General Requirements

Environment

- Python 3.9+
- Jupyter Notebook

File Required

Place the following file in the same directory as the notebook you are running:
- reviews_segment.pkl – Original Amazon reviews (for Method 0 and Method 1)
- reviews_segment_cleaned.pkl – Preprocessed reviews with a clean_text column (for Method 2)

Method 0: Boolean Search (Baseline)

This method uses simple Boolean logic (AND/OR) to retrieve reviews that match aspect and opinion keywords.

Libraries

pip install pandas nltk matplotlib ipywidgets

Required NLTK Resources

```
import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
```

Stopwords Note

A built-in stopword list is used to avoid needing external files. The notebook includes:
extra_stopwords = { "i", "me", "my", ..., "now" }
stop_words = set(stopwords.words("english")).union(extra_stopwords)

How to Run

1. Make sure reviews_segment.pkl is in the same folder.
2. Open the notebook and run cells in order.

3. The following outputs will be generated:
  - reviews_segment_cleaned.pkl
  - reviews_segment_cleaned.csv
  - .txt files with matching review IDs for Tests 1–3
  - Printed precision summary table

Method 1: Classifier-Based Connotation Matching

This method uses sentence embeddings (all-MiniLM-L6-v2) and a sentiment classifier trained on review star ratings to evaluate both aspect presence and sentiment alignment.

Libraries

pip install pandas scikit-learn matplotlib seaborn tqdm sentence-transformers

How to Run

1. Ensure reviews_segment.pkl is in the working directory.
2. Run all notebook cells in order.
3. The notebook filters reviews based on:
  - Aspect match via .str.contains()
  - Sentiment match using the classifier

Output

- Folder: Task4.4_ClassifierMatch/
- Sentiment-filtered review IDs
- Summary of precision, recall, and F1 scores

Method 2: TextBlob Sentiment Matching

This method applies the TextBlob library to assess sentiment polarity and check if the review aligns with the connotation of the opinion term.

Libraries

pip install pandas nltk textblob tqdm

Required NLTK Resources

```
import nltk
nltk.download('punkt')
```

How to Run

1. Make sure reviews_segment_cleaned.pkl is in the same folder.
2. Run all notebook cells in order.

Output

- Folder: Task4_TextBlobStars/
- Retrieved review IDs per test
- Precision, recall, and F1 score summary