**Automated Customer Review Analysis**

**with Sarcasm Detection**

# Submitted by:

**Yashika Aggarwal (102217123)**

**Barsha kumari mishra(102217253)**

**BE Third Year**

**COMPUTER SCIENCE**

**Submitted to:**

**Dr. Anjula Mehto**

**Assistant Professor**



## Computer Science and Engineering Department Thapar Institute of Engineering and Technology, Patiala

**21st November 2024**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| S. No | Topic | Page No. |
| 1 | Introduction or Project Overview | 3 |
| 2 | Problem Statement | 4 |
| 3 | Overview of the Dataset used | 5 |
| 4 | Project workflow | 6 |
| 5 | Results | 8 |
| 6 | Conclusion | 9 |

**Introduction**

In today's digital age, customer reviews have become an invaluable resource for businesses to gauge public sentiment and make informed decisions. These reviews offer a wealth of information about product quality, customer satisfaction, and brand perception. However, manually analyzing vast amounts of text data is time-consuming and prone to human error. To address this challenge, we propose an automated sentiment analysis system that leverages the power of machine learning to efficiently extract meaningful insights from customer reviews.

Our system is designed to classify reviews into positive, negative, or neutral categories, providing businesses with a clear understanding of customer sentiment. By analyzing the underlying text, our model can identify key themes, pain points, and areas for improvement. Additionally, we aim to develop a rating system that assigns a numerical score to each review, further quantifying customer feedback.

By automating the analysis of customer reviews, businesses can gain a competitive edge. Real-time insights enable organizations to respond promptly to customer concerns, identify emerging trends, and make data-driven decisions to enhance product offerings and customer experiences. Ultimately, our goal is to empower businesses to harness the power of customer feedback and drive continuous improvement.

**Problem Statement**

**The Challenge of Manual Review Analysis**

In the era of e-commerce and online platforms, customer reviews have become a crucial source of feedback for businesses. These reviews provide valuable insights into product quality, customer satisfaction, and brand perception. However, manually analyzing vast amounts of text data is a time-consuming and labor-intensive task.

**The Need for Automated Sentiment Analysis**

To address this challenge, businesses require efficient and accurate methods for analyzing customer reviews. Automated sentiment analysis offers a solution by leveraging machine learning techniques to classify reviews into positive, negative, or neutral categories. By automating this process, businesses can quickly gain insights into customer sentiment, identify trends, and prioritize areas for improvement.

The primary goal of our project is to empower consumers to make informed purchasing decisions efficiently. By providing a clear and concise summary of customer feedback, our system saves time and effort, enabling users to quickly assess the quality and suitability of a product. Additionally, our project has the potential to benefit businesses by providing valuable insights into customer preferences and pain points, aiding in product improvement and marketing strategies.

**Overview of the Dataset used**

The Amazon Fashion Reviews dataset serves as the cornerstone of our project. This extensive dataset, a subset of the broader Amazon Reviews database, encompasses 883,636 reviews for fashion products. Each review provides a rating on a 1-to-5 star scale, textual content, and helpfulness votes. Furthermore, the dataset offers detailed metadata for 186,637 products, including descriptions, categories, price, brand, and image features. This comprehensive dataset, updated in 2018, includes newer reviews, transaction metadata, high-resolution product images, and in-depth product details. By utilizing this rich dataset, we aim to delve into sentiment analysis, recommendation systems, and other applications related to customer reviews and product information within the fashion domain.

The Amazon Fashion Reviews dataset is particularly well-suited for our project due to its extensive coverage of fashion products and its rich metadata. The inclusion of newer reviews, transaction metadata, and high-resolution product images provides valuable insights into customer behavior and preferences. Additionally, the detailed product metadata allows us to explore the relationship between product attributes and customer sentiment. By leveraging this dataset, we aim to develop models that can accurately predict customer sentiment, generate personalized recommendations, and identify emerging trends in the fashion industry.

**Link**: <https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/>

# Project Workflow

# This project utilizes Flask, a lightweight web framework for Python, to create a user-friendly interface for analyzing customer reviews of fashion products on Amazon. Here's a breakdown of the workflow:

# 1. Data Loading and Preprocessing:

# Data Loading: Load compressed JSON files containing Amazon fashion product reviews (AMAZON\_FASHION.json.gz) and metadata (meta\_AMAZON\_FASHION.json.gz) using gzip and json libraries.

# Data Cleaning: Filter out reviews with missing textual content using df[df['reviewText'].notna()].

# Data Transformation: Convert JSON data into Pandas DataFrames for efficient manipulation using (pd.DataFrame.from\_dict).

