SENTIMENT ANALYSIS OF CUSTOMER REVIEWS FOR KINDLE BOOKS STORE

INDIVIDUAL REPORT

Joseph Do

5/8/24

MSc BUSINESS ANALYTICS

Table of Contents

I. INTRODUCTION TO DATA AND RESEARCH QUESTION

The electronic book (e-book) industry represents a substantial and continuously expanding market, with a global growth rate averaging 3.52% annually since 2017 (Wordsrated, 2023). Within this market, Amazon holds an estimated 67% share, solidifying its position as the dominant player. However, Kindle's share of Amazon book sales has experienced a decline in recent years. This report delved into a dataset encompassing over 5 hundred reviews of the Kindle Store on Amazon, aiming to gain insights into the sentiments expressed by reviewers. By analyzing these reviews, the report seeks to uncover the underlying sentiments of consumers toward digital reading materials, offering insights into customer satisfaction and content reception.

1. Research Justification

Kindle books, much like other digital goods, rely heavily on customer reviews for consumer engagement and sales. These reviews not only influence purchasing decisions but also reflect reader satisfaction and expectations. High engagement in the form of reviews can enhance a book's visibility and perceived value, driving further sales. Unlike physical goods, digital books provide a unique challenge in that consumers cannot preview the full content before purchase, making the reviews even more critical to potential buyers. Studies such as those by Katole (2022) highlight the increasing reliance on online platforms for customer purchases, where reviews form a crucial part of the consumer decision-making process.

Sentiment analysis has been widely applied in various domains to predict market trends, from stock market movements to product success in retail (Nguyen et al., 2015; Shazuli',2023) However, specific research focusing on sentiment analysis in the Kindle book market is not received much attention, presenting a unique opportunity to explore this area. The insights gained could inform not only marketing strategies but also content development and customer engagement practices.

2. Research Methodology

In this study, we employ a set of robust methodologies tailored for analyzing customer sentiment within the Amazon Kindle store, leveraging big data analytics and machine learning techniques. The objective is to extract and analyze sentiments from customer reviews to better understand consumer behavior and preferences in the digital book market.

The methodology for sentiment analysis of Kindle store's reviews through includes the following steps:

- Data Collection: The review dataset of Kindle store on Amazon is used.

- Extracting Features: Utilizing natural language processing techniques, we'll pinpoint crucial elements from the text data, encompassing word frequency, co-occurrence, and sentiment analysis.
- Data Preprocessing: The data is cleared by eliminating stop words, stemming, and converting all text to lowercase.
- Dataset Division: The dataset will be partitioned into training and testing subsets, adhering to a 70:30 ratio.
- Training ML Models: Random Forest Classifier from the scikit-learn library to predict the helpfulness of reviews based on their textual content and sentiment scores.
- Model Assessment: Performance evaluation of the models will be conducted using standard metrics like accuracy, precision, recall, and F1-score.
- Sentiment Analysis: Employing the selected model, reviews will be classified as positive, negative, or neutral.
- Visualize the outcomes: Libraries like Matplotlib, Seaborn, and WordCloud in Python are used. These tools help illustrate the distribution of sentiments across different books, the frequency of specific terms, and their association with positive or negative sentiments.

3. Research Aim and Questions

The primary research questions are:

1. How accurately can machine learning tools predict the helpfulness of Kindle book reviews based on textual content?
2. What are the sentiment trends through time?
3. Which words and phrases are most associated with Kindle book reviews, and what do they indicate about consumer sentiment?

By answering these questions, the study intends to provide a comprehensive overview of consumer sentiment in the Kindle book market, facilitating better strategic decisions for content creators and marketers in the digital publishing industry.

II. DATA PROCESSING AND EXPLORATION

The review dataset about Kindle Store on Amazon (Ni, 2018) is used in this report. The dataset has more than 500 hundred reviews with 12 variables from 2000 to 2018. More detail information about the dataset can be found in the table below:

Variables	Description	Used (Y/N)	Notes on usage
Overall	Rating given by the reviewer, usually from 1 to 5.	Y	Useful for sentiment analysis
Verified	Whether the review is from a verified purchaser.	N	
Review Time	Date the review was posted.	Y	Track the change of review rate through time
Reviewer ID	Unique identifier for the reviewer.	N	
asin	Unique code for the product listed on Amazon.	Y	Finding top 10 most reviewd products
style	Product variations like color or size.	N	
Reviewer Name	Name of the reviewer as displayed.	N	
Review Text	Full text of the review.	Y	Useful for sentiment analysis
Summary	Brief headline of the review.	N	
unixReviewTime	Timestamp of the review in Unix time.	N	
vote	Number of helpful votes the review received.	N	
image	Links to any images included with the review.	N	

Table II-1 Notes and descriptions on variables.

The table displays the name and description of each variable. Not all variables are utilized in the report; for those that are, the rationale behind their usage is also presented.

