ABSA Case Study

You are a data scientist who has been tasked with developing a versatile ABSA model for extracting aspects, determining aspect polarity, and detecting aspect categories from textual data. The goal is to create a robust model/pipeline ensuring flexibility and accuracy across different contexts.

Aspect-Based Sentiment Analysis (ABSA) is a natural language processing (NLP) technique that involves extracting and analysing sentiment or emotion associated with specific aspects or features of a given target entity, such as a product, service, or topic.

Data Description:

Your toolkit will include dataset comprising of 3000 customer reviews for a restaurant, all in English, enriched with human-authored annotations. These annotations contain the mentioned aspects of the target entities and the sentiment polarity of each aspect.

Tasks:

* ABSA Model Development: Create an ABSA model that can extract aspect terms, determine aspect polarity, and identify aspect categories within a given text. Ensure that the model is generic and can be applied to different domains with ease.
* Results Analysis: Analyze the results generated by the ABSA model. Use matplotlib/ plotly to show the overall as well category wise sentiment distribution.
* Actionable insights: Based on the above analysis, derive the conclusions and identify action items for the restaurant to work-upon.
* Report: Discuss pros and cons of the approach you selected and further improvement which can be done to it given more time.

In particular, the task 1 consists of the following subtasks:

Definition 1: Aspect term extraction

For the given entity - restaurant, identify the aspect terms present in the sentence and return a list containing all the distinct aspect terms. An aspect term names a particular aspect of the target entity(Restaurant).

For example, "I liked the service and the staff, but not the food”, “The food was nothing much, but I loved the staff”. Here service, staff and food are aspects.

Note: Multi-word aspect terms (e.g., “hard disk”) should be treated as single terms (e.g., in “The hard disk is very noisy” the only aspect term is “hard disk”).   
   
Definition 2: Aspect term polarity

For a given set of aspect terms within a sentence, determine whether the polarity of each aspect term is positive, negative, neutral or conflict (i.e., both positive and negative).

For example:

“I loved their fajitas” → {fajitas: positive}   
“I hated their fajitas, but their salads were great” → {fajitas: negative, salads: positive}   
“The fajitas are their first plate” → {fajitas: neutral}   
“The fajitas were great to taste, but not to see” → {fajitas: conflict}   
   
Definition 3: Aspect category detection

Decide a predefined set of aspect categories (e.g., price, food, service etc.) for restaurant, identify the aspect categories discussed in a given sentence. Aspect categories are typically coarser than the aspect terms of Subtask 1, and they do not necessarily occur as terms in the given sentence.

For example, given the set of aspect categories {food, service, price, ambience, anecdotes/miscellaneous}:

“The restaurant was too expensive”  → {aspect -expensive, category - price}   
“The restaurant was expensive, but the menu was great” → { aspect -expensive, category - price, aspect- menu, category - food}

Definition 4:

Action Items: Areas which can be improved to improve the overall sentiment of the customer.  
 

Note:  Do not use in-built ML solution pipelines from cloud such as AWS Sagemaker .

Here are your Deliverables:

- Jupyter notebook (or equivalent) showcasing your work

- Powerpoint presentation clearly explaining the approach and findings.

- Architecture diagram (if applicable) and explanations.