

Unveiling Emotional Intensity in Online Reviews: Adopting Advanced Machine Learning Techniques

Sanghyub John Lee¹ and Rouxelle de Villiers²

Australasian Marketing Journal
1–12

© 2024 Australian and New Zealand
Marketing Academy



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14413582241244808

journals.sagepub.com/home/anz



Abstract

The digital revolution has spurred significant growth in online reviews and user-generated content. Traditional methods used in Marketing for analysing large datasets have limitations, emphasising the need for improved analytical approaches, particularly with the advent of artificial intelligence technology. This research used a state-of-the-art transformer model to analyse extensive online book reviews to accurately identify six specific emotions in the reviews of both fiction (hedonic) and nonfiction (utilitarian) genres. This study collected 3,157,703 reviews of 15,293 books voted ‘best book of the year’ on GoodReads.com over the past decade. Our findings reveal noticeable differences in emotional intensity across genres, with nonfiction displaying a slightly higher level of joy, and fiction showing higher levels of anger, sadness and surprise. Joy emerged as the dominant emotion across genres; however, it does not necessarily have a direct impact on book ratings. This study emphasises the intricacies of reader emotions, serving as a significant case study for marketers and publishers aiming to optimise their strategies in the contemporary literary market. The study contributes to the literature on the impact of consumers’ emotional responses, how they are reflected in social review commentary for high-involvement online products, and their impact on product ratings.

Keywords

emotional analysis, online book reviews, fiction and nonfiction genres, hedonic versus utilitarian aspects, e-marketing, social persuasion

Date received: 25 May 2023; accepted: 18 March 2024

Introduction

In the wake of the digital revolution, exemplified by Web 2.0, there has been an exponential rise in user-generated content (UGC) and electronic word-of-mouth (e-WoM). The marketing community responded to this by using e-WoM to gain deeper consumer insights and to influence consumer decisions. A popular tool is sentiment analysis, which automatically scrutinises people’s opinions, emotions and attitudes in written form. Although emotions have traditionally been classified as either positive, negative or neutral, this categorisation appears overly simplistic (Luyckx et al., 2012). As basic human emotions encompass a wide spectrum (Ekman, 1972), such a reductive approach neglects the multifaceted nature of human emotions. This highlights a need for more accurate analytical methodologies (Sailunaz et al., 2018).

Historical challenges to extracting specific emotions from extensive online texts arose from a lack of automated, accurate emotion detection techniques that could operate at scale. Emotion detection in text is challenging because of inherent subjectivity and the absence of non-verbal cues, like facial expressions and tone of voice (Chatterjee et al., 2019). Established methods such as hand coding (Berger & Milkman, 2012; Schindler & Bickart, 2012; Tellis et al., 2019), surveys (Kronrod & Danziger, 2013) and emotion lexicon-based text analysis software (Ludwig et al., 2013; Rocklage & Fazio, 2020; Yin et al., 2017) often falter when applied to ‘big data’ (Davenport, 2014). However, machine learning (ML) and natural language processing (NLP) present more efficient alternatives. Deep learning

models, including neural networks and transformer structures, can refine, extract and elevate the accuracy of vast text datasets (Abdul-Mageed & Ungar, 2017; Mohammad et al., 2018; Saravia et al., 2018). Today’s marketers investigating consumer behaviour regard emotion detection mechanisms as indispensable assets. Therefore, our first research objective (RO1) was:

RO1: Identifying various emotions within large online reviews using advanced machine learning techniques, contrasting emotions reflected in a range of product categories (i.e. utilitarian vs. hedonic categories).

To address RO1, we focussed on online book reviews. Books, as a product focus, are of particular value. They can be classified into two distinct categories with diverse emotional appeal (Chu et al., 2015), and their emotional intensity is likely to shape the success and popularity of both new and old books, in both fiction and nonfiction genres (Maity et al., 2018). Emotional intensity involves identifying and describing variations in emotional expressions, such as differentiating

¹University of Auckland Business School, New Zealand

²Auckland University of Technology, New Zealand

Corresponding author:

Sanghyub John Lee, Marketing Department, University of Auckland Business School, 12 Grafton Road, Auckland 1010, New Zealand.

Email: sanghyub.lee@auckland.ac.nz

between a 10% and a 90% intensity of sadness (Maity et al., 2018). Currently, book sales are declining worldwide. For example, the sales of print books in the United States declined by 6.5% in 2022 relative to 2021 (L. Brown, 2023). In 2021, 75% of American adults engaged with books in various formats: 65% read printed books, 30% used e-books and 23% listened to audiobooks. On the digital front, e-books (with Amazon's Kindle accounting for 72% of the market) witnessed a 3.7% sales increase in 2023, resulting in a revenue of \$85 million (Errera, 2023). As the digital era advances, both print and electronic books face competition from alternative digital commodities (Baron, 2015). According to Schultz (2022), by stimulating short-term dopamine reward, platforms like YouTube and SNS threaten book-based education and entertainment. Therefore, championing the act of reading as a fulfilling and pleasurable pursuit is crucial (Burns et al., 1999). Research, including that by Kesson and Smith (2016) and Maity et al. (2017, 2018), consistently demonstrates that the emotional intensity of books plays a pivotal role in determining their success and popularity.

Comparing fiction (hedonic) and nonfiction (utilitarian) genres goes beyond mere literary analysis; it taps into the essence of consumer psychology and decision-making. By examining the dichotomy between hedonic and utilitarian products, as highlighted by Kivetz and Simonson (2002) and Kronrod and Danziger (2013), it can be seen that motivations for selecting books may be deeply rooted in readers' intrinsic needs – either for escapism and pleasure, or pragmatic information and utility (Jacobs, 2011; Stokmans, 1999). By comparing book reviews from both genres, we can: (1) discern the dominant emotional themes associated with each genre and how emotional responses shape reader preferences; (2) Inform publishing strategies, (e.g. publishers may prioritise books based on their emotional resonance with target readers); (3) Give marketers insights for tailoring campaigns that accentuate the inherent emotional value of a book, thus appealing directly to the reader's hedonic or utilitarian inclinations; and (4) Enhance authorial intent, by guiding authors on how to fine-tune narratives likely to resonate more profoundly with their target audience, ensuring the intended emotional impact. Therefore, our second research objective (RO2) was:

RO2: Evaluate the impact of emotional intensity on sales and rating success and popularity of books, across both fiction (hedonic) and nonfiction (utilitarian) genres.

In the following sections, we first explore emotion analysis in marketing and the fiction (hedonic) and nonfiction (utilitarian) genres. Next, we examine methodology, dataset and emotion detection techniques, followed by a presentation of the results, including emotion analysis, correspondence analysis and the effect of emotions on ratings in subgenres. Finally, we discuss the implications of our findings on marketing strategies, tailoring content and promotional efforts to resonate with the target audience.

Literature review

The literature review first covers the prior studies and state of knowledge regarding emotions in consumer decisions, then investigates consumer emotions when purchasing hedonic versus utilitarian products, followed by a further examination of fiction and non-fiction books and concludes with the literature on using big data collection and machine learning as methods for studies of this nature.

Emotions in marketing and consumer decision-making

Previous research has shown that consumers' emotional responses and product reviews have considerable impact in consumer

decision-making over all stages of the retail and online buying processes (Grant et al., 2013; Mihart, 2012; Penz & Hogg, 2011; Stankevich, 2017). Several marketing studies have investigated emotion analysis in varying contexts. These have ranged from understanding the impact of intense emotions on the perceived usefulness of online reviews (Schindler & Bickart, 2012), to the influence of figurative language on emotional responses to both hedonic and utilitarian offerings (Kronrod & Danziger, 2013). Studies have ranged from the effects of positive and negative wording on sales conversion rates (Ludwig et al., 2013), to the role of emotions such as anxiety and anger in shaping perceptions during purchasing processes (Yin et al., 2017). Other research has addressed the degrees of positivity and emotionality in product choices (Rocklage & Fazio, 2020) and the implications of positive emotions on the shareability of video advertisements (Tellis et al., 2019).

Berger and Milkman (2012) posited that high arousal, for both positive and negative emotions, enhances the chances of online content gaining traction and virality. Chitturi et al. (2007) postulated that consumers' emotional responses vary depending on whether they are choosing hedonic or utilitarian products. According to Chitturi et al. (2007) selecting hedonic products may evoke feelings of guilt or anxiety, whereas opting for utilitarian items might result in feelings of sadness or disappointment. Positive emotions such as elation or exhilaration could be associated with hedonic selections, while sentiments of assurance or confidence could emerge from utilitarian decisions.

