Customer Review Summarization: An AI Tool for Summarizing Product Reviews

Abstract

This project focuses on developing an AI-based tool to summarize customer reviews for products. Given the vast number of reviews available online, manually reading them is impractical. This tool uses natural language processing (NLP) to extract and summarize the essence of user feedback, making it easier for consumers and companies to understand public sentiment.

Introduction

Online reviews significantly influence consumer decision-making. However, large volumes of reviews can overwhelm potential buyers. This research aims to address this problem by creating an automated summarization tool that distills lengthy reviews into concise summaries.

Related Work

Previous studies have explored extractive and abstractive text summarization techniques. Popular models like TextRank, BERT, and T5 have shown promising results. Our approach combines both extractive and abstractive methods for better accuracy and coherence.

Dataset and Preprocessing

We used the 'Amazon Product Reviews' dataset from Kaggle, which contains thousands of customer reviews across multiple product categories. The dataset was uploaded via the Kaggle API. Preprocessing steps included removing stopwords, punctuation, and performing tokenization and lowercasing.

Methodology

The methodology included data cleaning, tokenization, and vectorization of text. We used pre-trained transformer models to generate summaries. ROUGE scores were used to evaluate summary quality. Python libraries like NLTK, spaCy, and the Hugging Face Transformers library were used.

Evaluation

Summarization performance was evaluated using the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric. Higher ROUGE scores indicate better summary coverage and coherence. We also conducted qualitative evaluations.

Results and Discussion

The tool successfully summarized lengthy reviews into short, meaningful statements. The summaries captured the core sentiment and insights of the original text. Challenges included handling reviews with mixed sentiments and technical product descriptions.

Conclusion and Future Work

This project demonstrates the potential of AI in summarizing customer reviews efficiently. In future, we aim to improve semantic understanding, handle multilingual data, and integrate the summarizer into e-commerce platforms for real-time usage.

References

1. Lin, C.Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries.
2. Vaswani et al. (2017). Attention is All You Need.
3. Devlin et al. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
4. https://www.kaggle.com/datasets/arhamrumi/amazon-product-reviews