towardsdatascience.com/practical-ai-automatically-generate-multiple-choice-questions-mcqs-from-any-content-with-bert-2140d53a9bf5