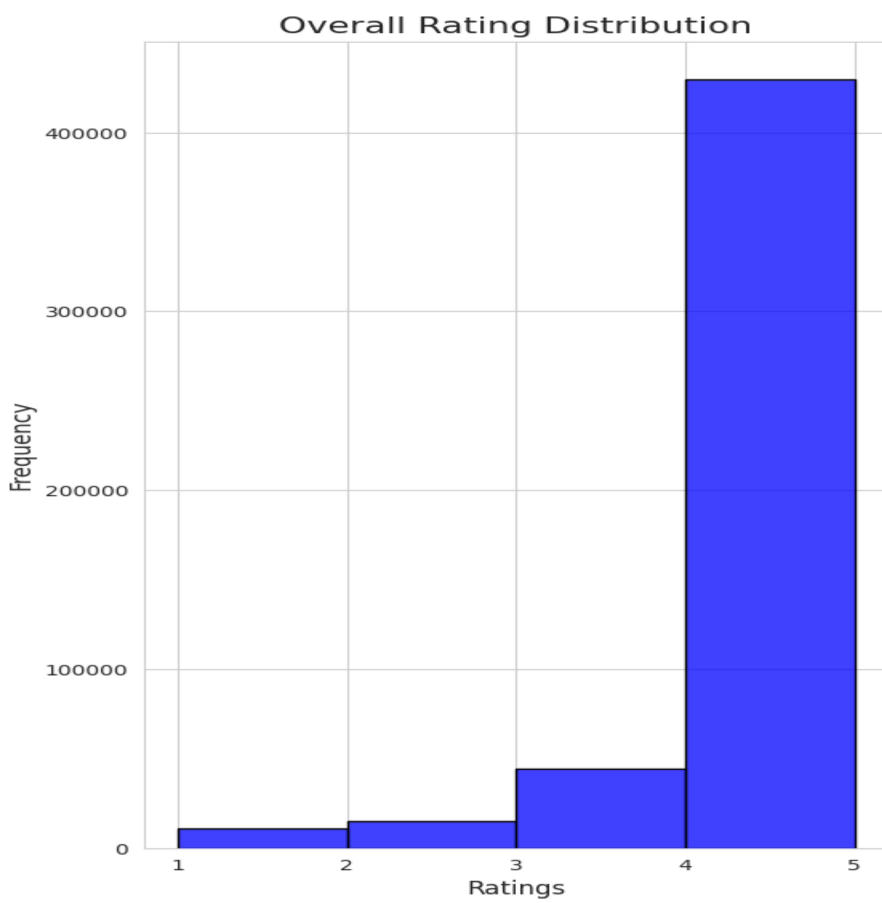

Figure II-1 Overall rating distribution.

Figure 1 demonstrates the distribution of the ‘overall’ variable – a rating from reviewers which ranges from 1 to 5. This indicates a predominantly positive reception of products, suggesting either genuine customer satisfaction or a tendency for users to review only when extremely satisfied. Ratings of 1 to 3 stars are notably less frequent, with 1-star ratings being the least common. The skew towards 5-star ratings could also suggest a review bias where feedback is predominantly positive, possibly affecting the overall perception of product quality. For a comprehensive understanding, further analysis on the review texts and the ratings across different product categories would be beneficial, alongside exploring how these ratings influence purchasing behaviors.

To gain a comprehensive understanding of the rating distribution in the dataset, a compound score is computed using the Vader Lexicon tool based on the 'reviewText' variable. This process involves transferring the 'reviewText' to a new variable called 'clean_review' before calculating the sentiment score. Initially, any user mentions in the reviews, typically denoted by '@' followed by a username in social media text, are removed. Subsequently, HTTP or HTTPS links are stripped from the text using a regular expression pattern. After these steps, special characters not part of standard English letters or apostrophes are replaced with spaces, ensuring proper word separation. The 'strip()' function is then utilized to remove leading and trailing whitespace from each review text. Finally, the cleaned reviews are filtered to retain only those exceeding 50 characters in length, aimed at excluding very short reviews that may lack sufficient context or meaningful content for further analysis.

Following that, the environment is configured in Google Colab, incorporating libraries such as NLTK, JSON, PySpark, Seaborn, WordCloud, among others, to analyze the dataset comprehensively.

