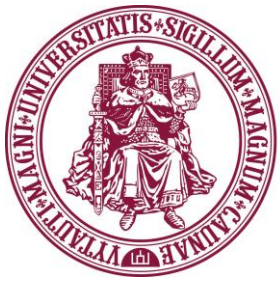


VYTAUTO DIDŽIOJO
UNIVERSITETAS
MCMXXII

TEXT MINING AND SENTIMENT ANALYSIS OF AMAZON ALEXA REVIEWS

Mohammed Irshad Kunnam Puthoor

SOCIAL MEDIA ANALYTICS
Vytautas Magnus University
Faculty of informatics

Table of Contents

1. Introduction 2
2. Methods 3
2.1 Dataset Description 3
2.2 Text Cleaning 3
2.3 Word Cloud Generation 3
2.4 Topic Modelling (LDA) 3
2.5 Visual Analysis 3
3. Results 4
3.1 Word Cloud 4
3.2 Rating Distribution 4
3.3 Sentiment Distribution 5
5. References 7

1. Introduction

One of the most significant types of user-generated content in modern e-commerce is online product reviews. Consumers create a large amount of unstructured textual data by sharing their experiences, thoughts, and satisfaction levels through reviews. Examining this data offers important insights into how consumers view a product, what features they value, and what problems they run into. With the increasing popularity of smart home devices especially voice-controlled assistants like Amazon Alexa, it is critical to comprehend consumer input to improve product design, improve user experience, and guide marketing choices.

This study uses text mining and social media analytics methods to examine 3,150 verified Amazon Alexa user reviews. The research uses natural language processing techniques to recognize often used phrases, reveal the main trends in customer conversations, and determine the emotional tone of user comments. The project aims to address important research concerns about user perception and customer happiness by integrating sentiment analysis, word frequency visualization, and subject modelling. The knowledge acquired contributes to a better comprehension of how consumers engage with Amazon Alexa and what features are most important to them.

Research Questions:
1. What is the overall sentiment of Amazon Alexa users?
2. Which words appear most frequently in feedback?

2. Methods

2.1 Dataset Description

The dataset consists of 3,150 verified Amazon Alexa customer reviews stored in TSV format. The review content, numerical rating, date, product variation, and feedback label are all included in each entry. The dataset is appropriate for sentiment analysis and text mining because of these characteristics. The panda's library was used to load the file into Python for analysis and preparation.

2.2 Text Cleaning

All review information was changed to lowercase to guarantee uniformity and prepare the text for analysis. To lessen data noise, URLs, punctuation, and special characters were eliminated. Only significant alphanumeric tokens were kept after extra spaces were normalized. The accuracy of subsequent NLP processes is increased by this preprocessing step.

2.3 Word Cloud Generation

To visually represent the most used terms in the dataset, a word cloud was made. For visualization, the cleaned text from every review was merged into a single corpus. Greater frequency and importance are indicated by larger words in the cloud. This step facilitates the rapid identification of recurrent themes and common user expressions.

2.4 Topic Modelling (LDA)

To find hidden themes in the review text, Latent Dirichlet Allocation (LDA) was used. By examining word co-occurrence patterns, LDA classifies semantically related words into topics. Three major topics that reflect the primary topics that customers discuss were extracted by the model. This method facilitates comprehension that goes beyond word frequency.

2.5 Visual Analysis

Two visualizations were produced to support interpretation of the results. A rating distribution chart shows how users rated the product from 1 to 5 stars. A sentiment distribution plot displays the number of positive, neutral, and negative reviews. These visual tools help compare numerical satisfaction with textual sentiment patterns.

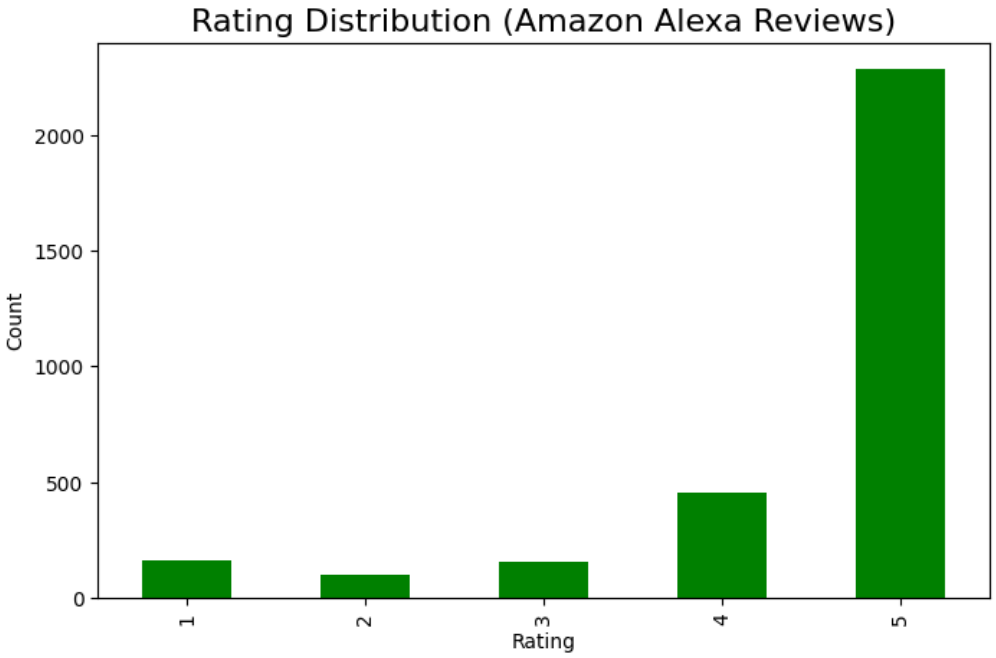
3. Results

3.1 Word Cloud

The word cloud highlights commonly used words such as 'love', 'great', 'use', 'alexa', 'echo', and 'easy'. These indicate that customers frequently describe positive experiences related to ease of use, sound quality, and device functionality.

