Assignment Instructions:
In this assignment, you will use the SemEval dataset to perform aspect-based sentiment analysis. Given a text and the corresponding entity in the "Aspect" column, your task is to predict the sentiment/polarity (positive, negative, or neutral) of the entity within the text. Do any preprocessing if it is required.

You must use the following three embedding methods:

1. **Frequency-based methods**: Use Bag of Words or TF-IDF (select the top 100 features (tokens) using the chi-square test to refine your feature set.
2. **Embedding-based methods**: You have to use vectors pre-trained on Wikipedia or other corpora (GloVe or Word2Vec embeddings).
3. **Sentence vectors**: Use pre-trained Sentence-BERT or Universal Sentence Encoder (USE).

Try different classifiers (Random Forest, SVM, Decision tree) for each embedding type and report the best one for comparison.

Evaluation:
Evaluate the results using the accuracy metric and cross validate each approach for 10 folds.

Deliverables:

Code: Submit a Jupyter Notebook or a Python script containing all the preprocessing, model implementation, and evaluation code.

Report: Provide a comprehensive report that includes:

Methodology: Detailed description of your preprocessing steps, choice of vectorization methods, and model implementation.

Findings: Summary of model performance, including accuracy comparisons and discussions on why certain embeddings performed better than others. You can show examples from data to support your discussion.

Submission Guidelines:

Ensure that all code is well-commented and organized.

Include all necessary libraries and dependencies required to run your code.

No cheating allowed. For plagiarism, we will be comparing your code with your peers and with the code generated by LLMs such as Chat-GPT.

Submit your report in PDF format, ensuring it is clearly structured and well-written.