Hedonic and utilitarian products

Hedonic and utilitarian products cater to different consumer needs and evoke distinct emotional intensity, which, in turn, influences purchasing decisions and satisfaction (Alba & Williams, 2013). Hedonic products are those that primarily provide pleasure, enjoyment and emotional satisfaction to consumers. These products are often characterised as indulgent, luxurious and experiential, appealing to consumers' desires for pleasure, excitement and entertainment. Examples of hedonic products include designer clothing, fine dining (e.g. Alba & Williams, 2013), movies and fiction books (e.g. Clement et al., 2006). Hedonic consumption is driven by the pursuit of pleasure and gratification, and marketing strategies for these products often focus on creating memorable, emotionally appealing experiences that resonate with consumers.

Utilitarian products, on the other hand, are primarily functional, practical and rational in nature. These products are designed to fulfil specific needs or solve particular problems. Their value is often based on their ability to effectively perform their intended function (Vieira et al., 2022). Utilitarian products include household appliances, documentaries (e.g. Lu et al., 2016) and nonfiction books (e.g. Klauda, 2009). Utilitarian consumption is driven by necessity, efficiency and the pursuit of practical solutions. Marketing strategies for utilitarian products typically focus on highlighting the functionality, reliability and cost-effectiveness of the product to appeal to consumers' rational decision-making processes.

Book genres: Fiction (hedonic) and nonfiction (utilitarian) genres

Fiction and nonfiction book genres cater to unique reader inclinations, with each serving a distinctive purpose. The two genres, reportedly seen as highbrow luxuries or lowbrow necessities (May & Irmak, 2014; Nathanson, 2006; Voss et al., 2003), are not mere literary categorisations but represent differing reader aspirations and experiences.

Fiction, characterised as a hedonic product, uses emotional narratives, providing an imaginative canvas for emotional exploration (Barnes, 2018). Such narratives are defined by storytelling and structure, the ability to connect with characters on an empathetic level and the allure of escapism and imagination. For instance, novels like *The Hunger Games* (Collins, 2008) and *To Kill a Mockingbird* (H. Lee, 1960) are respectively lauded for their emotive storytelling and the deep empathetic connection readers forge with the characters. The allure of immersive worlds, as in Tolkien's (1954) *The Lord of the Rings*, underscores the importance of imagination in literature, allowing readers to transcend their realities, a sentiment echoed by Moran (1994) and Merga (2017). Aldama (2015) and Kim and Klinger (2019) underscore the potency of narrative in eliciting a spectrum of emotions, while studies like those by John (2017) and Dill-Shackleford et al. (2016) emphasise the centrality of empathy in character engagement.

Nonfiction operates as a utilitarian product, giving readers knowledge and insights applicable to their lives (Gerard, 2017). Here, real-world relevance and impact are paramount. Works like Walker's (1982) *The Color Purple* emphasise human resilience and hope, resonating deeply with readers. Nonfiction content is particularly engaging when it offers readers an informative and transformative perspective, a characteristic prevalent in genres such as politics and social sciences or memoirs (Gerard, 2017). However, the emotional breadth in nonfiction may not be as varied as in fiction (S. Brown & Patterson, 2010; Driscoll & Rehberg Sedo, 2019).

In conclusion, the literature review has highlighted the nuances distinguishing fiction (hedonic) from nonfiction (utilitarian) genres. This can help the book industry by amplifying reader engagement and satisfaction. Building on this foundation, this study investigated two primary research questions using machine learning, with further rationales provided in the following section:

Research Question 1 (RQ1): *Do fiction (hedonic) books elicit a broader and more intense spectrum of emotions in readers compared to nonfiction (utilitarian) books, as reflected in their reviews?*

Research Question 2 (RQ2): *Is there a correlation between the emotions expressed in online reviews and the success (measured by sales or ratings) of a book within its respective genre (fiction or nonfiction)?*

Road-signs for machine learning and big data collection

Human emotions can be interpreted differently depending on the context, making comprehensive understanding challenging (Barrett, 2017). Accurate emotion detection is further complicated by subjectivity and the limitations of text sources, which lack contextual information such as facial expressions or tone of voice (Chatterjee et al., 2019). Therefore, it is essential to identify emotional intensity in text using the latest artificial intelligence methodologies that can account for various contexts (S. J. Lee et al., 2021).

The methodologies used in many of the aforementioned emotion-based studies – manual annotation (Berger & Milkman, 2012; Schindler & Bickart, 2012; Tellis et al., 2019), surveying (Kronrod & Danziger, 2013) and commercial text analysis software (Ludwig et al., 2013; Rocklage & Fazio, 2020; Yin et al., 2017) – have inherent limitations. Manual methods are ill-suited for use with large-scale datasets, and traditional text analysis tools, which rely on counting words from sentiment lexicons, may miss the depth and accuracy of the insights due to limited capabilities (S. J. Lee et al., 2021). Consequently, the advanced machine learning (ML)

algorithms available from computer science are being applied to detect emotions.

Despite ML's potential, marketing research has yet to explore emotion detection using ML algorithms in online reviews. Bougie et al.'s (2003) study was an early attempt to explore explicit emotion analysis using traditional survey techniques and human evaluators, rather than leveraging ML. Their findings challenged the oversimplified notion of categorising dissatisfied consumers under a broad 'negative emotion' umbrella. They highlighted the distinction between dissatisfaction, which triggers a drive to understand the cause of service shortcomings, and anger, which might prompt consumers to seek retribution against service providers at fault.

Methodology

Dataset

Our dataset was a comprehensive collection of 3,157,703 reviews, produced by a user base of 1,207,526 members, spanning 15,293 books. These books were chosen from those listed as 'best book of the year' on GoodReads.com. This selection was based on a methodical aggregation of the annual 'best book' votes conducted on Goodreads from 2010 to 2021. During this period, users of the platform voted for their preferred literary pieces, which formed the basis of our dataset.

GoodReads.com presents a comprehensive compendium of book reviews and ratings from a heterogeneous user base. As Maity et al. (2018) stated, 'Goodreads is a community-driven social cataloguing site that has exponentially grown into one of the most favoured social platforms for book reading and recommendations' (p. 118). For over a decade, the Goodreads' Readers Choice feature has allowed users to nominate and vote for their preferred books. The vast amount of reader-created content on Goodreads makes it an ideal resource for book reviews, yielding valuable insights into reader predilections and experiences.

In comparison to bestseller lists from bookstores such as Amazon, using yearly reader-chosen lists of best books offered several advantages for our analysis. Firstly, reader-chosen lists reflect the genuine preferences and opinions of a broader range of readers, whereas bestseller lists may be influenced by factors such as marketing strategies and promotional campaigns. Secondly, reader-chosen lists cover a diverse range of genres and subjects, offering a more comprehensive understanding of the emotional intensities across various types of literature genres. Lastly, by focussing on reader-chosen lists, we could gain a deeper understanding of the emotional intensity that resonates with readers, which is essential for customer-centric tailoring of content and promotional strategies for the marketing of books.

The main genres were fiction and nonfiction. Table 1 presents examples of best books from various genres and subgenres. These books represented a range of stories, themes and styles that have resonated with readers and have been voted as best books within their respective categories. The subgenres consisted of 20 types of fiction (Historical Fiction, Young Adult, Fantasy, Romance, Science Fiction, Thriller, Detective & Mystery, Adventure, Horror, Childrens Fiction, Dystopian, Contemporary Fiction, Drama, Short Stories, Poetry, Paranormal Romance (romance with supernatural elements such as vampires), LGBTQ+, Literary Fiction, Urban Fantasy, Mystery & Detective) and 10 categories of non-fiction (Memoir & Autobiography, Self-Help, History, Humour & Entertainment, Politics & Social Sciences, Religion & Spirituality, Biography, Philosophy, True Crime, Business & Money).

Table 1. Examples of Best Books by Genres.