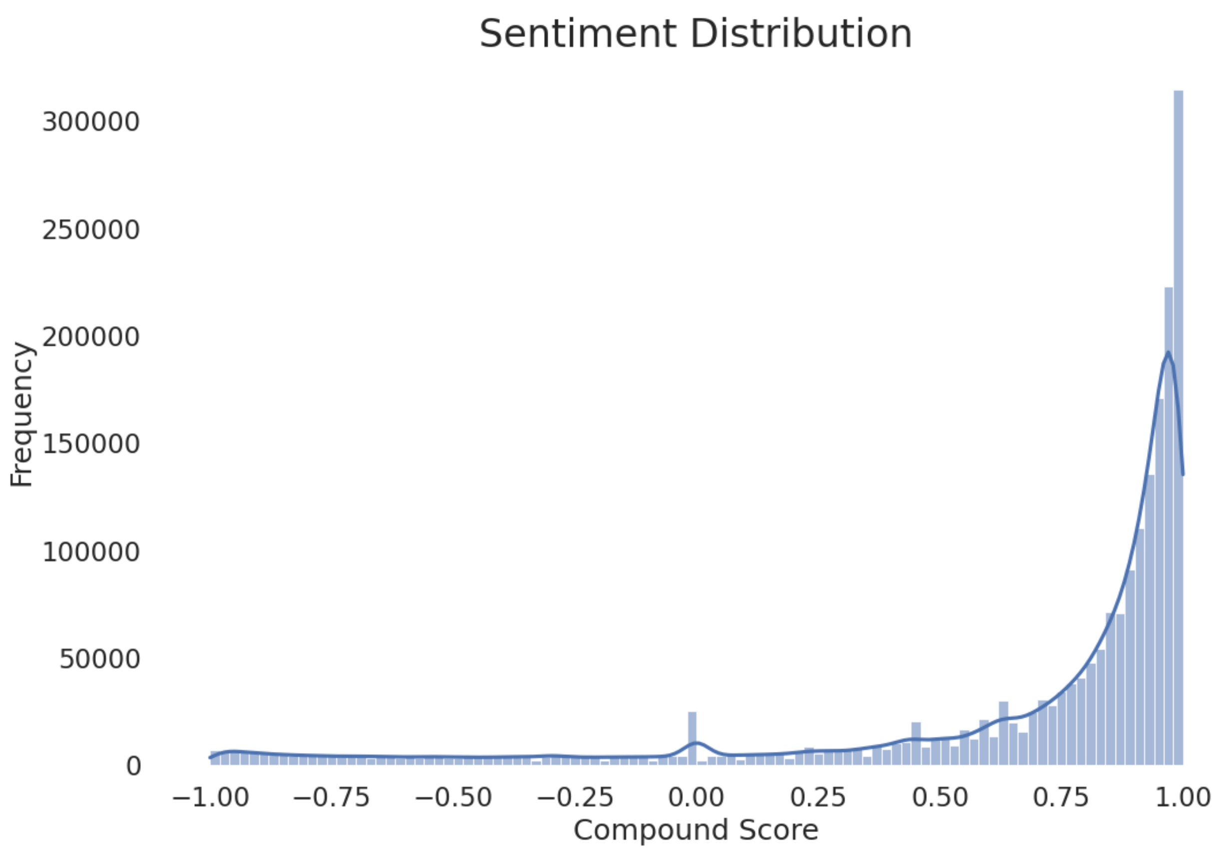

Figure II-2 Sentiment Distribution.

The new variable 'sentiment_score' which is a sentiment score from 'clean_review', stores value from -1 to +1. The minus score means negative reviews, the plus one means positive reviews. The sentiment distribution still displayed in the chart highlights a pronounced skew towards highly positive sentiment among the analyzed texts, with the compound scores clustering significantly at the extreme positive end (near 1.0), more than 250,000 reviews. Conversely, there are notably fewer instances of negative sentiment, as evidenced by the sparse bars toward the left side of the chart, particularly near the -1.0 mark. Neutral sentiments (scores around 0) and moderately positive or negative sentiments (scores around ±0.25 to ±0.75) are much less frequent compared to the extreme positive scores.

III. DATA VISUALISATION AND INTERPRETATION

In this segment, we delve into the advanced analytical outcomes, emphasizing the outcomes, elucidations, and understandings. Three research inquiries are thoroughly explored with visual aids and clarifications.

1. Research question 1 – Helpfulness of reviews

Machine learning algorithms offer scalable and efficient solutions to analyze large datasets and extract valuable insights from them. Predicting the helpfulness of reviews helps consumers make more informed decisions by highlighting which reviews are likely to provide useful information. This enhances the overall user experience and fosters trust in online platforms. Moreover, for businesses, understanding the helpfulness of reviews allows them to prioritize their responses and allocate resources effectively to address customer concerns or capitalize on positive feedback.

The dataset is organized based on the 'vote' column, retaining only instances where the value is numerical. A threshold is established using the quantile method to determine a point; if the 'vote' count surpasses this threshold, it is labeled as 1 means ‘helpful’; otherwise, it is labeled as 0. Subsequently, a new column is generated to indicate the assigned labels of 1 and 0. Then, the distribution of the new labels conducted as below:

