Singapore Airlines Review Analysis

Problem Statement

The objective of this analysis is to model topics from Singapore Airlines reviews using BERTopic combined with KeyBERT-Inspired embeddings and Maximal Margin Relevance (MMR). The aim is to uncover underlying themes in customer feedback and provide detailed interpretations of the results at each analysis step.

Required Steps

Data Preprocessing

The text should be cleaned through the following processes:
- Removal of unnecessary characters (punctuation, special symbols, numbers)
- Tokenization of reviews and elimination of common stopwords
- Reduction of words to their base forms through lemmatization
- Conversion of all text to lowercase for consistency

Expected Output: A thoroughly cleaned and preprocessed review dataset ready for modeling purposes.

BERTopic Model Training with KeyBERT-Inspired Embeddings

The following steps should be executed:
- Generation of dense and context-aware word and document representations using KeyBERT-Inspired embeddings
- Utilization of these embeddings to capture the most semantically relevant keywords for each review
- Training of the BERTopic model using the generated embeddings

Expected Output: A trained model capable of topic identification based on semantic information.

Topic Refinement through MMR

The following refinements should be implemented:
- Re-ranking of documents and topic terms using MMR
- Enhancement of topic coherence while minimizing redundancy
- Selection of representative and diverse topics through careful balance

Expected Output: A refined set of topics with enhanced coherence and clear differentiation.

Results Visualization and Analysis

The following visualizations should be created:
- Topic Modeling Results: Identified topics displayed with their associated keywords, topic distribution across reviews visualized, and a bar plot or pie chart representing topic frequencies

- Topic Embedding Visualization: High-dimensional embeddings reduced to 2D representation, UMAP or t-SNE plots created, clear topic labels added to the visualization, and relationships between topics depicted
- Topic Word Clouds: Word clouds for the top topics generated, term frequency and salience represented visually, and at minimum, the three most prevalent topics visualized

Expected Output: Comprehensive visualizations with detailed interpretations.

Topic Interpretation

For each identified topic:
- Assignment of labels based on keyword patterns
- Provision of interpretations
- Consideration of context in the analysis

Example: When keywords such as "staff," "helpful," and "friendly" are observed, the topic should be labeled as "Customer Service."

Expected Output: Clear, human-readable interpretations documented for each topic.

Final Analysis and Recommendations

The following elements should be included:
- A comprehensive summary of identified topics
- Analysis of topic prevalence
- Extraction of key insights from the analysis
- Identification of areas for improvement
- Highlighting of positive aspects of service
- Provision of specific, actionable recommendations to Singapore Airlines

Task Objective

The task is to apply topic modeling techniques to Singapore Airlines reviews, with results visualized and interpreted at each step. The final output should include actionable insights that can be implemented by the airline to enhance service quality. Note: Each step should be documented thoroughly, and any challenges or limitations encountered during the analysis should be noted.