Main genre	Sub-genre	Best books
Fiction	Historical fiction	The Book Thief, Where the Crawdads Sing, To Kill a Mockingbird
	Young adult	The Fault in Our Stars, Twilight, Catching Fire
	Fantasy	Harry Potter and the Sorcerer's Stone, The Night Circus, The Midnight Library
	Romance	Me Before You, Fifty Shades of Grey, Normal People
	Science fiction	Mockingjay, Ready Player One, The Martian
	Thriller	Gone Girl, The Girl on the Train, La paciente silenciosa
	Detective and mystery	The Curious Incident of the Dog in the Night-Time, The Guest List, And Then There Were None
	Adventure	The Alchemist, Life of Pi, Lord of the Flies
	Horror	Frankenstein, Dracula, The Shining
	Children's fiction	The Little Prince, Anne of Green Gables, The Secret Garden
	LGBTQ+	A Little Life, The Picture of Dorian Gray, Simon vs. the Homo Sapiens Agenda
	Dystopian	The Hunger Games, Divergent, 1984
	Contemporary fiction	A Man Called Ove, Eleanor Oliphant Is Completely Fine, The Casual Vacancy
	Drama	The Art of Racing in the Rain, My Sister's Keeper, Handle with Care
	Short stories	Olive Kitteridge, Breakfast at Tiffany's and Three Stories, Interpreter of Maladies
	Poetry	The Sun and Her Flowers, The Princess Saves Herself in This One, The Divine Comedy
	Paranormal romance	Dead Ever After, First Grave on the Right, Halfway to the Grave
	Literary fiction	An American Marriage, The Elegance of the Hedgehog, The Sense of an Ending
	Urban fantasy	Moon Called, Dead Witch Walking, Bitten
	Mystery and detective	Big Little Lies, The Lovely Bones, The Secret History
Nonfiction	Memoir and autobiography	Educated, Becoming, and Eat, Pray, Love: One Woman's Search for Everything Across Italy, India and Indonesia
	Self-help	The Subtle Art of Not Giving a F*ck: A Counterintuitive Approach to Living a Good Life, The Life-Changing Magic of Tidying Up: The Japanese Art of Decluttering and Organizing, and Quiet: The Power of Introverts in a World That Can't Stop Talking
	History	Sapiens: A Brief History of Humankind, The Devil in the White City, The Library Book
	Humour and entertainment	Where'd You Go, Bernadette, Anxious People, and The Hundred-Year-Old Man Who Climbed Out of the Window, Disappeared
	Politics and social sciences	White Fragility: Why It's So Hard for White People to Talk About Racism, The Art of War, The Prince
	Religion and spirituality	Mere Christianity, The God Delusion, and The Holy Bible: King James Version
	Biography	Unbroken: A World War II Story of Survival, Resilience and Redemption, The Immortal Life of Henrietta Lacks, Into the Wild
	Philosophy	The Prophet, Zen and the Art of Motorcycle Maintenance: An Inquiry Into Values, Meditations
	True crime	The Devil in the White City: Murder, Magic, and Madness at the Fair That Changed America, Just Mercy: A Story of Justice and Redemption, In Cold Blood
	Business and money	The Tipping Point: How Little Things Can Make a Big Difference, The 4-Hour Workweek, Start with Why: How Great Leaders Inspire Everyone to Take Action

Table 2 presents the descriptive statistics of the dataset, providing empirical evidence for the varying characteristics of book reviews across different genres and sub-genres. The results reveal clear variations in the book review characteristics depending on the genre. We speculate in this section on possible reasons why this may be so.

In fiction, the findings revealed that: (1) Historical Fiction and Young Adult had the highest number of reviews, indicating their popularity among readers, while sub-genres like Urban Fantasy and Mystery & Detective had fewer reviews, suggesting they are more niche; (2) Paranormal Romance and Children's Fiction received the highest average ratings, possibly due to their engaging and satisfying reading experiences, while LGBTQ+ and Contemporary Fiction had lower ratings, potentially due to their complex and divisive themes; and (3) LGBTQ+ and Drama showed the highest average likes for reviews, potentially attributed to their complex themes, resulting in uncertainty or ambiguity, which therefore encouraged in-depth discussions. Increased uncertainty about a book's content amplifies the significance of book reviews in directing reader choices. Conversely, Young Adult and Contemporary Fiction displayed the lowest average

likes likely because their broad appeal and familiar themes may not elicit as strong or distinctive emotional responses from readers compared to more niche genres.

In nonfiction, the findings showed that: (1) Memoir & Autobiography had the highest number of reviews, indicating readers' interest in personal stories and willingness to engage in providing reviews, while sub-genres like Business & Money and True Crime, catering to specific audiences, were less popular; (2) Religion & Spirituality and Philosophy had the highest average ratings, likely because of their thought-provoking content, while Humour & Entertainment had the lowest average ratings, possibly due to its subjective nature; and (3) Philosophy and Politics & Social Sciences had the highest average likes for reviews, possibly because of their thought-provoking topics that require comprehensive reviews and uncertainty, while Business & Money and Self-Help had the lowest average likes, perhaps due to their practical nature and readers' clearer expectations. Our statistics (as shown in Table 2) indicated that the characteristics of book reviews, such as average rating, likes of reviews and number of reviews, varied significantly depending on the genre.

Table 2. Descriptive Statistics of Online Book Reviews Across Genres.

Main genre	Sub genre	Avg. rating	Avg. pages	N of books	Likes of reviews	N of reviews
Fiction	Historical Fiction	3.90	408	2,446	0.91	533,244
	Young adult	3.93	341	1,966	0.44	495,044
	Fantasy	4.04	441	1,672	0.62	403,831
	Romance	3.87	329	1,491	0.70	310,928
	Science fiction	3.90	385	1,050	0.74	229,259
	Thriller	3.86	372	982	1.03	215,344
	Detective and mystery	3.93	333	512	0.68	117,683
	Adventure	3.93	340	471	0.53	105,084
	Horror	3.84	339	465	0.92	97,799
	Children's fiction	4.09	176	305	0.47	69,891
	LGBTQ+	3.77	263	206	1.29	37,834
	Dystopian	3.92	365	104	0.76	25,351
	Contemporary fiction	3.72	356	82	0.46	22,581
	Drama	3.86	193	75	0.93	18,624
	Short stories	3.84	246	77	1.26	17,245
	Poetry	4.10	263	98	1.46	16,873
	Paranormal romance	4.13	385	59	0.70	15,688
	Literary fiction	3.79	279	50	1.13	13,560
	Urban fantasy	3.98	338	37	0.54	9,808
	Mystery and detective	3.83	338	46	1.22	9,208
	Total	3.92	365	12,194	0.73	2,764,879
Nonfiction	Memoir and autobiography	3.89	293	741	0.60	153,169
	Self-help	3.96	250	247	0.45	43,014
	History	3.90	410	226	0.59	41,334
	Humour and entertainment	3.62	252	193	0.59	38,247
	Politics and social sciences	3.82	299	215	0.92	28,570
	Religion and spirituality	4.08	372	235	0.85	27,279
	Biography	3.91	349	158	0.51	25,037
	Philosophy	4.03	360	68	1.24	15,723
	True crime	3.92	395	58	0.54	11,627
	Business and money	3.87	248	60	0.34	8,824
	Total	3.89	313	2,201	0.63	392,824
	Total	3.92	357	14,395	0.71	3,157,703
Total	Total	3.92	357	14,395	0.71	3,157,703

Emotion detection

The Transformer Transfer Learning (TTL) method (S. J. Lee et al., 2023) for emotion detection is a novel approach developed using Transformer models (Devlin et al., 2018; Liu et al., 2019), which have proven to be highly effective in NLP tasks. A recent paper by S. J. Lee et al. (2021) highlights the use of transformer-based models such as BERT and RoBERTa, which significantly outperform traditional ML algorithms. Traditional methods, including sentiment lexicon or earlier ML models like RNN and CNN, demonstrate lower classification accuracy, as they are less capable of capturing the complex spectrum of human emotions from textual data. For example, S. J. Lee et al. (2021) found that fear was the most dominant emotion in COVID-19-related tweets, a finding that contrasts with previous studies, which concluded that positive emotions like trust and happiness were more prevalent. This discrepancy highlights the limitations of traditional methods in accurately identifying and classifying emotions in textual data.

The TTL method can identify the six emotions – anger, disgust, fear, joy, sadness and surprise – outlined in Ekman's (1972) basic human emotion theory. TTL was devised to address the limitations of existing emotion detection approaches that rely on either small, human-annotated datasets or large, self-reported emotion datasets. While human-annotated datasets may be subject to biases from the

annotators, self-reported emotion datasets may not capture the subtleties of social emotions that humans can easily discern. The TTL method is designed to train emotion detection models in a manner that mimics human developmental stages. It consists of two main steps: (1) detecting emotions reported by the authors in the text and (2) synchronising the model with social emotions identified in annotator-rated emotion datasets. By using this two-step approach, the TTL method seeks to improve the performance of emotion detection models in various contexts.

We adopted the TTL method of S. J. Lee et al. (2023) for our analysis of the specific emotions expressed in the Best Book Reviews because of its ability to capture a more accurate and nuanced representation of emotions. This model uses a two-step learning method, and was initially trained on over 3.6 million instances of four self-reported emotion datasets. The first stage captures a broad spectrum of emotions as directly expressed by individuals. The model was trained again with over 60,000 instances of seven annotator-rated emotion datasets, synchronising them with socially agreed emotions. The TTL model achieved an overall classification accuracy of 84% across the 11 datasets (S. J. Lee et al., 2023).