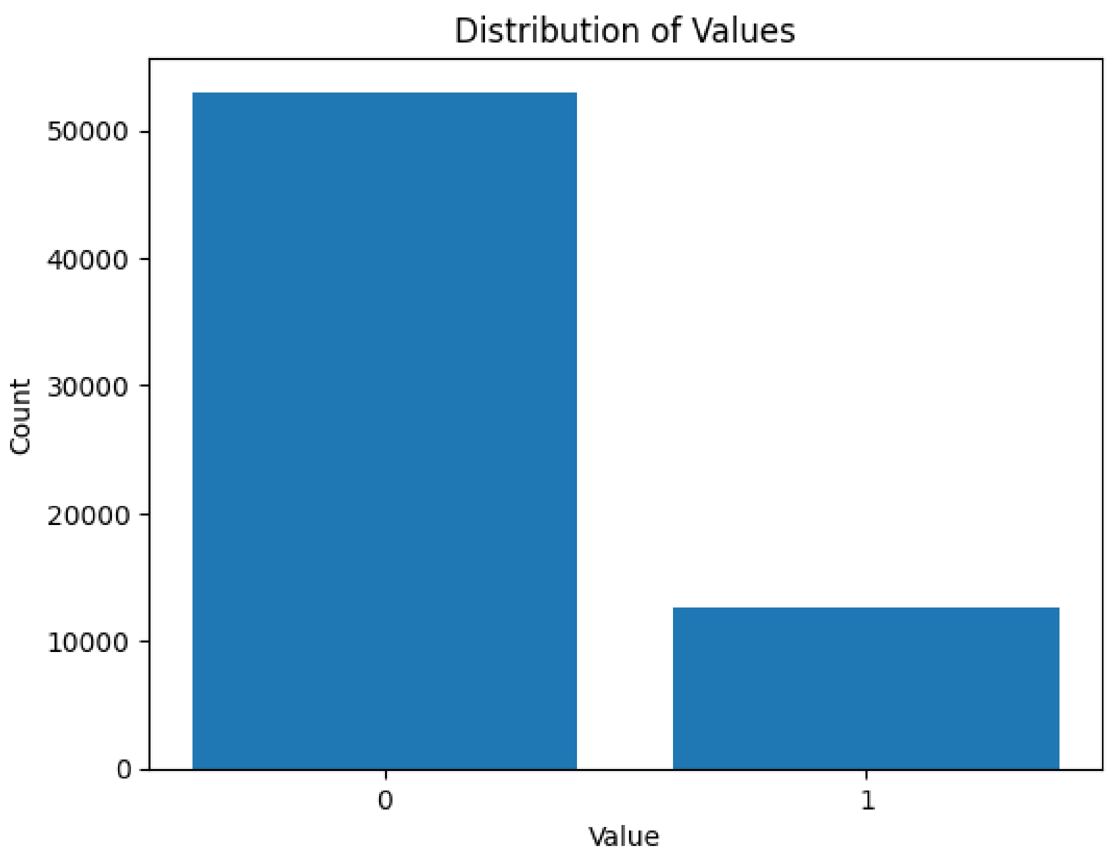

Figure III-1 Distribution of values 0 and 1.

The imbalance within the data is evident, potentially impacting the accuracy of predictions. Employing oversampling techniques on the minority class is a strategy utilized to rebalance the dataset. The ensuing outcome is presented below:

```
Accuracy: 0.7807372357475387
Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.95      0.87     10560
           1       0.30      0.10      0.15      2543

    accuracy                           0.78     13103
   macro avg       0.56      0.52      0.51     13103
weighted avg       0.71      0.78      0.73     13103
```

Figure III-2 Result from machine learning run.

The model achieves an accuracy of approximately 78.07%, which indicates that it correctly predicts the usefulness of a review about 78% of the time. While this is a decent accuracy, the more detailed metrics in the classification report reveal underlying issues with the model's performance, particularly with the minority class. For 'Non helpful reviews', at 0.81, the precision for class 0 is relatively high, meaning that when the model predicts a review is not helpful, it is correct 81% of the time. Combined with the recall index is 0.95 and the F1-Score is 0.87, it strongly suggests that the model is well-tuned to recognize reviews that are not helpful.

However, when it comes to helpful reviews, the precision drops significantly for class 1 to 0.30, meaning that only 30% of reviews predicted as helpful are helpful. This low precision indicates a high rate of false positives for helpfulness. Besides, the recall for class 1 is also low at 0.10, suggesting that the model fails to identify 90% of the truly helpful reviews. This is indicative of a model that is biased towards predicting reviews as not helpful.

The model shows good performance in identifying not useful reviews but performs poorly in recognizing useful reviews. This disparity could stem from several factors such as class imbalance, where not useful reviews significantly outnumber useful reviews, or insufficient or inadequate features that fail to capture what makes a review useful. Despite addressing the issue of imbalanced data, the model's performance fails to show improvement. This suggests that employing machine learning techniques for predicting review helpfulness may not yield effective results.

2. Research question 2 – Trend of sentiments

Monitoring the sentiment trend over time within online review datasets is crucial for every company. It provides insights into the evolving perceptions and attitudes of customers towards products or services. By analyzing sentiment trends, businesses can identify patterns, such as seasonal fluctuations or long-term shifts, which can inform strategic decision-making. Understanding these trends enables companies to proactively address emerging issues or capitalize on positive sentiment to enhance brand reputation and customer satisfaction. Besides, continuous monitoring of sentiment trends empowers businesses to stay agile and responsive in meeting the evolving needs and expectations of their customers in the dynamic online environment.

