Smart Review Analyzer: Enhancing Online Shopping with NLP-Based Sentiment Analysis and Summarization

Yashwanth Samudrala
yash0512@bu.edu

Anna Finn
afinn12@bu.edu

Abstract

Online shopping expansion created important review roles for customer judgments while the massive amount of reviews exceeds human analytical capabilities. The project makes use of Natural Language Processing (NLP) to create an automated review assessment through the analysis of sentiment combined with topic recognition and summary generation. The Amazon Reviews datasets from Kaggle enable our model to carry out 3 tasks: positive or negative review assessment, attribute recognition, and structured summary creation with praise and criticism themes. The review processing will utilize the combination of BERT, TF-IDF, and TextRank models to achieve effective insight extraction. Users obtain brief yet comprehensive reviews through this tool which helps them make efficient shopping decisions and build solid buying choices and improve their e-commerce journey. The structured feedback process enables businesses to improve their products by assessing customer sentiments. The system will obtain multilingual functionality and aspect-based sentiment investigation with e-commerce market integration as upcoming development steps to improve convenience and accessibility.

1 Introduction

Online shopping growth necessitates that product reviews serve as vital tools which users rely on before buying products. The massive number of product reviews creates a problem because customers need extended periods to process product quality through analysis. The application uses Natural Language Processing methods to analyze Amazon product evaluation texts and then classify them while automatically generating summaries and extracting important points.

The model conducts sentiment analysis to produce both review classifications of positive and negative feedback along with percentage distributions. Review topic extraction creates groups that organize feedback according to what buyers frequently mention in their reviews concerning aspects like durability, quality, and ease of use. The final segment includes summarization that presents essential product merits and drawbacks in a short summary.

The training data will come from multiple data sets of Amazon Reviews provided on Kaggle (containing over 100k-1M samples each) while sentiment analysis and key phrase detection derives from star ratings along with textual content. The summarization will be trained using this data as well. The tool will help customers reach buying decisions by replacing manual review reading with automated assessment.

2 Project Goal

The core purpose of this project delivers streamlined shopping online through structured assessments of product reviews. The new system generates an organized review summary consisting of popular compliments and complaints about the product which eliminates the need for manual search through numerous customer feedback.

Customers buying a smartphone seek information about its battery performance together with camera capabilities and durability features. Our platform automatically interprets multiple reviews for users by extracting common themes which include:

- "Great camera but poor battery life."

- "Lightweight and durable, but the screen scratches easily."

- "Amazing performance, but it overheats quickly."

The systematic overview of client feedback makes analysis more effective because it reveals both frequent negative comments and positive aspects found within customer commentary. The systematic customer opinion presentation will provide companies with organized feedback so they can understand their products better for service enhancement.

3 User Interaction

Our system will function by taking in a collection of customer reviews and producing a structured output that highlights key insights.

Example Input:
A list of product reviews, such as:

- "The phone has a great camera but the battery life is terrible."

- "Very durable and lightweight, but the screen scratches easily."

- "Amazing performance! The battery could be better, though."

- "The phone overheats after long use, but the display is fantastic."

- "Excellent camera, fast performance, but the charging speed is slow."

Example Output:
Overall Product Summary:
This smartphone is praised for its camera quality, lightweight design, and fast performance. However, customers have reported issues with battery life, overheating, and screen durability.

Sentiment Breakdown:
Positive Reviews: 75%
Negative Reviews: 25%

Key features identified:
Positive Features:

- "High-quality camera with sharp image clarity."

- "Lightweight and durable build, easy to carry."

- "Fast performance for multitasking and gaming."

Negative Features:

- "Battery life drains quickly with heavy usage."

- "Overheating issues were reported after prolonged use."

- "The screen scratches easily, requiring a protective case."

With this structured output, shoppers can quickly grasp a product's strengths and weaknesses without spending excessive time reading reviews.

4 Technical Approach

The reviews will be analyzed with Natural Language Processing (NLP) and Machine Learning (ML) processing since they act as the project's main text analytic components. Our key steps will include:

Data Collection and Preprocessing
We will use the Kaggle web platform to acquire Amazon product reviews. Text data must undergo cleaning and preparation by removing stopwords and punctuation and removing sections with no significance for analysis. The processed data will then be normalized and tokenized to make it suitable for sentiment analysis along with topic extraction.

Sentiment Analysis
The text contains labeled sentiment values where 1-2 stars represent negative sentiment while 4-5 stars deliver positive sentiment. The model will also consider word connotation in its analysis. A model for classifying reviews should be trained with Naïve Bayes, Logistic Regression, or Transformer-based models (BERT) as its algorithm. Each product receives analysis to determine its percentage of positive feedback together with negative comments.

Topic Extraction
Crucial keywords and phrases within the dataset can be identified through NLP methods that include TF-IDF (Term Frequency-Inverse Document Frequency) in conjunction with Named Entity Recognition (NER). The important product features, including "battery life," "screen quality" and "customer service," should be organized under separate topics. The analysis tool groups together these related phrases under their respective logical meaning groups.

Summarization

The application implements TextRank alongside T5 Text-to-Text Transfer Transformer or TextRank for structured summary generation. The model will construct a simple paragraph which contains major beneficial and adverse factors and establish a point-list organization for fast review assessment.

Edge Cases

This model will not consider "fake" or bot reviews, but will identify edge cases of extreme negative effects such as:

- "It exploded when I plugged it in."

- "The package was empty."

- "The product came moldy."

A severity rating will be computed through the text and star ratings, while a frequency rating will measure how often these critical issues are reported. By balancing these two factors, the model will decide whether to include such reviews in the final output.

5   Applications

This project has valuable applications for both consumers and businesses.

For Customers: Users save time when they can determine product reputation without reading through an extensive number of reviews. The decision-making process becomes better because customers can evaluate products through summarized attribute data. Retrieve a better shopping experience because it neutralizes disagreements that emerge from ambivalent customer reviews.

For Businesses: A structured feedback process enables brands to detect recurring problems from their customers. Through product improvement companies can find and solve essential problems in their products. The identification of consumer preferences enables organizations to reach optimal product designs for competitive market advantages.

6   Future Scope

Multilingual Support: Expanding the model to analyze reviews in multiple languages.

Aspect-Based Sentiment Analysis (ABSA): Categorizing sentiment based on product aspects (e.g., battery, display, sound).

Personalized Summaries: Customizing summaries based on user preferences.

Integration with E-commerce Websites: Developing browser extensions or API integrations for real-time review analysis.

7   Conclusion

Modern shoppers depend increasingly on online shopping, making it essential for product review assessments to be more accessible. Using NLP technology, we are developing a system designed to provide both shoppers and businesses with an efficient way to evaluate and enhance products.

The project aims to strengthen customer experience and support decision-making through sentiment analysis, topic extraction, and review summarization in order to enhance the online shopping experience.