The TTL method provides valuable insights into the emotional intensity of literary genres. The TTL model weighs emotion within each response through probabilistic scoring. As an example, for a reader's review that stated, 'This book opened my eyes to how

humans make decisions, and how easily they can be influenced by their peers and by the way choices are presented to them'. TTL identified surprise as the dominant emotion with a 0.72 score, followed by joy at 0.24, and sadness at 0.03. Recognising that human emotions often comprise a blend of various feelings (S. J. Lee et al., 2021), this study also examines emotion distribution as a mixed-emotion outcome. Both the primary emotion (e.g. surprise) and the mixed-emotion results (e.g. surprise=0.72, joy=0.24 and sadness=0.03) are factored in. The specific emotions evaluated are documented as separate variables with sequential decimal values ranging from 0 (0%) to 1 (100%), with the sum of these values equating to 1 (100%).

Key variables. In this section, we clearly define the key variables in our study, to enable readers to understand their importance and connect them to the analyses and concrete conclusions drawn. To illustrate, we use the book reviews for *Big Fish*, a novel focussed on the stories a father told his son. This book involves the son remembering these stories as the father is dying.

The heart of our data-driven approach, the collected dataset variables, unfold layers of user interactions across genres and sub-genres. An in-depth classification scheme, these variables help to assign books to distinctive thematic, stylistic or content-based categories. As an instance, *Big Fish* aligns with the Fiction genre and more specifically, the Fantasy sub-genre. Quantification of user engagement is reflected through the total count of reviews associated with specific genres or sub-genres. Here, *Big Fish* elicited 1917 reviews. Ratings of books offer a snapshot of users' perceptions, scaled typically between 1 and 5 stars. In the context of our example, the review for *Big Fish* was accorded a 3.0-star rating. Average ratings provide a holistic view of the general sentiment. An average rating of 3.67 provides the collective appraisal of *Big Fish*. As a testament to the impact and acceptance of a review, the 'likes' serve as a key pointer. In our exemplar, the review garnered 2 likes.

Delving deeper, our exploration used *Python* coding, specifically the TTL method, to detect the emotions of the review text, obtaining the following: (1) Dominant Emotion: Serving as a mirror to the most resonant emotion within a review, this variable exposes the prime emotion invoked by the review text. The *Big Fish* review prominently evoked 'sadness'. (2) Mixed-Emotion: (For *Big Fish* the scores were Anger=0.004, Disgust=0.000, Fear=0.031, Joy=0.369, Sadness=0.416 and Surprise=0.198). A more granular perspective arises through this variable, mapping out the emotional spectrum in the review. Probabilistic weights are accorded to each fundamental emotion, summing up to a total of 1.

Results

Emotion and emotional intensity analyses

This study investigated the differences in emotion scores between fiction and nonfiction book reviews on GoodReads.com using an independent samples t-test, a statistical method employed to compare the means of two independent groups to ascertain the presence of significant differences (Ross & Willson, 2017). The dataset comprised 3,342,842 best book reviews, with 2,764,879 classified as fiction and 392,824 as nonfiction reviews. The emotions analysed were joy, anger, disgust, fear, sadness and surprise.

Range of emotions: Fiction versus non-fiction. The t-test results indicated significant differences in the emotion scores of anger, joy, sadness and surprise between fiction and nonfiction book reviews. The findings are as follows. (1) Joy scores were significantly higher in nonfiction reviews ($M=0.664$, $SD=0.393$) than

fiction reviews ($M=0.623$, $SD=0.401$), with a moderate effect size (Cohen's $d=0.400$). This outcome could be attributed to the nature of many nonfiction genres, such as Self-Help, Religion & Spirituality and Business & Money, which aim to offer practical guidance, inspiration and resources for personal and professional development, typically provoking positive emotions. However, it's important to consider that fiction often stimulates a broader range of emotions, including negative ones, potentially affecting its joy score. Fictional narratives often explore intricate character relationships, moral dilemmas and dramatic events, which can elicit an array of emotions in readers. (2) Fiction reviews showed higher levels of anger ($M=0.064$, $SD=0.186$) compared to nonfiction reviews ($M=0.063$, $SD=0.187$), with a small effect size (Cohen's $d=0.186$). This could suggest that the factors that trigger anger in readers are relatively consistent across both genres, possibly related to themes, characters or the quality of writing. (3) Sadness scores were significantly higher in fiction reviews ($M=0.157$, $SD=0.282$) compared to nonfiction reviews ($M=0.129$, $SD=0.263$), with a small effect size (Cohen's $d=0.279$). Fiction often explores a broader range of emotions and themes, including drama, tragedy and loss, which can evoke sadness in readers. (4) Surprise scores were also significantly higher in fiction reviews ($M=0.088$, $SD=0.188$) than nonfiction reviews ($M=0.075$, $SD=0.178$), with a small effect size (Cohen's $d=0.186$). Fiction often involves unexpected twists, turns and imaginative elements, while nonfiction tends to focus on real-life events, facts and knowledge. (5) No significant differences were found in the emotion scores of disgust and fear between fiction and nonfiction book reviews. This indicates that these emotions are similarly expressed across both main genre of literature.

Emotion intensity: Fiction and non-fiction. Table 3 reports the average values of the six emotions for each genre and sub-genre, showing a clear variation in the intensity of emotions depending on the genre. In fiction, the genres with the highest joy scores are Children's Fiction (0.792), Adventure (0.674) and Fantasy (0.663), while Literary Fiction (0.507), Drama (0.530) and Horror (0.522) have lower scores. This may be attributed to the nature of the content and the target audience. Children's Fiction is designed to entertain and educate children, often evoking positive emotions. Adventure and Fantasy books typically provide an escape for readers, immersing them in imaginative worlds and exciting experiences. Conversely, Literary Fiction, Drama and Horror often explore darker themes, leading to higher scores in emotions such as fear and sadness. Horror, for instance, has the highest fear score (0.171) due to its focus on eliciting fear and suspense in the reader. In nonfiction, the highest joy scores are found in Self-Help (0.750), Religion & Spirituality (0.729) and Business & Money (0.743), while True Crime (0.538), Politics & Social Sciences (0.573) and Philosophy (0.606) show lower joy scores. This can be linked to the purpose and content of these genres. Self-Help, Religion & Spirituality and Business & Money books are often designed to provide practical guidance, inspiration and tools for personal and professional growth, leading to positive emotions. In contrast, True Crime, Politics & Social Sciences and Philosophy often deal with complex, controversial and challenging topics that may evoke negative emotions such as anger, fear and sadness.

In summary, the emotional intensity in book reviews varies between fiction and nonfiction genres, with Fiction exhibiting higher levels of anger, sadness and surprise, while nonfiction has a slightly higher average joy score. These differences may be attributed to the content, themes and the intended audience of the books within each

Table 3. Emotion Analysis of Online Book Reviews by Genre.

Main genre	Sub genre	Anger	Disgust	Fear	Sadness	Surprise	Joy
Fiction	Historical fiction	0.059	0.008	0.057	0.184	0.081	0.611
	Young adult	0.065	0.006	0.047	0.164	0.089	0.629
	Fantasy	0.061	0.006	0.043	0.141	0.085	0.663
	Romance	0.086	0.010	0.041	0.161	0.080	0.622
	Science fiction	0.059	0.008	0.063	0.137	0.105	0.627
	Thriller	0.070	0.012	0.111	0.143	0.110	0.555
	Detective and mystery	0.055	0.008	0.073	0.131	0.108	0.625
	Adventure	0.055	0.007	0.050	0.130	0.085	0.674
	Horror	0.062	0.015	0.171	0.142	0.089	0.522
	Children's fiction	0.031	0.005	0.026	0.090	0.058	0.792
	LGBTQ+	0.077	0.014	0.054	0.179	0.084	0.591
	Dystopian	0.074	0.010	0.089	0.167	0.098	0.561
	Contemporary fiction	0.061	0.012	0.041	0.224	0.080	0.582
	Drama	0.101	0.012	0.056	0.208	0.094	0.530
	Short stories	0.056	0.013	0.057	0.201	0.088	0.586
	Poetry	0.062	0.006	0.046	0.177	0.078	0.631
	Paranormal romance	0.078	0.006	0.044	0.150	0.084	0.638
	Literary fiction	0.078	0.015	0.057	0.243	0.101	0.507
	Urban fantasy	0.083	0.007	0.051	0.142	0.099	0.619
	Mystery and detective	0.058	0.008	0.070	0.188	0.112	0.564
	Total	0.064	0.008	0.060	0.157	0.089	0.623
Nonfiction	Memoir and autobiography	0.052	0.008	0.058	0.161	0.066	0.656
	Self-help	0.062	0.005	0.045	0.077	0.061	0.750
	History	0.055	0.008	0.078	0.129	0.095	0.635
	Humour and entertainment	0.068	0.012	0.026	0.130	0.063	0.701
	Politics and social sciences	0.120	0.012	0.106	0.097	0.093	0.573
	Religion and spirituality	0.066	0.005	0.045	0.073	0.082	0.729
	Biography	0.054	0.007	0.050	0.140	0.083	0.666
	Philosophy	0.096	0.007	0.055	0.126	0.109	0.606
	True crime	0.072	0.016	0.124	0.151	0.099	0.538
	Business and money	0.063	0.006	0.051	0.060	0.077	0.743
	Total	0.063	0.008	0.059	0.129	0.075	0.664
	Total	0.064	0.008	0.060	0.153	0.087	0.628
	Total	0.064	0.008	0.060	0.153	0.087	0.628

genre. Responses can vary significantly, with some genres evoking strong emotions such as anger, disgust or fear, while others evoke somewhat weaker positive emotions, such as joy or surprise.