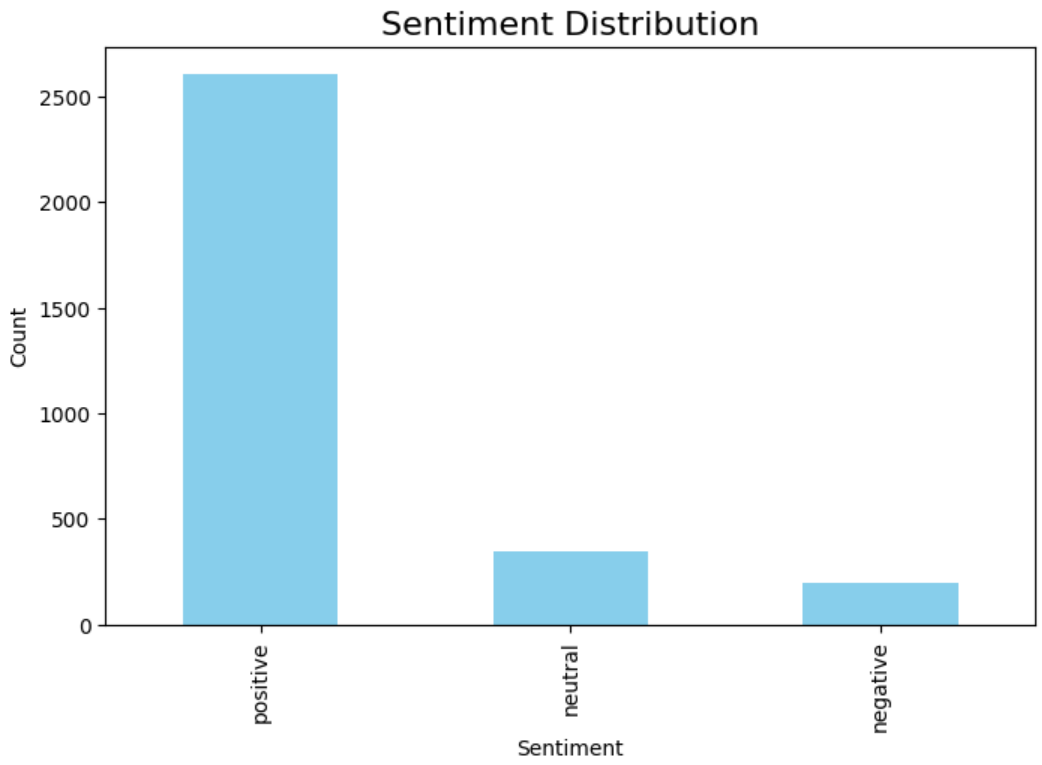
3.2 Rating Distribution

The rating distribution shows that most users gave 4 or 5 stars, reflecting high customer satisfaction. A small number of users gave 1 or 2 stars, indicating a low level of dissatisfaction.



3.3 Sentiment Distribution

The sentiment distribution confirms strong positive sentiment across reviews. Over 2,600 reviews are positive, around 350 neutral, and fewer than 220 negative. This correlates with the rating distribution and indicates an overall positive customer experience.



4. Conclusions

The results of this analysis demonstrate that customer sentiment toward Amazon Alexa is strongly positive. Sentiment classification revealed that most reviews fall into the positive category, with only a small percentage showing neutral or negative tone. This is further supported by the rating distribution, where most users awarded four or five stars. The combination of these findings suggests that customers are highly satisfied with the product's performance, reliability, and ability to meet day to day needs. The word cloud also highlights commonly used positive terms such as "love," "great," "easy," and "use," which reinforces the view that users associate Alexa with convenience and a positive experience.

Topic modelling provided more detailed insight into what customers value most. The identified topics show that users frequently discuss ease of use, sound quality, and general device functionality. These themes point to the features that matter most to customers and indicate that Amazon Alexa effectively delivers on them. Overall, the use of text mining techniques. Including sentiment analysis, word frequency visualization, and topic modelling. successfully uncovered meaningful insights into how customers perceive the device. The results demonstrate that Amazon Alexa is well-received and that customer feedback is largely enthusiastic and supportive of the product's capabilities.

5. References

[1] S. Kumar, S. Kansal, V. Sharma, and R. Talwar, "Sentiment Analysis on Amazon Alexa Reviews Using NLP and Classification," in *Proc. 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Dec. 2022, doi: **10.1109/ICAC3N56670.2022.10074086**.

[2] N. N. Vesaokar, "Sentiment Analysis of Amazon Alexa Reviews using Machine Learning," *Journal*, Aug. 2023, doi: **10.1729/Journal.35771**.

[3] S. A. Mostafa, W. Al-Dayyeni, A. Al-Shammari, A. Abdulazeez, and B. A. Khalaf, "Classification and Sentiment Analysis of Amazon Alexa Reviews," in *International Conference on Computer Science and Business*, Aug. 2024.

[4] Y. Hao, "The Use of Sentiment Analysis of User Reviews to Predict Customer Satisfaction of the Amazon Echo," *Medium*, May 17, 2020. [Online]. Available: https://medium.com.

[5] T. Escalona, "Detect sentiment from customer reviews using Amazon Comprehend," *AWS Machine Learning Blog*, Jan. 26, 2018. [Online]. Available: https://aws.amazon.com/blogs.