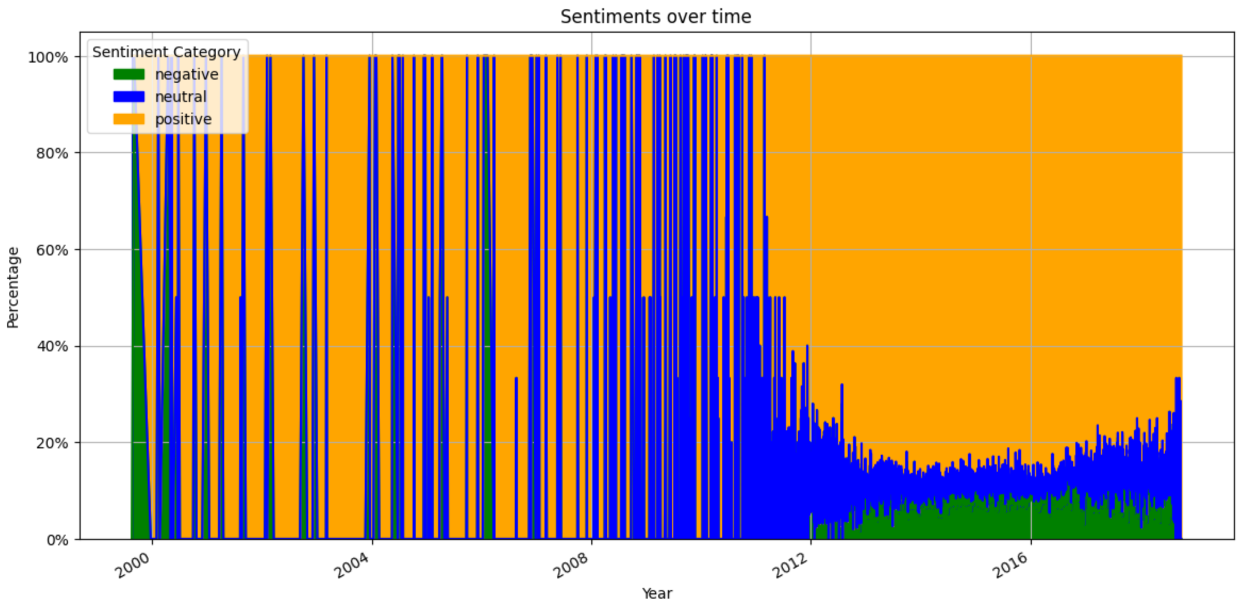

Figure III-3 Sentiment trend over time.

The graph provides a visual representation of sentiment trends over time, with data spanning from 2000 to 2018, categorized into negative (blue), positive (orange), and neutral (green) sentiments. Initially, from 2000 to around 2004, the sentiments are quite mixed, with no single category showing clear dominance. From approximately 2005 onwards, there is a marked shift, with positive sentiments (orange) becoming the overwhelmingly dominant category. This trend persists through to the end of the dataset in 2020, indicating a consistent high volume of positive sentiment over time. In contrast, negative sentiments (blue) significantly decrease after 2005 and remain minimal through the rest of the timeline. Neutral sentiments (green), while never dominant, are present throughout but remain consistently low compared to positive sentiments. This pattern may reflect Amazon's focused efforts to analyze customer feedback systematically and use these insights to enhance product offerings, specifically in the Kindle Store. By leveraging detailed review analytics, Amazon can identify and address common customer concerns and preferences, leading to improvements in the quality of books and related products.

The predominance of positive sentiments over an extended period also indicates that these improvements are well-received by customers, contributing to a higher satisfaction level with the products purchased from the Kindle Store. The consistent decrease in negative sentiments could be indicative of successful interventions by Amazon to rectify issues that previously led to customer dissatisfaction.