Correspondence analysis

To explore associations between two categorical variables (Greenacre, 2017), we performed a Correspondence Analysis (CA) of the relationship between the six emotions and book genres (divided into 30 sub-genres). This allowed for an in-depth exploration of how literary genres correlate with their emotional content.

The largest proportion of the inertia, which reflects the variance explained by each dimension (Clausen, 1998), was accounted for by the first dimension (D1: anger, disgust and fear, 63.6%): the emotion of fear had a high positive score (1.426). The second dimension (D2) contributed to 21.8% of the inertia, in that the emotion of sadness had a high negative score (−0.617), while the emotion of joy had a positive score (0.148). In total, the first two dimensions captured 85.4% of the total inertia, indicating a strong association between genres and emotions. The most frequently experienced emotion across all genres was joy ($n=2,092,638$), followed by sadness ($n=479,037$) and anger ($n=188,649$). The row profiles showed that joy was the most dominant emotion in most genres, with the highest proportion in children's fiction (82.0%) and the lowest in

thriller fiction (58.8%). The column profiles indicated that historical fiction contributed most to the anger (15.2%) and disgust (16.9%) emotions, while the young adult genre contributed the most to fear (12.0%) and sadness (16.7%) emotions.

Figure 1 shows the genres positioning map according to the emotions. Among the genres, horror was highly correlated with the first dimension (D1: score: 1.408), followed by thrillers (score: 0.712). This suggests that these two genres were most closely associated with emotions such as anger, disgust and fear. Children's fiction (score: −0.603) and self-help (score: −0.381) were negatively correlated with the first dimension (D1: anger disgust and fear), indicating that these genres were more closely associated with emotions such as joy, surprise and sadness. In the second dimension, children's fiction (score: 0.812) and religion & spirituality (score: 0.783) had the highest positive scores, whereas contemporary fiction (score: −0.723) and LGBTQ+ (score: −0.335) had the highest negative scores. This suggests that the emotions experienced in these genres were different from those in the other genres. In conclusion, the CA revealed a strong association between genres and emotions in literature. The analysis highlighted the dominance of joy as the most frequently experienced emotion across most genres and identified genres that were more closely associated with specific emotions.

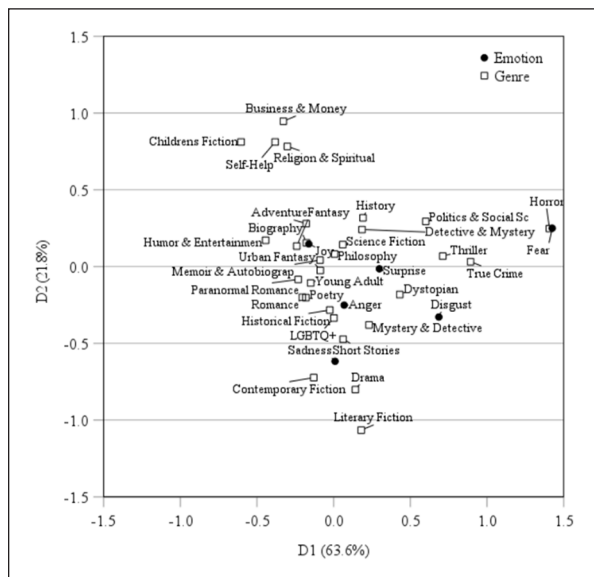


Figure 1. Genres positioning map according to reviewers' emotions.

The effect of emotions on ratings in subgenres

A stepwise multiple regression analysis was performed to assess the influence of emotions identified in the best book reviews on the average rating of the best books within each subgenre. The dependent variable in this analysis is the average rating of the best books in each subgenre, while the independent variables are the six emotions (i.e. anger, disgust, fear, sadness, surprise and joy) detected in the reviews for each best book. The dataset comprised 14,395 books across 30 subgenres, and the number of books (i.e. the number of independent variables) in each subgenre is presented in Table 2.

The R^2 values exhibited varying degrees of model fitness across different book genres. The R^2 values varied across genres, with some genres demonstrating a stronger relationship between emotions and book ratings, such as business & money ($R^2=.75$) and religion & spirituality ($R^2=.61$). Other genres showed weaker relationships, such as urban fantasy ($R^2=.13$) and literary fiction ($R^2=.26$). These varying R^2 values indicate the diverse influence of emotions on book ratings across genres, reflecting the complexity of reader preferences and emotional engagement in different types of books. The Durbin-Watson (D-W) statistics, which assess the presence of autocorrelation in the residuals, ranged between 1.57 and 2.05 across different genres, with most values close to 2. This range indicates that our models generally exhibited minimal autocorrelation, which supports the validity of the regression results. The stepwise regression procedure entered only those emotion variables into the model that met the criterion of a probability-of-F-to-enter $\leq .050$. Overall, the model fitness in our study demonstrates that emotions in book reviews play a significant role in predicting book ratings across various genres.

In the fiction genre, joy had a consistent positive impact on book ratings, with the exception of the Dystopian and Paranormal Romance subgenres. Disgust, on the other hand, appeared to have a significant negative impact on ratings across most fiction subgenres. Specifically, historical fiction had a significant relationship between disgust ($\beta=.64$), fear ($\beta=.11$) and joy ($\beta=1.20$) and book ratings, with an R^2 of .39. This indicates that readers appreciate the presence of these emotions in historical fiction books. Young adult fiction displayed a negative relationship with anger ($\beta=-.29$), disgust ($\beta=-.35$), sadness ($\beta=-.25$) and a weak positive relationship with

joy ($\beta=.05$), with an R^2 of .45. This suggests that readers of this genre prefer fewer negative emotions and may appreciate a moderate presence of joy. Fantasy books exhibited a negative relationship with anger ($\beta=-.39$) and positive relationships with disgust ($\beta=.28$), fear ($\beta=.12$) and joy ($\beta=.59$), with an R^2 of .46. Romance books showed a negative relationship with anger ($\beta=-.14$), disgust ($\beta=-.24$) and a positive relationship with fear ($\beta=.15$) and joy ($\beta=.38$), with an R^2 of .32. This indicates that readers of fantasy and romance books enjoy a mix of emotions, with a preference for lower levels of anger and disgust.

In nonfiction, disgust also emerged as a consistent negative predictor of book ratings. However, unlike in fiction genres, joy was not as consistently linked to higher ratings in nonfiction subgenres. Instead, other emotions, such as sadness and surprise, had significant effects on book ratings depending on the subgenre. Specifically, memoir and autobiography books displayed a negative relationship with anger ($\beta=-.32$), disgust ($\beta=-.38$) and surprise ($\beta=-.27$), and a positive relationship with fear ($\beta=.21$), with an R^2 of .52. This suggests that readers of this genre appreciate a balance of emotions, with a preference for fear and lower levels of anger, disgust and surprise. Self-help books had a negative relationship with anger ($\beta=-.56$), disgust ($\beta=-.17$) and sadness ($\beta=-.23$), with an R^2 of .51. This implies that readers prefer self-help books with less negative emotions. History books demonstrated a negative relationship with anger ($\beta=-.48$), disgust ($\beta=-.30$), surprise ($\beta=-.18$) and a positive relationship with fear ($\beta=.28$), with an R^2 of .45. This indicates that readers appreciate historical books that evoke fear while minimising other negative emotions (Table 4).

