Product Review Analysis

Topic Analysis of Review Data

A leading mobile wants to understand the customer's voice by analyzing the reviews of the product on Amazon and the topics that customers are talking about.

Problem Statement

The company Lenovo, the mobile phone manufacturer has launched its budget smartphone in the Indian market. The client wants to understand the VOC (voice of the customer) on the product. This will be useful to not just evaluate the current product but to also get some direction for developing the product pipeline. The client is particularly interested in the different aspects that customers care about. Product reviews by customers on a leading e-commerce site should provide a good view.

Solution

Dataset

The filename ‘K8 Reviews v0.2.csv’ contains the required dataset for this analysis. The CSV file contains the following columns

- Sentiment: The sentiment against the review.
- Reviews: The main text of the review.

Implementation

Discover the topics in the reviews and present them to businesses in a consumable format. Employ techniques in syntactic processing and topic modelling. Perform specific cleanup, POS tagging, and restricting to relevant POS tags, then, perform topic modelling using LDA. Finally, give business-friendly names to the topics and make a table for business.

1. Read the .csv file using Pandas. Take a look at the top few records.
2. Normalize casings for the review text and extract the text into a list for easier manipulation.
3. Tokenize the reviews using NLTKs word_tokenize function.
4. Perform parts-of-speech tagging on each sentence using the NLTK POS tagger.
5. For the topic model, we should want to include only nouns.

   - Find out all the POS tags that correspond to nouns.
   - Limit the data to only terms with these tags.
6. Lemmatize.
   - Different forms of the terms need to be treated as one.
   - No need to provide a POS tag to the lemmatizer for now.
7. Remove stopwords and punctuation (if there are any).
8. Create a topic model using LDA on the cleaned-up data with 12 topics.
   - Print out the top terms for each topic.
   - What is the coherence of the model with the c_v metric?
9. Analyze the topics through the business lens.
   - Determine which of the topics can be combined.
10. Create a topic model using LDA with what you think is the optimal number of topics
   - What is the coherence of the model?
11. The business should be able to interpret the topics.
   - Name each of the identified topics.
   - Create a table with the topic name and the top 10 terms in each to present to the business.

Output

The output is included in lenovo_topic_analysis_reviews_output.pdf.

Src

The code is included in lenovo_topic_analysis_reviews_src.pdf.