

New Smartphone Review Analysis/Topic Modelling

Description:

Help a leading mobile brand understand the voice of the customer by analyzing the reviews of their product on Amazon and the topics that customers are talking about. You will perform topic modeling on specific parts of speech. You'll finally interpret the emerging topics.

Problem Statement:

A popular mobile phone brand, Lenovo has launched their 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.

Introduction:

The aim of this project was to analyze customer reviews of Lenovo's budget smartphone on Amazon to understand the Voice of the Customer (VOC). By performing topic modeling on specific parts of speech (POS), the goal was to identify key themes and topics discussed by customers, which could provide valuable insights for Lenovo's product development and marketing strategies.

Data Preparation:

- The dataset 'K8 Reviews v0.2.csv' was loaded using Pandas, and initial exploratory data analysis (EDA) was conducted to understand its structure and contents.
- Duplicates in the dataset were identified and removed to ensure data integrity.
- Basic statistics such as the distribution of sentiment (positive vs. negative reviews) were analyzed to gain insights into overall customer feedback.

Text Preprocessing:

- Text preprocessing is a crucial step in natural language processing (NLP) tasks. The following preprocessing steps were performed:
- Normalization: Convert text to lowercase to ensure consistency.
- Tokenization: Split reviews into individual words or tokens.

- Part-of-Speech (POS) Tagging: Tag each word with its corresponding POS (noun, verb, adjective, etc.) using the NLTK library.
- Filtering for Nouns: Retain only nouns from the POS-tagged text, as they often represent key topics or entities.
- Lemmatization: Reduce words to their base or root form to standardize vocabulary.
- Removal of Stopwords and Punctuation: Exclude common stopwords (e.g., 'the', 'is') and punctuation marks to focus on meaningful content.
-

Topic Modeling with LDA:

- Latent Dirichlet Allocation (LDA) was chosen for topic modeling, a popular technique for identifying topics in a collection of text documents.
- A dictionary was created and filtered to include only relevant words based on frequency thresholds.
- LDA models were trained with varying numbers of topics, and the optimal number of topics was determined using the coherence score, a measure of topic coherence.
- The final LDA model with the optimal number of topics was generated, and the top terms for each topic were printed out for analysis.

Topic Interpretation:

- The topics generated by the LDA model were interpreted from a business perspective to identify meaningful themes and insights.
- Each topic was named based on the key terms associated with it, reflecting the main themes discussed by customers in the reviews.

Visualization and Interpretation:

- PyLDAvis was used to visualize the topics in an interactive manner, enabling exploration of topic distributions and word relevance within each topic.
- Word clouds were generated to visually represent the most frequent terms in the reviews, providing additional insights into customer sentiment and preferences.
- t-SNE dimension reduction was applied to visualize the distribution of topics in a lower-dimensional space, facilitating a deeper understanding of topic relationships.

Results and Recommendations:

- The LDA models successfully identified key topics discussed by customers in Lenovo's mobile phone reviews, providing valuable insights into customer sentiment and preferences.
- By analyzing the topics, Lenovo can gain a better understanding of customer needs and preferences, informing product development, marketing strategies, and customer engagement initiatives.

- Continuous monitoring of customer reviews and updates to the topic model can help Lenovo stay informed about evolving trends and sentiments, allowing for timely adjustments to product offerings and marketing campaigns.

Conclusion:

- This project demonstrated the effectiveness of text analytics and topic modeling techniques in extracting meaningful insights from customer reviews.
- By understanding the Voice of the Customer, Lenovo can make informed decisions to improve product quality, customer satisfaction, and brand loyalty.
- Further research and analysis could focus on sentiment analysis, trend detection, and sentiment-driven product improvements to enhance the overall customer experience.

Future Directions:

- Explore advanced NLP techniques such as sentiment analysis, aspect-based sentiment analysis, and emotion detection to gain deeper insights into customer feedback.
- Incorporate data from other sources such as social media, customer surveys, and customer support interactions to enrich the analysis and capture a more comprehensive view of customer sentiment.
- Implement machine learning models for predictive analytics to anticipate customer needs and preferences, enabling proactive decision-making and personalized customer experiences.
- members for further inquiries, collaborations, or feedback.

This detailed project report provides a comprehensive overview of the methodology, findings, and recommendations derived from the analysis of customer reviews for Lenovo's mobile phone product. It serves as a valuable resource for stakeholders, decision-makers, and researchers interested in understanding customer sentiment and leveraging data-driven insights for business decision-making.

- Paras Agrawal

Prefinal yearite at Vellore Institute of Technology, Vellore

Computer Science and Engineering Specialization in Information Security