AI-Powered-Insights-Sentiment-and-Topic-Analysis-of-Movie-Reviews

Most Popular Topics in Positive and Negative Sentiments in Amazon Movies and TV Reviews Dataset

Author: Rishikesh Kakde

Introduction

Analyzing customer reviews has become indispensable for businesses to understand consumer preferences, product quality, and service satisfaction. With the rise of e-commerce platforms like Amazon, customers share their experiences widely, offering a goldmine of data for analysis. Sentiment analysis and topic modeling are two effective techniques to extract actionable insights from large-scale textual data.

Previous studies, such as Yasen and Tedmori (2019), demonstrated that analyzing movie reviews can uncover patterns in audience satisfaction and dissatisfaction. Their research showed how sentiment classification and topic identification could provide valuable insights into user feedback. Building on this, the present study focuses on identifying the most popular topics in both positive and negative sentiments using reviews from Amazon's *Movies and TV* subset for the year 2023. The dataset, sourced from McAuley and Leskovec (2023), includes millions of user reviews, making it an ideal resource for this analysis. A dual approach was adopted, combining supervised sentiment classification using BERT and unsupervised topic modeling using LDA to explore the themes that dominate customer feedback.

Research Question

What are the key topics in positive and negative sentiments in Amazon *Movies and TV* reviews from 2023?

Methodology

Data

The dataset used in this study is the *Movies and TV* subset of the Amazon Reviews dataset, collected in 2023 by the McAuley Lab. The complete dataset comprises over 17 million reviews, providing a rich and diverse range of opinions on movies, TV shows, and related products. For the purpose of this study, first 1,500 reviews was extracted to serve as the test dataset, ensuring a manageable sample size for analysis. Additionally, next 2,000 reviews were collected to create a training dataset for supervised sentiment classification.

The choice of the *Movies and TV* subset was motivated by its relevance to the entertainment industry, where customer reviews play a crucial role in shaping audience behavior and production standards. The reviews in this dataset are user-generated and span a wide variety of movie genres, TV shows, and products related to entertainment. Since the original dataset lacked explicit sentiment labels, the EMPATH library was used to generate initial sentiment annotations, providing positive, negative, and neutral sentiment tags to the training dataset.

Data collection and extraction were performed using Python libraries such as PANDAS for data handling and EMPATH for sentiment labeling. The timeframe of the dataset is constrained to 2023, ensuring that the insights are recent and relevant to current consumer trends. By using this focused timeframe, the study captures contemporary opinions, allowing for an up-to-date understanding of audience preferences and sentiments.

Analysis

The analysis process involved a dual approach: supervised sentiment classification using a pretrained BERT model and unsupervised topic modeling using Latent Dirichlet Allocation (LDA).

In the supervised sentiment classification phase, the following steps were performed:

1. Data Labeling: Initial sentiment labels for the training dataset (2,000 reviews) were generated using the EMPATH library, which classified reviews into positive, negative, and neutral sentiments.
2. Model Fine-Tuning: A pretrained BERT model was fine-tuned on the labeled training set. The model was trained for three epochs using optimal hyperparameters, including learning rate, batch size, and weight decay. Details for the hyperparameters can be found in the supplementary materials. Fine-tuning was performed to adapt the BERT model to the nuances of the *Movies and TV* dataset, balancing computational constraints with model performance.
3. Evaluation: The fine-tuned BERT model was evaluated using accuracy and F1-score metrics, achieving approximately 80% alignment with the EMPATH sentiment tags. This evaluation can be found in the supplementary material.
4. Classification: The trained BERT model was applied to the test dataset of 1,500 reviews, categorizing each review into positive, negative, or neutral sentiments.

The topic modeling process consisted of the following steps:

1. Text Preprocessing: The review text was preprocessed to improve model accuracy and performance. Preprocessing steps included, Stopword removal, Lemmatization, Tokenization, and TF-IDF Vectorization.
2. Optimal Topic Selection: The coherence score was calculated to determine the optimal number of topics for both positive and negative reviews. The coherence score measures the interpretability and semantic consistency of the topics, ensuring meaningful clustering of terms.
3. Topic Modeling: Latent Dirichlet Allocation (LDA), an unsupervised topic modeling algorithm, was applied to the positive and negative subsets of the reviews. The LDA model identified the dominant topics within each subset, providing insight into the most discussed themes.
4. Visualization: The results of the LDA topic modeling were visualized using PyLDAvis,. PyLDAvis displays an Intertopic Distance Map, which shows the relationship between topics, and a ranked list of the Top-30 Most Relevant Terms for each topic. The visualizations illustrate the prominent themes in positive and negative sentiments. The PyLDAvis results were exported as HTML files for further exploration and are included in the supplementary materials.

Results

The sentiment classification phase categorized the reviews into three classes – Positive (37%), Negative (16%), and Neutral (47%). This analysis focuses on **positive** and **negative** sentiments, as these categories provide actionable insights for stakeholders.

1. Positive Sentiments

The analysis of positive sentiments revealed dominant themes centered around "great movies," "engaging stories," and "well-made productions," with the most frequently occurring terms being "great," "movie," "love," "character," and "story." Topic modeling identified three distinct topics: Engaging Storylines, where reviews praised captivating plots and well-developed characters, highlighting emotional connections through terms like "story," "love," and "character"; Cinematic Quality, where words like "great," "film," and "awesome" reflected appreciation for production value, direction, and visual appeal; and Entertainment Value, where terms like "enjoyable," "fun," and "well-made" underscored the entertainment factor of the content.

2. Negative Sentiments

The analysis of negative sentiments revealed dominant themes related to "repetitive content," "predictable storylines," and "poor quality productions," with frequently occurring terms such as "quot," "boring," "show," "watch," and "time." Topic modeling identified three key areas: Lack of Originality, where many reviewers criticized clichéd plots and unoriginal themes, using words like "boring," "predictable," and "repetitive"; Poor Execution, where terms like "show," "watch," and "time" reflected frustration with underwhelming pacing, acting, or low production value; and Customer Frustration, where specific words like "andy," "life," and "could" indicated disappointment with content that lacked emotional or narrative engagement.

Conclusion and Limitations

This study identified key themes in consumer sentiments within the Movies and TV category on Amazon. Positive reviews underscored the importance of cinematic quality, while negative reviews highlighted dissatisfaction with predictability and product presentation. These findings align with existing literature and highlight the value of customer feedback in shaping industry standards.

However, like every study, the study has limitations. The analysis was based on a subset of 1,500 reviews, which may not fully represent the dataset's diversity. Additionally, while the BERT model effectively classified sentiments, its performance is limited by the quality of initial labels provided by EMPATH. The limited computing power and RAM that I have on my computer limited the number of epochs and batchsize that could've proven helpful in better classification of the sentiments. The model results align about 80% with EMPATH library results on the same dataset (See supplementary material for code and results). More training would've helped the model learn better features and context to classify the reviews. Future research could leverage a larger dataset, explore multilingual reviews, and incorporate advanced models for deeper insights.

References

- Yasen, M., & Tedmori, S. (2019). *Movies Reviews Sentiment Analysis and Classification*. 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT).
- McAuley, J., & Leskovec, J. (2023). *Amazon Reviews Dataset*. Retrieved from: https://amazon-reviews-2023.github.io.