

Technical Report: Aspect-Based Sentiment Analysis of Headphone Reviews

Brand Battle – Final Project

1 Objective and Scope

The objective of this project is to analyze customer reviews of a headphone product in order to identify the product aspects that drive positive and negative customer sentiment. Instead of relying solely on overall ratings, the analysis focuses on extracting specific aspects mentioned in reviews and associating sentiment with them. This allows for a more detailed understanding of customer preferences and pain points.

2 Overall Workflow and Design Rationale

The analysis follows a sequential pipeline where each step prepares the data for the next stage. The workflow consists of:

1. Text preprocessing
2. Aspect extraction using RAKE
3. Adjective-based sentiment scoring
4. Negation handling
5. Aspect-level sentiment aggregation
6. Weakness identification using low-rated reviews

This order was deliberately chosen to reduce noise early in the pipeline, ensure sentiment is computed only on meaningful opinion-bearing words, and enable clear identification of weaknesses that may be hidden in the overall dataset.

3 Text Preprocessing

Text preprocessing was performed before aspect and sentiment analysis to improve extraction quality. Reviews were converted to lowercase and cleaned by removing URLs, numbers, punctuation, and extra whitespace. Tokenization was applied to standardize input for downstream processing.

Stopword removal was applied with a key modification: negation words such as *not*, *no*, and *never* were retained. Removing these words would distort sentiment polarity and lead to incorrect interpretations. Stemming and lemmatization were intentionally avoided, as modifying word forms can interfere with accurate sentiment lookup in SentiWordNet.

4 Aspect Extraction using RAKE

Aspect extraction was carried out using Rapid Automatic Keyword Extraction (RAKE), implemented entirely from scratch without importing any external RAKE libraries.

RAKE was selected because it is unsupervised, performs well on short review texts, and naturally extracts multi-word aspect phrases. The algorithm identifies candidate phrases by splitting text using stopwords as boundaries. Each word is scored using a degree-to-frequency ratio, and phrase scores are computed as the sum of their constituent word scores.

Implementing RAKE from scratch provides transparency, avoids reliance on black-box implementations, and allows customization of scoring and stopword handling. Aspect extraction was performed before sentiment analysis so that sentiment could be associated with meaningful product features rather than isolated words.

5 Sentiment Analysis using SentiWordNet

Sentiment analysis was restricted strictly to adjectives, as adjectives are the primary carriers of opinions in product reviews. Other parts of speech were ignored to reduce ambiguity and noise.

Only adjective synsets from SentiWordNet were considered. For each adjective, sentiment polarity was computed as the difference between its positive and negative scores. When multiple adjective synsets existed, the sentiment values were averaged. This approach aligns with the project constraints and ensures consistent sentiment estimation.

6 Negation Handling

Negation handling was explicitly incorporated to avoid incorrect polarity assignment. A fixed three-word window was used, where sentiment scores of adjectives appearing within the window following a negation word were inverted.

This method is simple, interpretable, and effective for short e-commerce review sentences. Negation handling was applied after adjective extraction but before sentiment aggregation to ensure polarity adjustments were reflected correctly in the final aspect sentiment scores.

7 Aspect-Level Sentiment Aggregation

For each review, multiple aspects were extracted and a sentiment score was computed based on adjectives present in the review. Aspect-level sentiment was obtained by aggregating sentiment scores across all occurrences of each aspect.

Both frequency and average sentiment were tracked. Frequency indicates how often an aspect is discussed, while average sentiment captures customer opinion toward that aspect. This combination helps identify aspects that are both important and emotionally significant.

8 Weakness Identification using Low-Rated Reviews

To identify weaknesses more clearly, reviews with ratings less than or equal to three were analyzed separately. This focused analysis helps isolate recurring complaints that may be overshadowed by positive reviews in the full dataset.

This step was performed at the end of the pipeline to build upon previously extracted aspects and sentiment scores rather than reprocessing the entire workflow.

9 Pipeline Flowchart Description

The overall pipeline can be summarized as:

Raw Reviews → Text Cleaning → RAKE-based Aspect Extraction → Adjective Filtering → SentiWordNet Sentiment Scoring → Negation Handling → Aspect-level Aggregation → Weakness Analysis

A visual flowchart based on this sequence can be included to clearly illustrate data flow and processing stages.

10 Conclusion

The proposed pipeline offers a transparent and interpretable approach to aspect-based sentiment analysis. Key design choices such as custom RAKE implementation, adjective-only sentiment scoring, and explicit negation handling were made to align with project requirements and improve interpretability. The final results provide actionable insights into product strengths and weaknesses, demonstrating the practical usefulness of ABSA for customer review analysis.