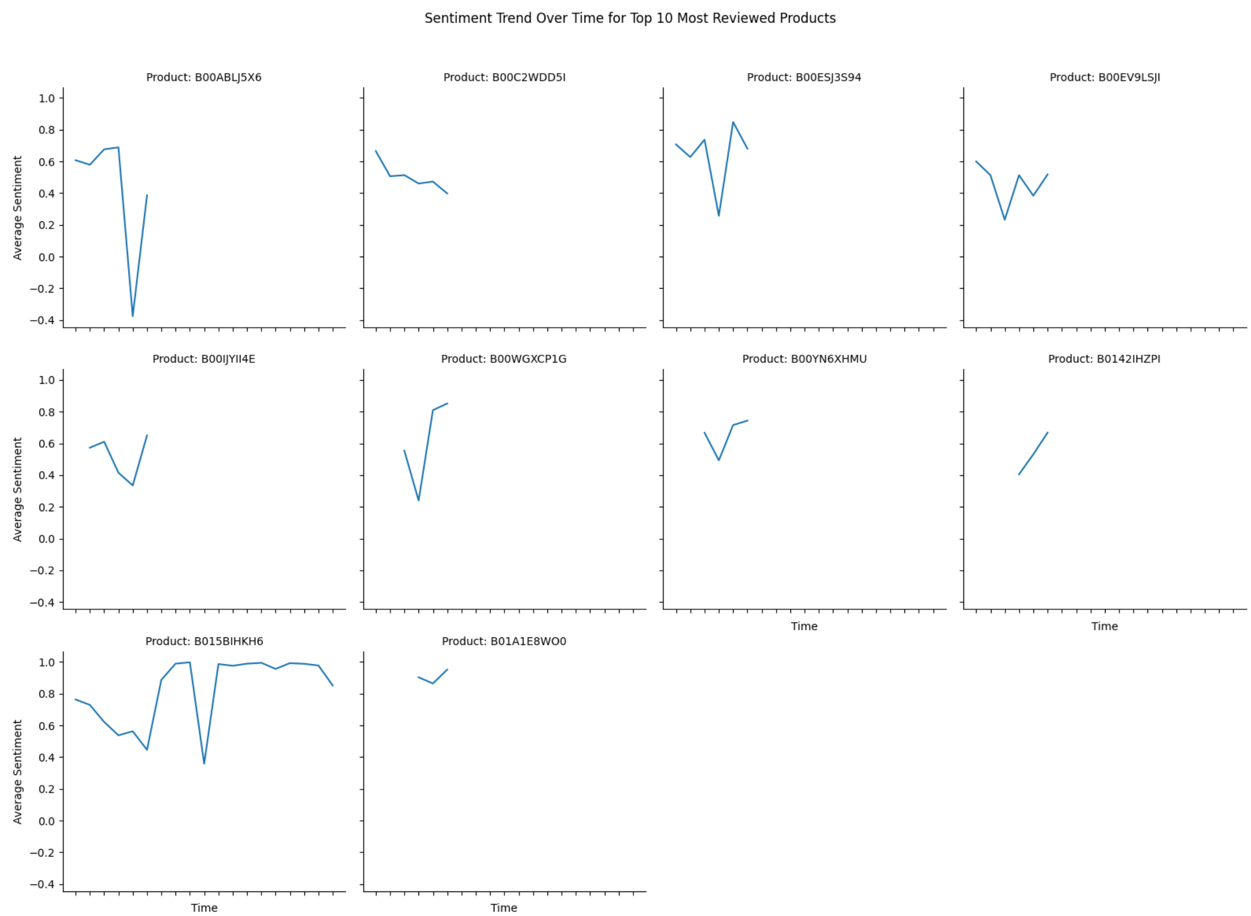

Figure III-4 Sentiment Trend Over Time for Top products.

The collection of graphs portrays the sentiment trends over time for the top 10 most reviewed products can be seen in figure 2.

Each product exhibits a distinct sentiment trend over time, characterized by fluctuations in sentiment levels, including noticeable dips and subsequent recoveries. This variability suggests shifts in customer satisfaction or responses to the product across different time periods. Most of them demonstrate a V-shaped sentiment trend, indicating an initial decline in sentiment followed by a subsequent recovery. This pattern may imply product enhancements or the resolution of negative issues following initial feedback. Certain products, particularly those with ID codes "B01A1E8WOO" and "B0124IHZPI," maintain relatively stable and consistent positive sentiment over time, indicating sustained customer satisfaction with these offerings. Other products, such as those represented by "B00AJ5J594" and "B00YGXCP1G," experience sharp shifts in sentiment, characterized by significant peaks and troughs. These abrupt fluctuations may reflect inconsistencies in product quality, varying customer expectations, or alterations in product versions or updates.

This graph proves the point that the company always continuously takes the reviews seriously and improves the quality of books in the store.

3. Research question 3 – Pairs of words

The study of word frequencies and their pairs in customer reviews is more than just counting words; it's about uncovering the voice of the customer in a structured, quantitative manner that informs strategic decisions. This method bridges the gap between qualitative feedback and quantitative data analysis, providing a robust foundation for improving business practices, enhancing customer experience, and driving innovation. Let's look at the most frequency words in a whole dataset.

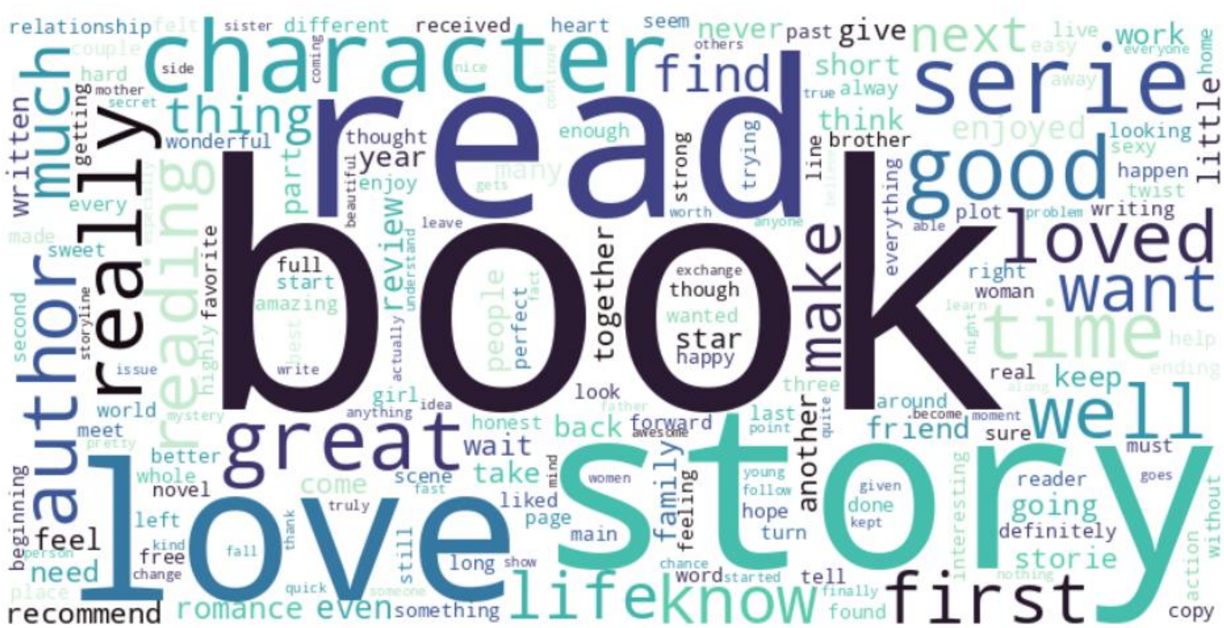

Figure III-5 Word cloud of the most frequent words.