Interpretation of results

In an age of dwindling book sales and the rise of diverse media, understanding the emotional impact of books and its link to books' popularity and success is an important insight for those aiming to promote reading. This discussion focusses on the pivotal findings related to the emotional experiences of readers in fiction (hedonic) and nonfiction (utilitarian) genres. We break down these experiences into five segments: Storytelling and Narrative Structure, Empathy and Connection with Characters, Imagination and Escapism, Real-world Relevance and Impact of Emotional Depth over all genres.

Storytelling and structure. Firstly, our research indicates that fiction books generate stronger emotional responses (see Table 3). This is probably due to their intricate storytelling and narrative structures, particularly in genres such as historical fiction, fantasy and romance. Conversely, nonfiction genres, including history and philosophy, tend to be more informational, leading to subdued emotional reactions. Thus, for marketing purposes, it is beneficial to emphasise the narrative facets that align with a specific genre's inherent emotions.

Empathy and connection with characters. Character development plays a vital role in emotional engagement (Eekhof et al., 2023; Keen, 2006). Genres with pronounced character development, including young adult fiction and detective stories, result in profound emotional experiences as readers establish empathy with the characters. Emphasising character-driven narratives can thus be a potent marketing tool.

Imagination and escapism. Fiction genres offering a rich tapestry of imagination and escapism, like fantasy and adventure, were associated with heightened joy levels. Such genres offer readers a sanctuary from reality, a feature that should be highlighted in marketing campaigns (Klimmt, 2008; Merga, 2017). In contrast, nonfiction genres,

Table 4. The Effect of Emotions of Book Reviews on the Rating of Books in Each Subgenre.

Genres	F	Sig.	R ²	D-W	Anger (β)	Disgust (β)	Fear (β)	Sadness (β)	Surprise (β)	Joy (β)
Fiction										
Historical fiction	383.36	***	.39	1.89	ns	ns	.48***	.64***	.11***	1.20***
Young adult	391.96	***	.45	1.88	-.29***	-.35***	ns	ns	-.25***	.05*
Fantasy	348.46	***	.46	1.92	ns	-.39***	.28***	.12***	ns	.59***
Romance	167.77	***	.32	1.82	-.14***	-.24***	.15***	ns	ns	.38***
Science fiction	169.17	***	.40	1.82	-.13***	-.15***	.23***	ns	ns	.52***
Thriller	128.21	***	.35	1.74	-.12***	-.07*	.31***	ns	ns	.50***
Detective and mystery	83.93	***	.40	1.85	ns	-.23***	.41***	-.15*	ns	.42***
Adventure	87.82	***	.44	1.91	-.17***	-.29***	.17***	ns	ns	.40***
Horror	91.72	***	.38	1.71	ns	-.09*	.45***	ns	ns	.68***
Children's fiction	35.23	***	.32	1.86	-.19***	ns	.11*	ns	-.18**	.40***
LGBTQ+	54.27	***	.45	1.57	-.46***	-.37***	ns	ns	-.15**	ns
Dystopian	52.87	***	.62	1.69	-.53***	-.39***	ns	-.22***	ns	ns
Contemporary fiction	35.09	***	.47	1.86	-.45***	ns	ns	ns	ns	.32**
Drama	20.48	***	.47	2.03	ns	-.49***	.35**	ns	ns	.52***
Short stories	28.73	***	.29	1.82	ns	ns	ns	ns	ns	.54***
Poetry	37.60	***	.44	1.91	ns	-.51***	ns	ns	ns	.25**
Paranormal romance	35.66	***	.66	1.69	-.35***	-.68***	ns	ns	-.29***	ns
Literary fiction	2.90	ns	.26	2.05	ns	ns	ns	ns	ns	ns
Urban fantasy	0.89	ns	.13	1.64	ns	ns	ns	ns	ns	ns
Mystery and detective	16.13	***	.55	1.60	ns	-.37**	.53***	ns	ns	.56***
Nonfiction										
Memoir and autobiography	197.32	***	.52	1.97	-.32***	-.38***	.21***	ns	-.27***	ns
Self-help	80.10	***	.51	1.54	-.56***	-.17***	ns	-.23***	ns	ns
History	42.57	***	.45	1.65	-.48***	-.30***	.28***	ns	-.18***	ns
Humour and entertainment	20.25	***	.31	1.82	-.30***	-.40***	ns	-.26***	.22**	ns
Politics and social sciences	41.81	***	.29	1.70	-.48***	ns	ns	-.29***	ns	ns
Religion and spirituality	85.35	***	.61	1.81	-.19**	-.20***	.23***	ns	ns	.62***
Biography	32.71	***	.47	1.70	-.20**	-.24***	ns	-.23***	-.52***	ns
Philosophy	22.83	***	.41	1.73	-.34**	-.40***	ns	ns	ns	ns
True crime	13.66	***	.43	1.98	-.43***	-.29*	ns	ns	-.23*	ns
Business and money	51.80	***	.75	1.89	-.41***	-.37***	ns	ns	-.36***	ns

Note. F = The F-statistic, Sig. = Significance, R² = The coefficient of determination, D-W = Durbin-Watson statistic, and β = The coefficients, ns = $p > .05$ (not significant), * $p \leq .05$ (significant), ** $p \leq .01$ (highly significant) and *** $p \leq .001$ (extremely significant).

anchored in real-world scenarios, do not offer the same level of emotional escape, providing a more detached emotional experience.

Real-world relevance. Nonfiction genres with pronounced real-world relevance, such as politics or memoirs, elicited stronger emotional reactions, establishing that readers value content that is both informative and transformative. It's crucial for marketers promoting nonfiction to stress its relevance and potential self-insights (transformative value) and potential informative value (insights gained) (Rice, 2000).

Impact of the genre, and depth of emotional response on enjoyment. The stepwise multiple regression analysis showed, in the fiction genre, a surprising negative relationship between anger and ratings in fantasy ($\beta = -.39$) and young adult ($\beta = -.29$) genres. This suggests that readers of these genres may be more sensitive to anger, which could impact their overall enjoyment of the books. Marketers (and pedagogues) should consider highlighting other emotions, such as joy or fear, in promotional materials for these genres to appeal to their target audiences. The paranormal romance genre in fiction showed a strong negative relationship with disgust ($\beta = -.68$), indicating that readers of this genre may be particularly averse to disgust, which could negatively impact their enjoyment and ratings of the

books. To target this audience effectively, marketers should avoid promoting elements that could evoke disgust and instead emphasise the romantic and supernatural aspects of these stories. In the nonfiction genre, the memoir and autobiography genre showed a negative relationship with anger ($\beta = -.32$), disgust ($\beta = -.38$) and surprise ($\beta = -.27$), while exhibiting a positive relationship with fear ($\beta = .21$). This was an interesting finding, as it implies that readers appreciate a balance of emotions in these books, preferring fear over other negative emotions. Marketers should focus on promoting the emotional depth and range of these books while emphasising the fear aspect to attract readers. The business and money genre had a high *R*-square value of .75, indicating a strong relationship between the emotional variables and book ratings. The negative relationships with anger ($\beta = -.41$), disgust ($\beta = -.37$) and sadness ($\beta = -.36$) suggest that readers of this genre appreciate books that minimise negative emotions. Marketers should emphasise the practical, informative and positive aspects of business and money books to appeal to this audience.

Conclusion

This study incorporated advanced machine learning techniques into marketing and consumer decision-making, focussing on the analysis

of emotions extracted from extensive sources of online reviews. With a primary goal of understanding the emotional landscape in both fiction and nonfiction books, the research addresses two key research objectives (RO1 and RO2) related to insights gained from emotions and their correlation with the success of books.

Insights from emotions in online reviews

Emotional intensity in marketing research: Fiction and nonfiction. The study successfully applied the TTL method of machine learning to accurately identify and interpret consumers' emotive responses from a vast base of online reviews, specifically those on Goodreads. This sheds light on the profound emotional intensity that readers associate with both fiction and nonfiction books. The use of machine learning in this area enables accurate emotion detection despite the emotional complexity present in large-scale online reviews, highlighting the significant potential of this tool for marketing research focussed on online customer feedback.

Emotional responses to fiction and nonfiction. Addressing Research Question 1 (RQ1), the analysis found that fiction books elicited a diverse range of emotions, including negative ones such as anger, sadness and surprise. In contrast, nonfiction prompts slightly elevated levels of joy. Fiction's propensity to evoke a broader spectrum of emotions, including negative ones, potentially impacts its joy score (S. Brown & Patterson 2010; Driscoll & Rehberg Sedo, 2019). This emotional richness in fiction is attributed to its nuanced exploration of complex character dynamics, ethical quandaries and dramatic events. Understanding these emotional patterns holds considerable value for authors, publishers and readers, enhancing their ability to appreciate the unique emotional complexities of different genres. Fiction, inherently hedonic, frequently evoked a rich array of emotions, answering our first research question (RQ1), with a resounding 'Yes' and expands scholars' understanding of diverse emotional responses to fiction and non-fiction books.

