The impact of readers' sentiments on book sales: an applied study of sentiment analysis of bestselling book reviews on Amazon Egypt

Dr. Ahmed Mohamed Ali Abdel Mokhtar
Teacher of Library and Information Sciences,
Faculty of Arts, Minia University, Egypt
ahmed.mokhtar@mu.edu.eg

Abstract

The study aimed to analyze the sentiments of reviews for the bestselling books on Amazon to explore the positive, negative, and neutral emotions expressed in these reviews and assess their impact on book sales. This was achieved through the use of machine learning and natural language processing techniques available in the Orange Data Mining environment, supported by Python libraries and algorithms such as Vader and LDA. The study followed a descriptive-analytical approach and collected a total of 4,145 reviews using web scraping techniques from Amazon. These reviews pertained to the top 20 bestselling books on Amazon during the period from August 18 to August 25, 2024.

The study yielded several key findings, the most notable being that the book Good Energy ranked first among the top 20 bestselling books on Amazon, followed by Imminent in second place and Atomic Habits in third. The study also demonstrated that bestselling books are not limited to a specific genre but span various fields of knowledge. One of the main conclusions was that readers' emotions towards books are a significant factor influencing purchasing decisions and book sales. The study revealed that positive sentiments dominated the reviews, with approximately 3,603 positive reviews, representing 87% of the total. Most of these reviews were classified under the "Joy" category, accounting for 3,474 reviews or 83.3%, according to Ekman's model.

The study concluded with a set of recommendations and proposed mechanisms directed at authors, publishers, and digital platforms. One key recommendation is to leverage sentiment analysis results in marketing campaigns and develop recommendation algorithms that do not solely rely on star ratings but also incorporate sentiment analysis of reviews. It also suggested displaying books with higher positive sentiment at the top of the browsing lists in specific subject areas. Additionally, the study recommended issuing periodic reports listing books that recorded the highest percentage of positive sentiment in reader reviews, categorized by subject. It also proposed providing a brief graphical summary for each book, showing the evolution of reader sentiments and how they reflect on the book's sales performance.

Keywords

Natural Language Processing (NLP), Sentiment Analysis, Book Sales, Digital Book Platforms, Orange Data Mining, Amazon

1. Introduction

This study builds on prior research by Abdul Mukhtar (2024), which explored the potential of machine learning and natural language processing (NLP) techniques to understand readers' preferences and opinions about information sources. That research resulted in the design of a workflow model utilizing machine learning techniques via Orange Data Mining to analyze readers' reviews of books. The model was tested on reviews of library and information science books on LibraryThing. Key recommendations from the study emphasized the need for libraries, information institutions, and publishers to expand their use of artificial intelligence (AI) and NLP to enhance marketing strategies, refine acquisition policies, and ensure the alignment of information sources with user expectations and needs.

Building on those findings, this study investigates the impact of readers' sentiments on the sales performance of best-selling books on Amazon. It aims to explore the significance of sentiment analysis in understanding reviews and opinions about information sources on digital platforms. This study also highlights the relevance of these technologies for information institutions, publishers, and distributors in the digital age.

2. Study Problem and Questions

Readers' reviews on digital platforms like Amazon significantly influence purchasing decisions. Best-selling books are shaped not only by the reputation of publishers and authors or the quality of content but also by the engagement, ratings, and reviews of readers.

This study aims to analyze the influence of positive and negative reviews on the sales of Amazon's best-selling books. It also seeks to explore how sentiment analysis and text processing can improve marketing strategies, support digital publishing, and drive the commercial success of books.

The study addresses the following questions:

1. What are the numerical and temporal characteristics of Amazon's best-selling books?

2. Are best-selling books on Amazon concentrated in specific subject areas, or are they diverse?

3. What insights can be drawn from sentiment analysis of reviews for Amazon's best-selling books?

4. Which books receive the highest numbers of positive and negative reviews?

5. How do reviews reflect emotional trends using the Profile of Mood States (POMS) and the Ekman model?

6. What insights emerge from topic modeling of reviews for Amazon's best-selling books?

7. What are the most frequent keywords in reviews of Amazon's best-selling books?

8. How can the analysis of readers' sentiments inform strategies for improving book marketing?

3. Importance of the Study

This study is significant for several reasons:

1. It provides a curated dataset of reviews for Amazon's best-selling books, enabling further research on text analysis algorithms and machine learning model development.

2. It examines the relationship between reader sentiment (positive, negative, neutral) and book sales, shedding light on how opinions shape purchasing decisions.

3. It underscores the importance of engaging with readers in digital book marketplaces to strengthen the digital publishing industry.

4. Sentiment analysis results offer actionable insights for publishers and authors to understand readers' needs and preferences, leading to more effective marketing strategies and higher book sales.