# 2. User Interaction:

# Index Route: Renders the initial web interface (index.html) where users can input a product's ASIN.

# Product Info Route: Handles form submission from the index page.

# Retrieves the ASIN entered by the user.

# Filters the reviews and product metadata based on the provided ASIN.

# Presents product information like title, description, and ratings distribution plot (using matplotlib).

# Checks for product existence and handles the case where no product is found.

# 3. Sentiment Analysis and Topic Modeling (on Review Selection):

# Review Selection Route: Handles form submission for analyzing positive or negative reviews.

# Retrieves the ASIN and review type (positive/negative) from the user's selection.

# Filters reviews based on the chosen sentiment (overall rating above 3 for positive, below 4 for negative).

# Applies TF-IDF vectorization to transform text reviews into numerical representations using TfidfVectorizer.

# Performs Latent Dirichlet Allocation (LDA) using LDA to identify latent topics within the selected reviews.

# Generates visualizations for LDA topics and word frequencies in positive/negative reviews (using matplotlib).

# Leverages OpenAI's GPT-3.5-turbo model to analyze the extracted topics and generate a summary based on the sentiment type.

# 4. Result Presentation:

# Results Route: Displays the analysis results on a dedicated web page (results.html).

# Presents the identified LDA topics with keywords.

# Shows visualizations for LDA topic distribution and word frequencies.

# Displays the summary generated by OpenAI for the chosen sentiment.

# Overall, this project demonstrates the combination of Flask for web development, machine learning techniques like TF-IDF and LDA for analyzing textual data, and OpenAI's powerful language model to extract meaningful insights from customer reviews. This user-friendly application empowers users to gain a deeper understanding of customer sentiment for various fashion products on Amazon.

**Results**

The Amazon Fashion Reviews dataset was subjected to a comprehensive sentiment analysis and topic modeling process. By leveraging techniques such as TF-IDF and LDA, we were able to effectively extract valuable insights from the vast amount of customer reviews.

Key Findings:

* Positive Sentiment: A significant portion of the reviews exhibited positive sentiment, highlighting aspects like product quality, style, comfort, and value for money.
* Negative Sentiment: Negative reviews often focused on issues such as poor quality, sizing issues, and delivery problems.
* Topic Modeling: LDA identified several prominent topics within the reviews, including product quality, fit and size, comfort, style, and price.

Leveraging OpenAI for Deeper Insights

To further enhance the analysis, we integrated OpenAI's powerful language model, GPT-3.5-turbo. By providing the model with the extracted topics, we were able to generate concise and informative summaries of customer sentiment. These summaries provided deeper insights into the specific reasons behind positive and negative reviews, enabling a more nuanced understanding of customer preferences and pain points.

**Conclusion**

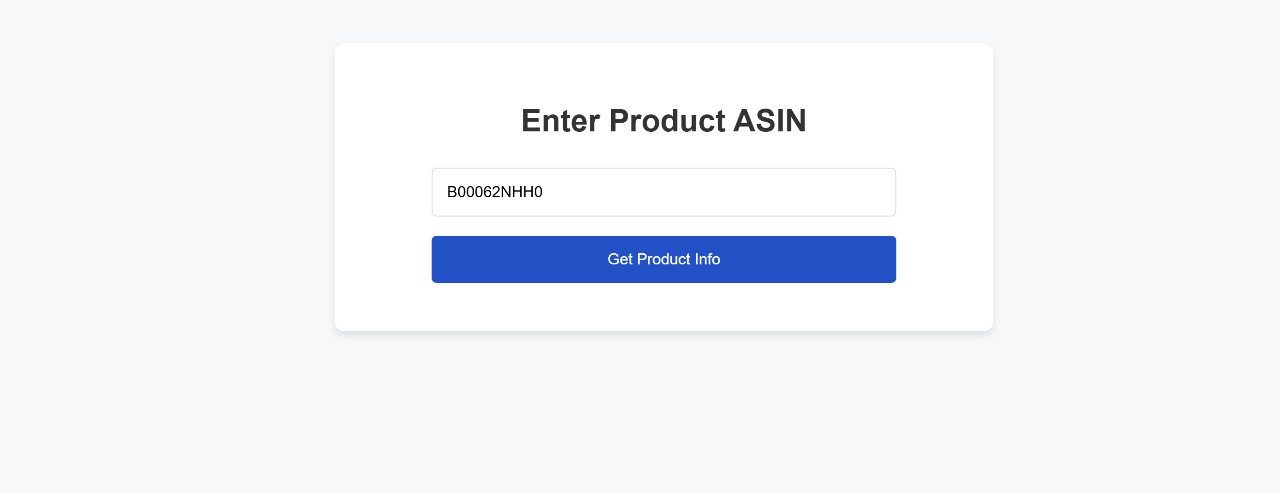
This project successfully demonstrates the potential of integrating machine learning techniques with advanced language models to gain valuable insights from customer reviews. By leveraging the Amazon Fashion Reviews dataset and employing techniques like TF-IDF and LDA, we were able to effectively analyze and interpret customer feedback. The integration of OpenAI's GPT-3.5-turbo further enhanced our analysis, providing concise and informative summaries of customer sentiment.