Correlation between emotional themes and book success

In alignment with RQ2, the study meticulously examined the influence of emotional intensity on the success and popularity of books across fiction and nonfiction genres.

Genre-specific emotional preferences. An in-depth examination of top book reviews in subgenres provided insights into the unique emotional predilections of readers across different genres. In fiction, genres such as historical fiction and fantasy demonstrated a taste for a blend of emotions, with heightened levels of disgust and fear and reduced levels of anger. On the other hand, young adult fiction, romance and science fiction genres showed a predilection for less negative emotions (anger and disgust), suggesting a desire for a more emotionally balanced and enjoyable reading experience by their readers. Nonfiction genre reviews also exhibited variable emotional preferences. Memoirs and autobiographies were preferred for fear (low anger, disgust and surprise), while self-help and history books indicated a preference for less negative emotions, with a particular leaning towards fear in history books. This suggests that readers of nonfiction genres value a balance of emotions that align with the subject matter, with fear playing a more pronounced role in certain genres.

Authors, movie makers and content creators who combine various genres can gain insight into which emotional responses are likely to generate either positive, negative or destructive emotional

responses. This information will enable authors, publishers and marketers to tailor their offerings to their audience's preferences.

Correlation between emotions and success. Research Question 2 (RQ2) investigated the correlation between emotions expressed in online reviews and the success of a book within its respective genre. A compelling relationship was found between the dominant emotional theme in reader reviews and a book's success within its genre. The correspondence analysis unveils a robust correlation between genres and emotions in literature, with certain genres closely aligned with specific emotions. The findings further illuminate nascent theory on the link between emotional impact and sales: readers' perceptions of a book's value are closely linked to the book's subsequent success.

Theoretical and managerial implications

Theoretical implications. The study's academic implications challenge traditional beliefs about the relationship between fiction and nonfiction genres and emotional responses. It expands scholars' understanding of the different types of literature and which emotional responses they evoke. Emotion analysis in literary studies is highlighted as a valuable tool, fostering an interdisciplinary dialogue between the humanities and data science (machine learning), social sciences (marketing) and humanities (literature and visual arts).

Using insights from consumer reviews online to understand consumer emotional responses to information sources and other forms of story-telling and entertainment, obtained in empirical research based on large data sets online, prompts extended studies into other knowledge/entertainment products such as movies, computer games, television programmes, eSports and various documentaries. Of particular interest to scholars, educators and teachers might be simulations, cases, online courses and other digital development forums, where the online reviews might be fewer, but the basic research questions and methodology will still produce valid, robust results.

Managerial implications. Significant managerial implications are evident, as emotional intensity and review commentaries align well with the popularity of books. The methodology and proposed system for detecting emotions from early reviews can anticipate books' popularity and sales, guiding book marketers and sellers in potential strategies for niche marketing using emotional cues. This knowledge can be used to customise content and promotional strategies to cater to the emotional needs of the target audience, enhancing reader engagement and satisfaction.

Market sensing – gathering market data and using it to guide planning and decision-making – is becoming increasingly important to the success of any business enterprise, including book-writing, book-publishing and niche marketing (Flahive, 2017; Kinberg, 2014; Ramdarshan Bold, 2018). Authors and publishers can tailor the storyline to evoke specific emotions that resonate with the target audience (S. Brown, 2011), while cover designers can create visuals that accurately represent the emotional themes of the book.

Research limitations and future directions

Despite its robust methodology and significant findings, this study acknowledges certain limitations inherent in online review research. Dependence on Goodreads reviews introduces a potential selection bias as the platform's users may not fully represent overall readership or all books, and the study's restriction to publicly available information limits a more comprehensive analysis, with a lack of control variables. Research could expand on this study by investigating the

emotional intensity of books in other languages and cultural contexts, and exploring factors such as author gender, book format and publication date. Additionally, researchers could explore the role of emotional intensity in predicting book sales and reader ratings, and provide further insights for the literary industry. Various factors such as the impact of celebrity influencers, book clubs and other social cues could be considered in configurations of conditions likely to affect consumer choice. Further, the impact of complex factors on non-liking and consumer dissonance or dissatisfaction (Hamza & Zakkariya, 2014) could be explored. Deep dives into consumers' emotional responses and behavioural data will be more important in the future than now or in the recent past. Understanding which emotional levers will encourage specific consumers or consumer segments to spend can make a significant contribution to both bottom-line and top-line numbers.

Finally, this study contributes to the integration of machine learning into marketing and consumer decision-making, providing insights into the intricate emotional complexities associated with fiction and nonfiction books. The findings challenge traditional beliefs, inform marketing strategies and shape academic discourse. Despite its limitations, the study suggests future research directions to explore the role of emotions in predicting book success and extends the epistemology to various knowledge/entertainment products beyond books. Ultimately, the study enhances our understanding of the emotional experiences of readers, guiding content creators, marketers and educators in meeting the diverse needs of the book-loving community.

Author contributions

Sanghyub John Lee drafted the manuscript and designed the study. All authors considered the results and approved the final manuscript.

Acknowledgements

We would like to express our gratitude to our colleagues and peers who provided insight and expertise that greatly assisted the research.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate

Not applicable

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Sanghyub John Lee  <https://orcid.org/0000-0001-6714-0225>

Rouxelle de Villiers  <https://orcid.org/0000-0001-5015-9915>

References

- Abdul-Mageed, M., & Ungar, L. (2017, July). *Emonet: Fine-grained emotion detection with gated recurrent neural networks* [Paper presentation]. 55th Annual Meeting of the Association for Computational Linguistics, (Volume 1: Long papers), Vancouver, Canada, 30 July-4 August 2017.
- Alba, J. W., & Williams, E. F. (2013). Pleasure principles: A review of research on hedonic consumption. *Journal of Consumer Psychology*, 23(1), 2–18.
- Aldama, F. L. (2015). The science of storytelling: Perspectives from cognitive science, neuroscience and the humanities. *Projections*, 9(1), 80–95.
- Barnes, J. L. (2018). Imaginary engagement, real-world effects: Fiction, emotion, and social cognition. *Review of General Psychology*, 22(2), 125–134.
- Baron, N. S. (2015). *Words onscreen: The fate of reading in a digital world*. Oxford University Press.
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1), 1–23.
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205.
- Bougie, R., Pieters, R., & Zeelenberg, M. (2003). Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. *Journal of the Academy of Marketing Science*, 31(4), 377–393.
- Brown, L. (2023). US sales of print books dropped 6.5% in 2022, but adult fiction sees increase. *The Bookseller*. <https://www.thebookseller.com/news/us-sales-of-print-books-dropped-65-in-2022-but-adult-fiction-sees-increase>.
- Brown, S. (2011). And then we come to the brand: Academic insights from international bestsellers. *Arts Marketing: An International Journal*, 1(1), 70–86.
- Brown, S., & Patterson, A. (2010). Selling stories: Harry Potter and the marketing plot. *Psychology & Marketing*, 27(6), 541–556.
- Burns, M. S., Griffin, P., & Snow, C. E. (1999). *Starting out right: A guide to promoting children's reading success*. National Academy Press.
- Chatterjee, A., Narahari, K. N., Joshi, M., & Agrawal, P. (2019, June). *SemEval-2019 task 3: EmoContext contextual emotion detection in text* [Conference session]. 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, 6-7 June 2019 (pp. 39–48).
- Chitturi, R., Raghunathan, R., & Mahajan, V. (2007). Form versus function: How the intensities of specific emotions evoked in functional versus hedonic trade-offs mediate product preferences. *Journal of Marketing Research*, 44(4), 702–714.
- Chu, W., Roh, M., & Park, K. (2015). The effect of the dispersion of review ratings on evaluations of hedonic versus utilitarian products. *International Journal of Electronic Commerce*, 19(2), 95–125.
- Clausen, S. E. (1998). *Applied correspondence analysis: An introduction*. Sage.
- Clement, M., Fabel, S., & Schmidt-Stolting, C. (2006). Diffusion of hedonic goods: A literature review. *The International Journal on Media Management*, 8(4), 155–163.
- Collins, S. (2008). *The hunger games*. Scholastic.
- Davenport, T. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Harvard Business Review Press.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv pre-print. arXiv:1810.04805*.
- Dill-Shackleford, K. E., Vinney, C., & Hopper-Losenicky, K. (2016). Connecting the dots between fantasy and reality: The social psychology of our engagement with fictional narrative and its functional value. *Social and Personality Psychology Compass*, 10(11), 634–646.
- Driscoll, B., & Rehberg Sedo, D. (2019). Faraway, so close: Seeing the intimacy in Goodreads reviews. *Qualitative Inquiry*, 25(3), 248–259.
- Eekhof, L. S., Van Krieken, K., Sanders, J., & Willems, R. M. (2023). Engagement with narrative characters: The role of social-cognitive abilities and linguistic viewpoint. *Discourse Processes*, 60(6), 411–439.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), *Symposium on motivation, 1971* (Vol. 19, pp. 207–280). University of Nebraska Press.
- Errera, R. (2023). *Printed books vs eBooks statistics, trends and facts*. TonerBuzz. <https://www.tonerbuzz.com/blog/paper-books-vs-ebooks-statistics/>
- Flahive, G. (2017). Digital self-promotion for the underdog author: Creative opportunities and experimentation. *Interscript*, 1(2), 23–42.
- Gerard, P. (2017). *Creative nonfiction: Researching and crafting stories of real life*. Waveland Press.
- Grant, R., Clarke, R. J., & Kyriazis, E. (2013). Modelling real-time online information needs: A new research approach for complex