4. Study Methodology

4.1 Research Design
The study employs a descriptive analytical approach to assess the influence of readers' sentiments on sales performance for Amazon's best-selling books.

4.2 Study Sample
The study analyzes 4,154 reviews from the top 20 best-selling books on Amazon during the third week of August 2024 (August 18–25).

4.3 Study Tools
The following tools were employed:

1. MS Excel: For storing and preprocessing review data.

2. Orange Data Mining: An open-source software for data mining and machine learning, integrating with Python for analysis, model building, testing, and visualization (Salem et al., 2021).

3. Python Libraries: Including NumPy and NLTK for text processing.

4. VADER Algorithm: For classifying and analyzing text sentiment.

5. Ekman Model: For analyzing emotional trends in reviews.

6. Topic Modeling: Using the Latent Dirichlet Allocation (LDA) algorithm.

5. Study Limitations

Objective Limitations: The focus is on sentiment analysis of reviews for Amazon's best-selling nonfiction books. Fiction books were excluded.

Temporal Limitations: Data was collected from Amazon's weekly best-seller report for the third week of August 2024 (August 18–25).

Qualitative Limitations: The analysis is limited to published reviews of nonfiction books, excluding those for fiction titles.

7. Study Results

1. The study analyzed 4,145 reviews of the top 20 best-selling books on Amazon during the period from August 18 to 25, 2024.

2. Good Energy ranked as the best-selling book, followed by Imminent and Atomic Habits. Shameless was the least-selling book during the same period.

3. Several books, including Atomic Habits, The Body Keeps the Score, Forgiving What You Can't Forget, Spooky Cutie, What Happens Next, and The Devil at His Elbow, achieved a rating of 4.8, indicating high popularity among readers.

4. The best-selling books cover a variety of topics:

   o   Good Energy specializes in nutritional sciences.

   o   Imminent focuses on national and international security.

   o   Atomic Habits addresses organizational behavior, business operations, and infrastructure.
       This diversity reflects readers' interest in a wide range of subjects.

5. Book prices ranged from $7.99 to $79.95. Chaos was the most expensive book ($79.95), while Spooky Cutie was the least expensive ($7.99).

6. The publication dates of the top 20 books span from 2000 (The 48 Laws of Power) to 2024.

7. Forgiving What You Can't Forget and The Body Keeps the Score received the highest number of positive reviews, with six positive reviews each in the initial 20 reviews.

8. The Body Keeps the Score had the highest negative sentiment score, with a compound factor of -0.4767.

9. Positive sentiments dominated reader reviews, totaling 3,603 (87%), compared to 408 negative reviews (9.8%) and 134 neutral reviews (3.2%). This highlights the influence of reader sentiment on book sales.

10. Based on Ekman's model, most reviews were classified as Joy (3,474 reviews, 83.3%). Anger appeared in only 15 reviews (0.4%), and no reviews expressed Disgust. This underscores the prevalence of positive emotions in reviews.

11. The study identified 11,626 keywords in reader reviews. The most frequently used terms were "book" (5,944 occurrences), "read" (2,459), and "habits" (1,501). A word cloud analysis confirmed that the majority of terms reflect positive sentiments.

8. Study Recommendations

8.1 Recommendations for Publishers

- Leverage sentiment analysis results to craft promotional campaigns, emphasizing features that readers value most and incorporating positive review excerpts into marketing materials.

- Encourage readers to share their reviews after purchasing and reading books by offering discounts or incentives for future purchases.

- Periodically monitor shifts in reader sentiment to understand how external factors, such as seasons or events, affect sales.

- Respond to negative reviews to demonstrate a commitment to addressing reader concerns and improving services.

- Use sentiment analysis to identify demographic trends among audiences expressing positive opinions, enabling targeted and effective promotional efforts.

8.2 Recommendations for Authors

- Utilize sentiment analysis to identify areas for improvement and refine future works based on reader feedback.

- Engage with readers by responding positively to their reviews and comments, fostering a stronger author-audience relationship.

- Encourage readers to discuss and share their books on social media and forums to amplify visibility and readership.

- Draw inspiration for new ideas from positive and negative sentiments expressed in reader reviews to guide future creative endeavors.

8.3 Recommendations for Commercial Digital Book Platforms

- Enhance recommendation algorithms by incorporating sentiment analysis alongside star ratings.

- Highlight books with positive sentiments when users browse specific subject areas.

- Provide authors and publishers with dashboards summarizing sentiment analysis of their works.

- Display charts tracking sentiment trends and their impact on book sales for each title.

- Offer summary reports for each book, summarizing the general sentiment trends to assist in quicker purchasing decisions.

- Implement tools to notify authors or publishers of recurring negative reviews, facilitating proactive responses.

- Publish periodic reports showcasing books with the highest positive sentiment percentages, categorized by subject area.

- Develop mechanisms to minimize fake or unreliable reviews, maintaining the credibility of the review system.