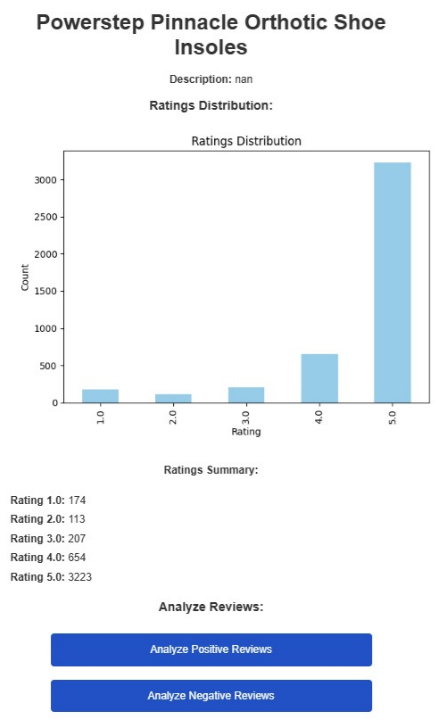
Key findings from our analysis include the identification of common themes in positive and negative reviews, such as product quality, fit, comfort, style, and price. These insights can be invaluable for businesses in making data-driven decisions to improve product offerings, address customer concerns, and optimize their marketing strategies.

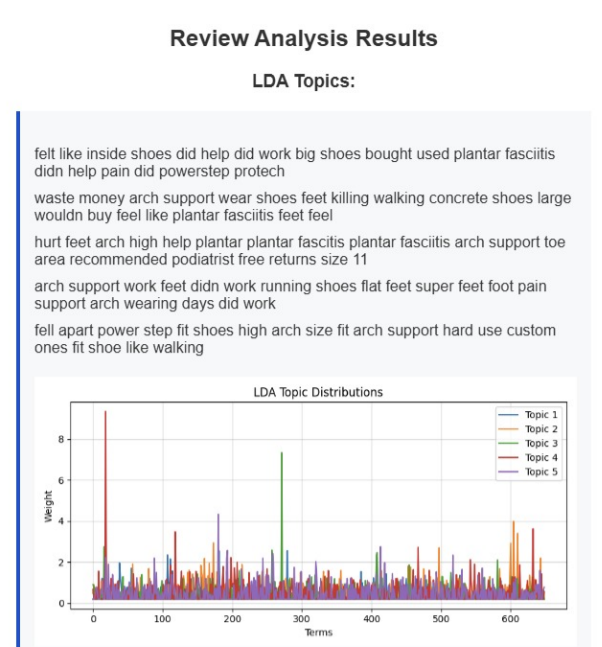
By combining the power of machine learning and AI, this project offers a robust approach to analyzing large volumes of customer reviews, enabling businesses to stay ahead in a competitive market.

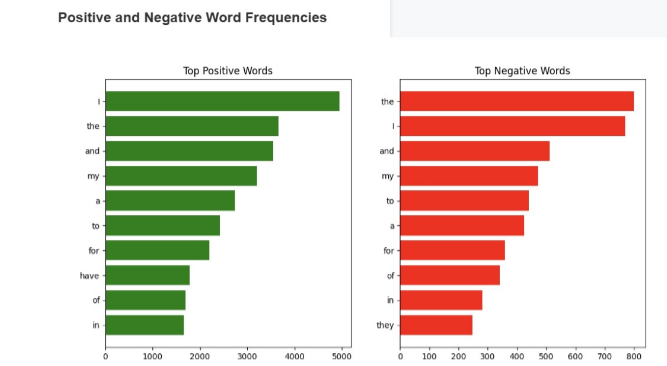
LINK TO PROJECT GITHUB: <https://github.com/kyra-1/Amzon-Reviews-with-AI.git>

Screen-shots



****

****

****