- consumer behaviour. *Journal of Marketing Management*, 29(7–8), 950–972.
- Greenacre, M. (2017). *Correspondence analysis in practice*. CRC Press.
- Hamza, V. K., & Zakkariya, K. A. (2014). A study on the dimensions of customer expectations and their relationship with cognitive dissonance. *Journal of Management*, 8(1), 1–13.
- Jacobs, A. (2011). *The pleasures of reading in an age of distraction*. Oxford University Press.
- John, E. (2017). Empathy in literature. In H. Maibom (Ed.), *The Routledge handbook of philosophy of empathy* (pp. 306–316). Routledge.
- Keen, S. (2006). A theory of narrative empathy. *Narrative*, 14(3), 207–236.
- Kesson, A., & Smith, E. (2016). Introduction: Towards a definition of print popularity. In A. Kesson (Ed.), *The Elizabethan top ten: Defining print popularity in early modern England* (pp. 1–15). Routledge.
- Kim, E., & Klinger, R. (2019). An analysis of emotion communication channels in fan fiction: Towards emotional storytelling. *arXiv preprint. arXiv:1906.02402*.
- Kinberg, M. (2014). Market sensing as a tool for fiction authors. *Journal of Marketing & Management*, 191, 45–57.
- Kivetz, R., & Simonson, I. (2002). Earning the right to indulge: Effort as a determinant of customer preferences toward frequency program rewards. *Journal of Marketing Research*, 39(2), 155–170.
- Klauda, S. L. (2009). The role of parents in adolescents' reading motivation and activity. *Educational Psychology Review*, 21(4), 325–363.
- Klimmt, C. (2008). Escapism. In W. Donsbach (Ed.), *The international Encyclopedia of communication* (Blackwell). Wiley Publishing.
- Kronrod, A., & Danziger, S. (2013). "Wii will rock you!" The use and effect of figurative language in consumer reviews of hedonic and utilitarian consumption. *Journal of Consumer Research*, 40(4), 726–739.
- Lee, H. (1960). *To kill a mockingbird*. Harper & Row.
- Lee, S. J., Kishore, S., Lim, J., Paas, L., & Ahn, H. S. (2021, December 7–10). *Overwhelmed by fear: Emotion analysis of COVID-19 vaccination tweets* [Conference session]. IEEE Region 10 Conference (TENCON), Auckland, New Zealand (pp. 429–434). IEEE.
- Lee, S. J., Lim, J., Paas, L., & Ahn, H. S. (2023). Transformer transfer learning emotion detection model: Synchronizing socially agreed and self-reported emotions in big data. *Neural Computing and Applications*, 35, 10945–10956.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint. arXiv:1907.11692*.
- Lu, J., Liu, Z., & Fang, Z. (2016). Hedonic products for you, utilitarian products for me. *Judgment and Decision Making*, 11(4), 332–341.
- Ludwig, S., De Ruyter, K., Friedman, M., Brüggem, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, 77(1), 87–103.
- Luyckx, K., Vaassen, F., Peersman, C., & Daelemans, W. (2012). Fine-grained emotion detection in suicide notes: A thresholding approach to multi-label classification. *Biomedical Informatics Insights*, 5(1), 61–69. <https://doi.org/10.4137/BII.S8966>
- Maity, S. K., Kumar, A., Mullick, A., Choudhary, V., & Mukherjee, A. (2018, January). *Understanding book popularity on Goodreads* [Conference session]. 2018 ACM International Conference on Supporting Group Work, Sanibel Island, FL, USA, 7–10 January 2018 (pp. 117–121).
- Maity, S. K., Panigrahi, A., & Mukherjee, A. (2017, July 31–August 3). *Book reading behavior on Goodreads can predict the Amazon best sellers* [Conference session]. 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, Sydney, Australia. <https://doi.org/10.1145/3110025.3110138>
- May, F., & Irmak, C. (2014). Licensing indulgence in the present by distorting memories of past behavior. *Journal of Consumer Research*, 41(3), 624–641.
- Merga, M. K. (2017). What motivates avid readers to maintain a regular reading habit in adulthood? *The Australian Journal of Language and Literacy*, 40(2), 146–156.
- Mihart, C. (2012). Impact of integrated marketing communication on consumer behaviour: Effects on consumer decision-making process. *International Journal of Marketing Studies*, 4(2), 121–129.
- Mohammad, S., Bravo-Marquez, F., Salameh, M., & Kiritchenko, S. (2018, June). *Semeval-2018 task 1: Affect in tweets* [Paper presentation]. 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana, 5–6 June 2018.
- Moran, R. (1994). The expression of feeling in imagination. *The Philosophical Review*, 103(1), 75–106.
- Nathanson, S. (2006). Harnessing the power of story: Using narrative reading and writing across content areas. *Reading Horizons: A Journal of Literacy and Language Arts*, 47(1), 2.
- Penz, E., & Hogg, M. K. (2011). The role of mixed emotions in consumer behaviour: Investigating ambivalence in consumers' experiences of approach-avoidance conflicts in online and offline settings. *European Journal of Marketing*, 45(1/2), 104–132.
- Ramdarshan Bold, M. (2018). The return of the social author: Negotiating authority and influence on Wattpad. *Convergence*, 24(2), 117–136.
- Rice, A. M. (2000). The rise of 'good reading' over 'good writing': How and why women's magazine fiction changed in the 1950s and 1960s. *Media History*, 6(2), 139–150.
- Rocklage, M. D., & Fazio, R. H. (2020). The enhancing versus backfiring effects of positive emotion in consumer reviews. *Journal of Marketing Research*, 57(2), 332–352.
- Ross, A., & Willson, V. L. (2017). Paired samples T-test. In A. Ross (Ed.), *Basic and advanced statistical tests* (pp. 17–19). Brill.
- Sailunaz, K., Dhaliwal, M., Rokne, J., & Alhajj, R. (2018). Emotion detection from text and speech: A survey. *Social Network Analysis and Mining*, 8(1), 1–26.
- Saravia, E., Liu, H.-C. T., Huang, Y.-H., Wu, J., & Chen, Y.-S. (2018, October). *CARER: Contextualized affect representations for emotion recognition* [Paper presentation] 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018 (pp. 36687–3697). Association for Computational Linguistics.
- Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour*, 11(3), 234–243.
- Schultz, W. (2022). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, 18(1), 23–32.
- Stankevich, A. (2017). Explaining the consumer decision-making process: Critical literature review. *Journal of International Business Research and Marketing*, 2(6), 7–14.
- Stokmans, M. J. (1999). Reading attitude and its effect on leisure time reading. *Poetics*, 26(4), 245–261.
- Tellis, G. J., MacInnis, D. J., Tirunillai, S., & Zhang, Y. (2019). What drives virality (sharing) of online digital content? The critical role of information, emotion, and brand prominence. *Journal of Marketing*, 83(4), 1–20.
- Tolkien, J. R. R. (1954). *The lord of the rings: The fellowship of the ring*. Allen & Unwin.
- Vieira, V. A., Rafael, D. N., & Agnihotri, R. (2022). Augmented reality generalizations: A meta-analytical review on consumer-related outcomes and the mediating role of hedonic and utilitarian values. *Journal of Business Research*, 151, 170–184.
- Voss, K. E., Spangenberg, E. R., & Grohmann, B. (2003). Measuring the hedonic and utilitarian dimensions of consumer attitude. *Journal of Marketing Research*, 40(3), 310–320.
- Walker, A. (1982). *The color purple*. Harcourt Brace Jovanovich.
- Yin, D., Bond, S. D., & Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research*, 54(3), 447–463.