It is evident that the terms 'read,' 'story,' and 'book' dominate the word frequency in this dataset, reflecting its focus on the Kindle store. It can be inferred from this set of words that a significant portion of the reviews pertain to customers' experiences while reading books. Additionally, commonly encountered terms such as "quality," "price," "service," "shipping," and "design" indicate the primary aspects that customers prioritize when evaluating products. For instance, "quality" likely refers to product performance, while "service" may denote customer support, both of which are crucial for ensuring customer satisfaction. Similarly, the presence of "price" and "shipping" suggests that cost and delivery speed are important considerations for customers.

The prevalence of the term "design" underscores its significance in customer satisfaction, possibly relating to the aesthetic or functional aspects of the products. Overall, these frequently occurring words underscore the key factors influencing customer opinions and decision-making processes when interacting with products or services.

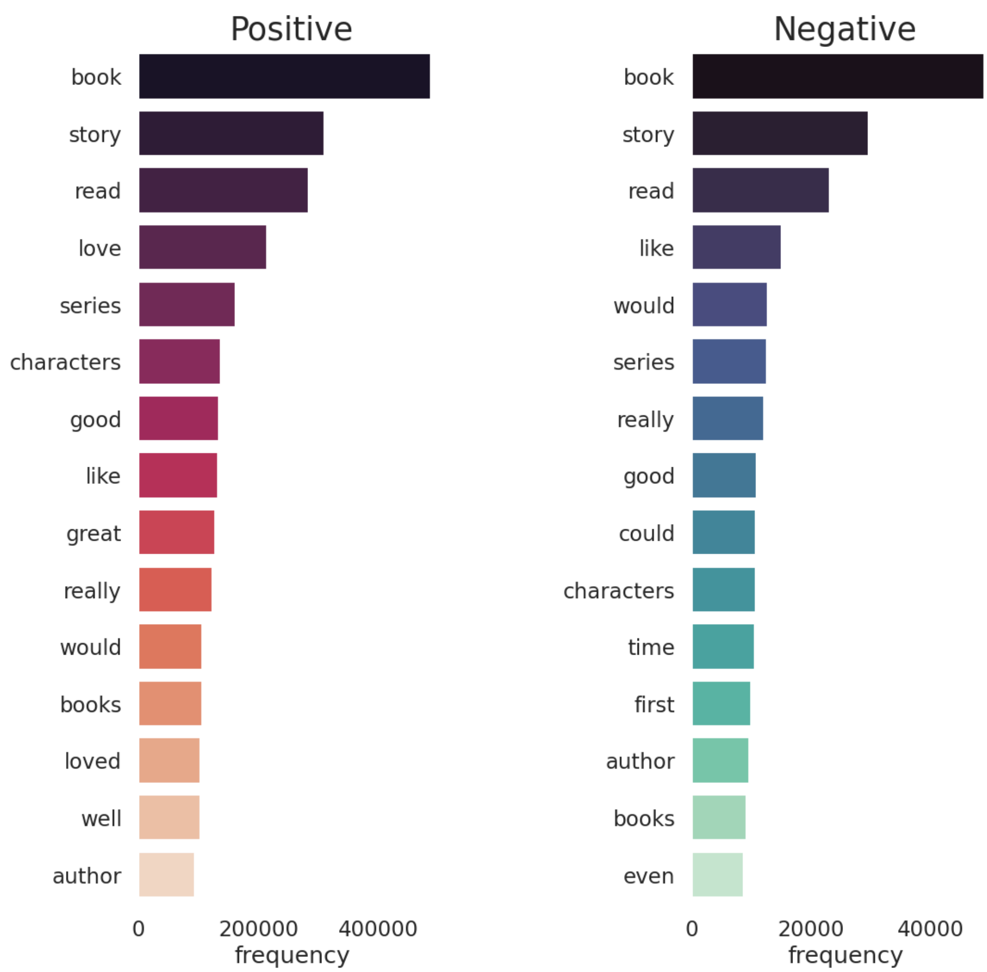

Figure III-6 Top 30 most frequent words for each category.

The provided bar graphs compare the most frequent words found in positive and negative reviews within a dataset, presumably of book reviews. The analysis of these words offers a fascinating insight into what drives positive and negative sentiments among readers. Interestingly, some words appear frequently in both positive and negative reviews ("book," "story," "read," "series," "characters," "good"), indicating key focus areas in reviews that can swing either positively or negatively based on the reader's expectations and experience. The emotional intensity seems higher in positive reviews with words like "love" and "great," whereas negative reviews include more conditional and comparative terms ("could," "would," "even"). To have a more detailed look, the bigram networks is used for analyzing.

