Input: Review like

“Camera quality is good but battery drains fast”

Output :

Camera: positive, Battery: negative

Process:

1. Load customer reviews.
2. Encode reviews.
3. Train the model using the Encoded reviews.
4. Identify specific topics in the data.
5. Classify the sentiment.
6. Use test data to see how well the model is working.
7. Give a review for prediction.
8. Map the results to the final output.

Assumptions: Data is labeled. Data involves aspect information within it. Data is cleaned, the best quality. Sufficient data will be available for training in a particular language.\*

Challenges: Model work for trained aspects/available in dataset only, for new aspects again training will be required..\*\*

Data Requirement:

1. Data Structure with Example:

| **Review** | **Aspect** | **Polarity** |
| --- | --- | --- |
| The Scene Shun Lee Palace is popular with midtown locals, possibly because the upscale room means you can impress a client and have Chinese for lunch at the same time. | Scene | Positive |
| The Scene Shun Lee Palace is popular with midtown locals, possibly because the upscale room means you can impress a client and have Chinese for lunch at the same time. | lunch | Neutral |
| To our surprise, this inquiry was interpreted as a personal offense by the waiter who told my wife to leave the store if she didn't want the pastries. | waiter | Negative |
| To our surprise, this inquiry was interpreted as a personal offense by the waiter who told my wife to leave the store if she didn't want the pastries. | pastries | Neutral |

1. Amount of Data Required

Per languages base: 45,000 \*\*\*

Per Aspect base: 150

File format: XML

Sample Data: [sample.xml](https://drive.google.com/file/d/1VDxeeYg80BSVk1JfKWuTy-WChOFXuPZH/view?usp=share_link)

\*hardware requirements for the system will be described later.

\*\*We are having some data for high-level languages from open source.

\*\*\*More languages will need more data for achieving reasonable accuracy. In the case of some languages(like Thai), more data can also be required for better training.