Figure III-7 Bigram Networks for words.

Figure III-5 compares the bigram networks of words from positive and negative reviews, revealing distinct patterns and focal points within each sentiment category. The network for positive reviews is characterized by clusters of interconnected nodes where central terms like "highly," "recommend," "absolutely," and "wonderful" stand out. These nodes act as hubs linking to other positive qualifiers and nouns, suggesting a strong emphasis on recommending and praising certain aspects of the books. Terms like "characters" and "loved" are also noticeable, indicating that character development and emotional attachment are significant to positive experiences. The network, although less dense than the negative, shows a clear pattern of reinforcement around enjoyment and recommendation, with several terms branching out into further descriptive positives.

In contrast, the network for negative reviews is denser and more interconnected, with "disappointed," "waste," "poor," and "money" forming central nodes. This suggests a focus on dissatisfaction related to the expenditure (money) and perceived value (waste, poor). Negative character evaluations and expressions like "boring," "no," "plot," and "characters" frequently appear, pointing to dissatisfaction with narrative and development. Terms such as "grammatical," "errors," and "poorly" highlight criticisms targeting the technical aspects of writing, reflecting concerns over the quality of editing and writing.

These bigram networks provide a vivid illustration of how readers articulate their experiences in reviews. Positive reviews tend to celebrate and recommend, using a variety of positive descriptors that connect less densely but cover a broad range of topics. Negative reviews, however, form a dense network of dissatisfaction focused on value and quality, indicating a more unified theme of critique and disappointment. These insights can be incredibly useful for authors and publishers to understand and improve the elements that matter most to their readers.

IV. DATA INSIGHTS AND CONCLUSIONS

The first research question explored the use of machine learning tools to predict the helpfulness of Kindle book reviews. The model achieved a predictive accuracy of approximately 78.07%, demonstrating a decent ability to classify reviews as either helpful or not. It excelled at identifying non-helpful reviews (high precision and recall for class 0) but struggled with correctly identifying helpful reviews, exhibiting low precision and recall for class 1. This indicates that while the model can effectively detect non-helpful reviews, its capacity to recognize truly helpful reviews needs improvement, likely due to issues like class imbalance and the need for more nuanced features to capture helpfulness.

Next to it, the analysis of sentiment trends over time provided insights into how customer perceptions have evolved. The data showed a pronounced shift towards positive sentiments around 2005, which continued to dominate through the end of the dataset. This suggests improvements in product quality or customer service might have influenced customer satisfaction positively. The trend analysis also highlighted the importance of continuous sentiment monitoring as a tool for businesses to adapt to changing customer expectations and maintain a favorable brand perception.

The third question delved into the most frequent words and their pairs in the reviews to uncover underlying consumer sentiments and preferences. Positive reviews frequently included words like "love" and "great," emphasizing strong customer satisfaction. In contrast, negative reviews highlighted terms related to disappointment and technical issues, providing critical feedback that could guide improvements. The bigram analysis further illustrated how specific word pairs could enhance understanding of customer sentiments, offering actionable insights for refining marketing strategies and product offerings.

In summary, the results highlight the significance of ongoing sentiment analysis as a strategic asset for digital market platforms. By promptly identifying negative reviews and assessing their helpfulness, companies can swiftly address issues. However, the study is constrained by uneven data distribution, impacting the efficacy of machine learning. Moving forward, collecting more comprehensive data will enable the effective application of machine learning methods, offering substantial benefits to the company.

V. REFERENCES

1. Zhang, H. (2022). Model Comparison in Sentiment Analysis: A Case Study of Amazon Product Reviews. *Highlights in Science, Engineering and Technology*, 16, 23-31.
2. Stuti, M., Meshram, M., & Sahu, N. (n.d.). Sentiment Analysis of E-Commerce Product Review Through Machine Learning. Retrieved from www.ijfmr.com
3. Ilham, A. A., Bustamin, A., & Wahyudiarto, E. (2023). Customer Satisfaction Assessment System on Transactions E-commerce Product Purchases Using Sentiment Analysis. *13*(3).
4. Katole, H. J. (2022). A research article on the Importance of Online Customer Reviews on Customer Purchase. In *The Business and Management Review* (Vol. 13).
5. Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603-9611. https://doi.org/10.1016/j.eswa.2015.07.052
6. Shazuli, S. A. N., & Head, I. (n.d.). Impact of Online Reviews on Consumer Purchase Decisions in E-Commerce Platforms. Retrieved from www.ijfmr.com
7. Ni, J., Li, J., & McAuley, J. (2019). Amazon Reviews. Justifying recommendations using distantly labeled reviews and fined-grained aspects. *Empirical Methods in Natural Language Processing (EMNLP)*. Available online at: http://deepyeti.ucsd.edu/jianmo/amazon/index.html
8. Yadav, R., Kumar, A. V., & Kumar, A. (2019). News-based supervised sentiment analysis for prediction of futures buying behaviour. *IIMB Management Review*, *31*(2), 157-166. https://doi.org/10.1016/j.iimb.2019.03.006